Jackknife and Bootstrap Resampling Methods in Statistical Analysis to Correct for Bias

Peter Young
Physics Department, University of California Santa Cruz, Santa Cruz CA 95064

I. THE BASICS

In this section we review the standard statistical method for estimating the mean of a set of data with noise, as well as the error bar in the mean.

We are given a set (sample) of N random variables, $x_i,\ i = 1, 2, \cdots, N$, which are uncorrelated. We denote *sample* averages by an overbar, so, for example, the sample mean and variance are given by

$$\overline{x} \ = \ \frac{1}{N}\sum_{i=1}^{N} x_i \,, \tag{1}$$

$$\sigma^2_{\rm samp} \ = \ \overline{x^2} - \overline{x}^2 \,. \tag{2}$$

We also denote *exact* averages by angular brackets, $\langle \cdots \rangle$, *e.g.* for the mean, which we also write as X,

$$X \equiv \langle x \rangle = \int x P(x)\, dx, \tag{3}$$

where $P(x)$ is the distribution of x from which the sample is drawn. Similarly the exact variance, σ^2, is given by

$$\sigma^2 \equiv \langle x^2 \rangle - \langle x \rangle^2 = \langle (x - \langle x \rangle)^2 \rangle. \tag{4}$$

We have available the single set of data x_i. However, it is conceptually useful to imagine performing *many repetitions* (of the experiment or simulation). If the number of repetitions is large enough, then an average over them will, to a good approximation, correspond to the exact average. For example, if we average the sample mean over many repetitions we get the exact mean of x since

$$\langle \overline{x} \rangle = \frac{1}{N}\sum_{i=1}^{N} \langle x_i \rangle = \langle x \rangle \equiv X. \tag{5}$$

This result is obvious but also useful because it tells us that the sample mean is an *unbiased estimate* for the exact mean, in the sense that the average of the sample mean, over many repetitions, is the exact mean. An unbiased estimate will become more and more accurate as the number of data points is increased. However, a biased estimate not continue to improve with increasing N once the error is smaller than the bias. Hence *we should work with unbiased estimators.*

We will also be interested in the variance of the sample mean (again averaged over many repetitions). We find that

$$\sigma^2_{\overline{x}} \equiv \langle \overline{x}^2 \rangle - \langle \overline{x} \rangle^2 = \frac{1}{N^2}\langle \sum_{i,j} x_i x_j \rangle - \frac{1}{N^2}\langle \sum_i x_i \rangle^2 = \frac{N(N-1)}{N^2}\langle x \rangle^2 + \frac{N}{N^2}\langle x^2 \rangle - \langle x \rangle^2 = \frac{1}{N}\left(\langle x^2 \rangle - \langle x \rangle^2 \right) = \frac{\sigma^2}{N} \,. \tag{6}$$

However, this involves the variance of the exact distribution, which in general we don't know. Rather we want to get an estimate of the variance of the sample mean using only data in the sample. We shall now see that an *estimate* of the exact variance σ^2 can be obtained from the variance of the (single) sample, $\sigma^2_{\rm samp}$, defined in Eq. (2). Averaging $\sigma^2_{\rm samp}$ over many repetitions gives

$$\langle \sigma^2_{\rm samp} \rangle = \frac{1}{N}\langle \sum_i x_i^2 \rangle - \frac{1}{N^2}\langle \sum_{i,j} x_i x_j \rangle = \frac{N-1}{N}\left(\langle x^2 \rangle - \langle x \rangle^2 \right) \equiv \frac{N-1}{N}\sigma^2. \tag{7}$$

Hence, from Eqs. (6) and (7), an *unbiased* estimate for $\sigma^2_{\overline{x}}$ is $(\delta \overline{x})^2$ where

$$\boxed{(\delta \overline{x})^2 = \frac{\sigma^2_{\rm samp}}{N-1} \,.} \tag{8}$$

We will use $\delta\overline{x}$ to be a measure of the uncertainty in the sample mean. Hence our estimate of the exact average $\langle x\rangle$ from the single set of data, $x_i,\ (i = 1, \cdots, N)$, is

$$\boxed{\overline{x} \pm \delta\overline{x} = \overline{x} \pm \frac{\sigma_{\rm samp}}{\sqrt{N-1}}\,.} \tag{9}$$

Note that it just involves the sample mean $\overline{x}$ and sample standard deviation $\sigma_{\rm samp}$. Usually N is large, in which case the difference between $N-1$ and N is unimportant.

