Appendix E Bootstrap and permutation techniques

The purpose of this appendix is to clarify (a) the similarities and differences between the bootstrap and permutation confidence intervals (CIs), and (b) the role of both techniques in the broader context of parametric and non-parametric hypothesis testing. Note that the bootstrap has been advocated in the Biomechanics literature for trajectory-level analysis. The permutation technique is used in the main manuscript because it is more generalizable than the bootstrap. Interested readers may wish to consult Good (2005) for a more thorough treatment of these topics for 0D datasets, and to Nichols and Holmes (2002) for a discussion of how these techniques extend to 1D and higher-dimensional data.

Sections E.1 and E.2 below analyze the following eight-response dataset:

117 104 110 122 119 90 110 97

Section E.1 computes CIs for this dataset using three different techniques, and Section E.2 conducts one-sample hypothesis testing using four different techniques. Table E1 below summarizes the results of those analyses. Considering these results briefly, it is clear that all techniques produce similar, albeit non-identical CIs and p values. To emphasize why these results are similar but not identical, Section E.3 repeats the three CI techniques for thousands of random (Gaussian) datasets to demonstrate why all techniques may be regarded as theoretically equivalent when the data are normally distributed. This Appendix thus shows that it is sufficient in the main manuscript to compare a single parametric technique (which assumes normality) to a single non-parametric technique (which does not assume normality).

Table E1: Confidence intervals (CI) and one-sample hypothesis tests computed using four different techniques, based on the dataset above.

Class	Technique	95% CI	One-sample test
Non-parametric	Bootstrap	[98.4, 117.5]	$p = 0.07559$
Non-parametric	Permutation	[98.9, 118.3]	$p = 0.06250$
Non-parametric	Wlicoxon		$p = 0.06735$
Parametric	Student's t	[99.3, 117.9]	$p = 0.06411$

E.1 Confidence intervals (CIs)

E.1.1 Bootstrap CI

A simple bootstrap CI can be computed as follows:

(a) Compute the sample mean (in this case: 108.625).

(b) Label the responses as follows:

A	B	C	D	E	F	G	H
117	104	110	122	119	90	110	97

(c) Resample with replacement: select a random set of labels, allowing labels to repeat, then compute the mean for the resampled data. For example, a labeling of "AABBBCDE" has responses: [117, 117, 104, 104, 104, 110, 122, 119], and a sample mean of: 112.125.

(d) Repeat (c) many times and store all sample means. Stop either when (i) all possible resamplings have been made (i.e. AAAAAAAA through HHHHHHHH), or when (ii) a specified number of iterations (e.g. 1000) has been completed.

(e) After all sample means have been accumulated, find the value C_{upper} above which only 2.5% of all estimates traverse, and the value C_{lower} below which only 2.5% of all estimates traverse. The CI is $[C_{\text{lower}}, C_{\text{upper}}]$.

As specified in Table E1 above this procedure yields a CI of [98.4, 117.5].

E.1.2 Permutation CI

A simple permutation CI can be computed as follows:

(a) Compute the sample mean (in this case: 108.625) .

(b) Subtract the sample mean from all responses, then label each observation as "+1":

+1	+1	+1	+1	+1	+1	+1	+1
8.375	–4.625	1.375	13.375	10.375	–18.625	1.375	–11.625

(c) Resample without replacement: permute using either a "+1" or a "–1" label for each response, then multiply each response by each label. For example, a labeling of "+1 +1 +1 –1 –1 –1 +1 –1" produces the new sample "8.375 –4.625 1.375 –13.375 –10.375 18.625 1.375 11.625". For each new sample compute the one-sample t statistic. If there are n responses, there are 2^n possible labelings (256 in this case).

(d) Repeat (c) many times and store all t statistic values for all resamplings. Stop either when (i) all possible resamplings have been made (i.e. "+1 +1 +1 +1 +1 +1 +1 +1" through "–1 –1 –1 –1 –1 –1 –1 –1"), or when (ii) a specified number of iterations (e.g. 1000) has been completed.

(e) After all t statistic values have been accumulated, find the critical height above which only 2.5% of t statistic values traverse, then compute the CI according to Appendix F.

This results in a CI of [98.9, 118.3] (Table E1).

E.1.3 Parametric CI

The parametric CI can be computed using the critical height h^*, which is defined via the one-sample t statistic distribution (see Appendix F, and in particular the "One-sample" row of Table F3). This procedure yields a CI of [99.3, 117.9], which is very similar to both the bootstrap and permutation results.

E.1.4 Comparison of CI results

All three techniques yield similar, but non-identical results. Since the parametric technique assumes that the data come from a normal (Gaussian) distribution, all CI techniques should, by definition, converge to the identical value when (a) the data are normal and (b) the sample size is large. The different techniques will only produce precisely the same result when the sample size is infinitely large, as we will see in Section E.3 below. Investigators must therefore judge whether the discrepancies amongst the techniques is negligible or non-negligible. Evidence of departure from normality, for example, would be a good reason to choose one of the non-parametric techniques. For the results above (Table E1), the discrepancies amongst the different CI techniques are likely negligible for most applications. The main point is that the three CI techniques are theoretically equivalent when the data are normally distributed.