II. HOW CAN BIAS ARISE?

In the previous section we saw that $\overline{x}$ provides an *unbiased* estimate of the exact average and $\sigma_{\rm samp}/\sqrt{N-1}$ provides an unbiased estimate of the uncertainty in this average.

Now suppose that we generate random values for the x_i but want to estimate not the average of x itself but *some function of the average of* x, *i.e.* $f(\langle x\rangle) \equiv f(X)$. As an example, we may want to compute $\langle x\rangle^2$, in which case $f(X) = X^2$. A poor way to estimate this would be from

$$\overline{f} \equiv \overline{f(x)} = \frac{1}{N}\sum_{i=1}^{N} f_i \quad \text{where } f_i = f(x_i). \tag{10}$$

However, this is really an estimate for $\langle f(x)\rangle$, rather than $f(\langle x\rangle)$. As an estimate for the latter it is bad because it is *biased*, *i.e.* repeating the set of N measurements M times (where M is large) and averaging over them does give the exact answer for $M \to \infty$. In other words

$$\langle\, \overline{f}\,\rangle \neq f(X). \tag{11}$$

The difference is given by

$$\begin{aligned}
\langle\, \overline{f}\,\rangle - f(X) &= \int P(x)\,(f(x) - f(X))\,dx \\
&= f'(X)\int P(x)(x-X)\,dx + \frac{1}{2}f''(X)\int P(x)(x-X)^2\,dx + \cdots \\
&= \frac{1}{2}f''(X)\left[\langle x^2\rangle - \langle x\rangle^2\right] + \cdots \\
&= \frac{1}{2}f''(X)\sigma^2 + \cdots\ ,
\end{aligned} \tag{12}$$

which is of order *unity*, *i.e.* it does not vanish for $N \to \infty$. The third line follows because $X \equiv \langle x\rangle = \int P(x)x\,dx$ and $P(x)$ is normalized *i.e.* $\int P(x)\,dx = 1$. The neglected terms in Eq. (12) involve higher moments of the distribution of $x - X$. As expected, there is no bias if f is a linear function because then $f'' = 0$ and the neglected terms involve derivatives higher than the second which also vanish.

A better, *i.e.* less biased, estimate for $f(X)$ is clearly $f(\overline{x})$. Repeating the steps in Eq. (12) gives

$$\begin{aligned}
\langle\, f(\overline{x})\,\rangle - f(X) &= \frac{1}{2}f''(X)\left[\langle\overline{x}^2\rangle - \langle\overline{x}\rangle^2\right] + \cdots \\
&= \frac{1}{2N}f''(X)\sigma^2 + \cdots\ ,
\end{aligned} \tag{13}$$

where we used Eq. (6). In other words, the bias is of order $1/N$ rather than order 1. Since the statistical error in $f(\overline{x})$ is of order $1/\sqrt{N}$, which is much bigger than $1/N$ for large N, the bias in $f(\overline{x})$ can usually be neglected.

We have seen that a generally satisfactory estimate for $f(X)$ is given by $f(\overline{x})$. However, we would also like to estimate the error in this estimate. A convenient way to do this is to use jackknife and bootstrap ideas, which are discussed in the rest of this handout.

III. ESTIMATE OF ERROR WHEN THERE IS BIAS USING JACKKNIFE

In the last section we saw that if N is large. then a satisfactory estimate for $f(X)$ is $f(\overline{x})$. However, we do not yet know how to estimate the error in this value. The result in Eq. (9) is just for the average of the data $\langle x\rangle$, not

some function f of the average. In this section we will explain how estimate the error in using $f(\overline{x})$ as an estimate for $f(X)$, do this using jackknife methods.

Clearly $\overline{f^2} - \left(\overline{f}\right)^2$ is the variance of $\overline{f(x)}$ not $f(\overline{x})$, and so cannot be used to get the uncertainty in the latter, since we saw in the previous section that they are quite different. This is where the jackknife and bootstrap resampling methods comes in. This section discusses jackknife and the next section will discuss bootstrap.

We *define* the jackknife averages, x_i^J by

$$x_i^J \equiv \frac{1}{N-1}\sum_{j\neq i} x_j, \tag{14}$$

so x_i^J is the average of all the x-values *except* x_i. Similarly we define

$$f_i^J \equiv f(x_i^J). \tag{15}$$

We now state that the jackknife estimate of $f(X)$ is the average of the f_i^J, *i.e.*

$$\boxed{f(X) \simeq \overline{f^J} \equiv \frac{1}{N}\sum_{i=1}^{N} f_i^J,} \tag{16}$$

and the uncertainty in this estimate, $\sigma_{f(\overline{x})}$, is given by

$$\boxed{\sigma_{f(\overline{x})} = \sqrt{N-1}\,\sigma_{f^J}} \tag{17}$$

where

$$\boxed{\sigma_{f^J}^2 \equiv \overline{\left(f^J\right)^2} - \left(\overline{f^J}\right)^2.} \tag{18}$$

In the rest of this section we will prove these results. Note that the factor of $N-1$ in Eq. (17) is in the numerator, whereas in the corresponding expression for the original (i.e. unjackknifed) quantities, Eq. (9), there is a similar factor in the denominator. One can easily show that if $f(x) = x$, so there is no bias, the jackknife estimates are *identical* to the estimates from the original data points, x_i, *i.e.*

$$\overline{x^J} = \overline{x}, \qquad (N-1)\,\sigma_{x^J}^2 = \frac{\sigma_{\rm samp}^2}{N-1}\,. \tag{19}$$

There is a difference, however, in the presence of bias. We shall now show that Eqs. (16) and (18) give estimates for $f(X)$ and its error which are *unbiased* for large N.

The bias in f_i^J is given by $\langle f_i^J\rangle - f(X)$. Defining $\delta x_i = x_i - X$, then

$$x_i^J \equiv \frac{1}{N-1}\sum_{j\neq i} x_j = X + \frac{1}{N-1}\sum_{j\neq i}\delta x_j. \tag{20}$$

Now the typical value of $(N-1)^{-1}\sum_{j\neq i}\delta x_j$ is of order $1/\sqrt{N}$ for large N, and so is small. Hence it makes sense to expand f_i^J about $f(X)$ using a Taylor series, *i.e.*

$$\begin{aligned}
\langle f_i^J\rangle &= \left\langle f\left(X + \frac{1}{N-1}\sum_{j\neq i}\delta x_j\right)\right\rangle \\
&= f(X) + \frac{f'(X)}{N-1}\sum_{j\neq i}\langle\delta x_j\rangle + \frac{1}{2}\frac{f''(X)}{(N-1)^2}\sum_{j\neq i}\sum_{k\neq i}\langle\delta x_j\delta x_k\rangle + \cdots \\
&= f(X) + \frac{1}{2(N-1)}f''(X)\sigma^2 + \cdots\,.
\end{aligned} \tag{21}$$

Since this expression is independent of i we also have

$$\langle\overline{f^J}\rangle - f(X) = \frac{1}{2(N-1)}f''(X)\sigma^2 + \cdots\,. \tag{22}$$

Higher order terms are of higher order in $1/N$. We see that the bias vanishes for $N \to \infty$, and is essentially the same as the bias in $f(\overline{x})$, see Eq. (13), except that N is replaced by $N-1$. For large N we can, if we wish, eliminate the leading, $O(1/N)$, contribution to the bias and estimate $f(X)$ from

$$\boxed{f(X) \simeq N f(\overline{x}) - (N-1)\overline{f^J}.} \tag{23}$$

The resulting bias will then normally be of order $1/N^2$. However, this $1/N$ correction to the mean is usually negligible because the statistical error is of order $1/\sqrt{N}$, which is much larger if N is large, as is usually the case.