E.2 Hypothesis tests

Thorough descriptions of one-sample hypothesis tests using the bootstrap, permutation, Wilcoxon and parametric (one-sample t test) techniques can be found in many statistics textbooks so in interest of brevity are not repeated here. Additionally, as will be shown in Appendix F, CIs are equivalent to one-sample hypothesis tests, so re-describing the techniques here would be redundant. This section therefore just focusses on the results in Table E1 above.

Like the CI results, all four hypothesis test procedures produce similar, but non-identical p values. For classical hypothesis testing, the null hypothesis would not be rejected for any of the four tests at α=0.05. The next section explores why these four approaches are theoretically equivalent (when the data are normal) even when the results are not precisely equivalent.

E.3 Convergence of CIs

Repeating the bootstrap, permutation and parametric CI procedures on thousands of random datasets (drawn from the Gaussian distribution) of increasingly larger sample sizes yields the results in Fig.E1. The two non-parametric CIs clearly converge to the parametric CI as sample size increases, implying theoretical equivalence amongst the three procedures (when the data are normally distributed).

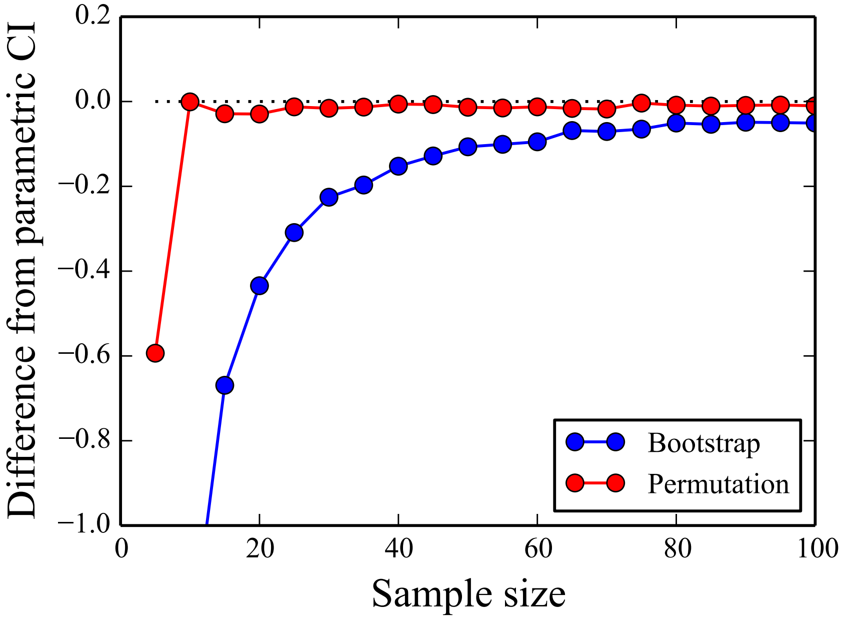

Figure E1: Convergence of the CI for three estimation procedures. These results were obtained by: (i) producing a random sample from the Gaussian distribution of the given sample size, with a mean of 100 and a variance of 10, (ii) estimating the CI using the three procedures indicated (Bootstrap, Permutation, Parametric), and (iii) repeating 500 times for each sample size. Single results depict the mean values across the 500 repetitions.

Summary

This Appendix has shown that there is fundamentally little difference between the bootstrap and permutation approaches, but that they might produce non-negligibly different numerical results in certain situations, like when sample sizes are very small. The larger point is that the bootstrap procedure is not particularly unique, as has been implied in the literature. Instead the bootstrap procedure yields a solution which can also be obtained using other techniques, and its scope is also relatively limited in the broader context of generalized hypothesis testing (Fig.E2).

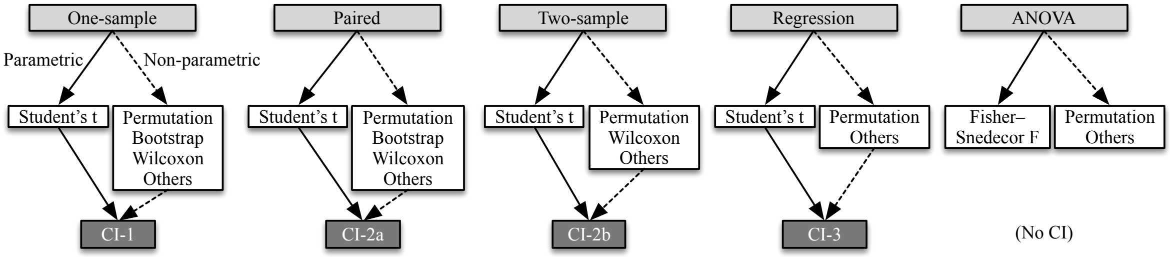

Figure E2: Context of the bootstrap (for both 0D and 1D tests). Light grey, white and dark grey boxes respectively depict: experimental designs, statistical inference procedures, and confidence intervals (CI).