Next we discuss the uncertainty in our estimate of $f(X)$ from $f(\overline{x})$. The standard deviation of $f(\overline{x})$, obtained by repeating the measurements many times, is given by $\sigma_{f(\overline{x})}$ where

$$\sigma^2_{f(\overline{x})} \equiv \langle f^2(\overline{x})\rangle - \langle f(\overline{x})\rangle^2 \,. \tag{24}$$

We evaluate the exact averages by assuming the N is large so $\overline{x}$ differs only slightly from X. This gives

$$\sigma^2_{f(\overline{x})} = \frac{1}{N} f'(X)^2 \sigma^2 + \cdots \,, \tag{25}$$

where the omitted terms are higher order in $1/N$.

We will now estimate estimate the right hand side of Eq. (25) from the variance of the jackknife estimates in Eqs. (18). Averaging Eq. (18) over many repetitions, gives

$$(N-1)\,\langle \sigma^2_{f^J}\rangle \;=\; \frac{1}{N} f'(X)^2 \sigma^2 + \cdots \,, \tag{26}$$

$$\;=\; \sigma^2_{f(\overline{x})} + \cdots \,, \tag{27}$$

where the last line used Eq. (25) and again the omitted terms of higher order in $1/N$. Eq. (27) shows that, for large N, an unbiased estimate for the error in $f(\overline{x})$ is $\sqrt{N-1}\,\sigma_{f^J}$, as stated earlier in Eq. (17).

Rather than using jackknife, one could estimate the error in $f(\overline{x})$, by "error propagation", *i.e.* by directly using Eq. (25). However, the jackknife has the advantage that it automatically takes care of error propagation even in complicated situations where there are many variables for which it would be tedious to compute all the terms contributing to the error.

The above method effectively puts the N data points into N groups each with 1 point. The jacknife acts on N datasets, each of which contains all groups but one. One can *generalize* the jackknife method to grouping the N data points into $N_{\rm group} = N/m$ groups each with m points. The jackknife then acts on $N_{\rm group}$ datasets each of which contains the $N - N_{\rm group}$ data points in all groups but one. The generalization of the expression for the error bar, Eq. (17), is

$$\sqrt{N_{\rm group}-1}\,\sigma_{f^J} \tag{28}$$

and that for the elimination of the leading piece of the bias in the mean, Eq. (23), is

$$N_{\rm group} f(\overline{x}) - (N_{\rm group}-1)\overline{f^J}. \tag{29}$$

We now give some example of the use of jackknife.

A. Example 1

We compute $\cos(\langle x\rangle)$ where $x_i = \pi/3 + \epsilon_i$ where ϵ_i is a Gaussian random variable with zero mean and standard deviation unity. We took a sample of $N = 1000$. The jackknife estimate of the mean was 0.4916, with an error estimate of 0.0280 which is consistent with the exact value of 0.5. For comparison $\cos\overline{x}$ was also equal to 0.4916 to this precision. Using Eq. (23) to get a less biased estimate of the mean, and using the full precision of the numbers in the computer, gives 0.4919. The difference between 0.4916 and 0.4919 is completely unimportant since the error bar of 0.0280 is very much larger.

B. Example 2

Jackknife (and bootstrap) can be used to compute error bars for quite general functions of the data set. For example, one can use the jackknife and bootstrap resampling schemes to estimate parameters describing the *shape* of the distribution from the data set. An example, which is also used in my research in phase transitions (where it is called the "Binder ratio"), is the "kurtosis", g, defined by

$$g = \frac{\langle \delta x^4 \rangle}{\langle \delta x^2 \rangle^2}\,, \tag{30}$$

where $\delta x = x - \langle x \rangle$. Since the total power of x in the numerator and denominator is the same, the kurtosis depends only on the shape of the distribution and not on its overall scale. It takes the value 3 for a Gaussian distribution.

In the example I took $N = 1000$ points from a Gaussian distribution and computed the kurtosis using the jackknife method. (Since I knew that the mean is zero I assumed $\langle x \rangle = 0$.) For each of the N jackknife data sets I computed g and obtained an average and error bar using Eqs. (16) and (17)–(18) above. The result is

$$g = 3.090 \pm 0.145 \tag{31}$$

which is consistent with the exact value of 3.

C. Example 3

It is also argued that the jackknife method can also be used to estimate error bars in fit parameters, see *e.g.* B. Efron *The Jackknife, the Bootstrap and Other Resampling Plans*, SIAM (1982), § 3.8. The idea is to fit all the data points minus one and repeat this for all choices of the missing point. If the i-th data point is omitted we denote the resulting jackknife estimate of the α-th fit parameter, a_α, by $a^J_{i,\alpha}$. According to Efron, the covariance matrix is then given by

$$(\mathrm{COV})_{\alpha,\beta} = \frac{N-1}{N} \sum_{i=1}^{N} \left[a_i^J - \overline{a^J} \right]_\alpha \left[a_i^J - \overline{a^J} \right]_\beta\,, \tag{32}$$

where

$$\overline{a^J_\alpha} = \frac{1}{N} \sum_{i=1}^{N} a^J_{i,\alpha}. \tag{33}$$

We illustrate this with some data. In a quadratic fit of 50 points with Gaussian noise to $3 - 2x + x^2$, the following results were obtained:

```
50 points: Jackknife fits
 fit parameters
     3.01221     1.97836    -1.00207
 error bars
     0.01907     0.08472     0.07789
 covariance matrix from jackknife
     0.00036    -0.00140     0.00113
    -0.00140     0.00718    -0.00642
     0.00113    -0.00642     0.00607
            Least squares fit
 fit parameters
     3.01218     1.97847    -1.00217
 error bars
     0.02123     0.09807     0.09496
 covariance matrix
     0.00045    -0.00180     0.00150
    -0.00180     0.00962    -0.00902
     0.00150    -0.00902     0.00902
chisq =     33.91606
```

Note that the results for the fit parameters from the jackknife and conventional least square fits are similar, though the jackknife somewhat underestimates the error bars (is this a general trend?). If we go to 500 data points, though, the differences in the error bars between the two approaches become tiny:

```
500 points: Jackknife fits
 fit parameters
     3.00663     1.98693    -0.99771
 error bars
     0.00680     0.03117     0.03040
 covariance matrix from jackknife
     0.00005    -0.00018     0.00015
    -0.00018     0.00097    -0.00091
     0.00015    -0.00091     0.00092
            Least squares fit
 fit parameters
     3.00663     1.98693    -0.99771
 error bars
     0.00671     0.03098     0.03000
 covariance matrix
     0.00005    -0.00018     0.00015
    -0.00018     0.00096    -0.00090
     0.00015    -0.00090     0.00090
chisq =    547.36826
```

Is there any advantage of estimating errors in fit parameters from Jackknife rather than a standard least square analysis? Presumably not in this example, but perhaps there could be in more complicated situations.

IV. THE BOOTSTRAP

The bootstrap, like the jackknife, is a resampling of the N data points x_i. Whereas jackknife considers N new data sets, each of containing all the original data points minus 1, bootstrap uses $N_{\rm boot}$ data sets each containing N points obtained by random (Monte Carlo) sampling of the original set of N points. During the Monte Carlo sampling, the probability that a data point is picked is $1/N$ irrespective of whether it has been picked before. (In the statistics literature this is called picking from a set "with replacement".) Hence a given data point x_i will, on average, appear once in each Monte Carlo-generated data set, but may appear not at all, or twice, and so on. The probability that x_i appears n_i times is close to a Poisson distribution with mean unity. However, it is not exactly Poissonian because of the constraint in Eq. (34) below. It turns out that we shall need to include the deviation from Poisson distribution even for large N. We shall use the term "bootstrap" data sets to denote the Monte Carlo-generated data sets.

More precisely, let us suppose that the number of times x_i appears in a Monte Carlo-generated data set is n_i. Since each bootstrap dataset contains exactly N data points, we have the constraint

$$\sum_{i=1}^{N} n_i = N\,. \tag{34}$$

Consider one of N variables x_i. Each time we generate an element in a bootstrap dataset the probability that it is x_i is $1/N$, which we will denote by p. It is then standard that the probability that x_i occurs n_i times is a binomial distribution

$$P(n_i) = \frac{N!}{n_i!\,(N-n_i)!}\,p^{n_i}(1-p)^{N-n_i}\,. \tag{35}$$

It is then also standard that the mean and standard deviation are given by

$$[n_i]_{\rm MC} \;=\; Np = 1\,, \tag{36}$$

$$[n_i^2]_{\rm MC} - [n_i]^2_{\rm MC} \;=\; Np(1-p) = 1 - \frac{1}{N}\,, \tag{37}$$

where $[\ldots]_{\rm MC}$ denotes an exact average over bootstrap samples (for a fixed original data set x_i). For $N \to \infty$, the binomial distribution goes over to a Possion distribution, for which the factor of $1/N$ in Eq. (37) does not appear. We

assume that $N_{\rm boot}$ is sufficiently large that the bootstrap average we carry out reproduces this result with sufficient accuracy. Later, we will discuss what values for $N_{\rm boot}$ are sufficient in practice. Because of the constraint in Eq. (34), n_i and n_j (with $i \neq j$) are not independent and we find from Eq. (37) that

$$[n_i n_j]_{\rm MC} - [n_i]_{\rm MC}[n_j]_{\rm MC} = -\frac{1}{N} \qquad (i \neq j)\,. \tag{38}$$

First we consider bootstrap averages of x, for which, if course, there is no bias, and so bootstrap is not necessary. However, it shows how to get averages and error bars in a simple case, which is then easily generalized to more complicated situations which do have bias.

We denote the average of x for a given Monte Carlo data set by x^B_α, where α runs from 1 to $N_{\rm boot}$, *i.e.*

$$x^B_\alpha = \frac{1}{N}\sum_{i=1}^{N} n_i^\alpha x_i\,. \tag{39}$$

Still for a fixed original data set x_i we compute the bootstrap average of the mean of x and the bootstrap variance in the mean, by averaging over all the bootstrap data sets. We assume that that $N_{\rm boot}$ is large enough for the bootstrap average to be exact, so we can use Eqs. (37) and (38). The result is

$$\overline{x^B} \equiv \frac{1}{N_{\rm boot}}\sum_{\alpha=1}^{N_{\rm boot}} x^B_\alpha \;=\; \frac{1}{N}\sum_{i=1}^{N}[n_i]_{\rm MC}x_i = \frac{1}{N}\sum_{i=1}^{N} x_i = \overline{x} \tag{40}$$

$$\sigma^2_{x^B} \equiv \overline{(x^B)^2} - \left(\overline{x^B}\right)^2 \;=\; \frac{1}{N^2}\left(1-\frac{1}{N}\right)\sum_i x_i^2 - \frac{1}{N^3}\sum_{i\neq j} x_i x_j\,, \tag{41}$$

where

$$\overline{(x^B)^2} \equiv \frac{1}{N_{\rm boot}}\sum_{\alpha=1}^{N_{\rm boot}}\left[\left(x^B_\alpha\right)^2\right]_{\rm MC}\,. \tag{42}$$

We now average Eqs. (40) and (41) over many repetitions of the original data set x_i. Averaging Eq. (40) gives

$$\langle\overline{x^B}\rangle = \langle\overline{x}\rangle = \langle x\rangle \equiv X\,. \tag{43}$$

This shows that the bootstrap average $\overline{x^B}$ is an unbiased estimate of the exact average X. Averaging Eq. (41) gives

$$\left\langle\sigma^2_{x^B}\right\rangle = \frac{N-1}{N^2}\sigma^2 = \frac{N-1}{N}\sigma^2_{\overline{x}}\,, \tag{44}$$

where we used Eq. (6) to get the last expression. Since $\sigma_{\overline{x}}$ is the uncertainty in the sample mean, we see that

$$\sqrt{\frac{N}{N-1}}\sigma_{x^B}\,, \tag{45}$$

is an unbiased estimate of this uncertainty. Remember, σ_{x^B} is the standard deviation of the bootstrap data sets. Usually N is sufficiently large that the square root in Eq. (45) can be replaced by unity.

As for the jackknife, these results can be generalized to finding the error bar in $f(X)$ rather than X. One computes the bootstrap estimates for $f(X)$, which are

$$f^B_\alpha = f(x^B_\alpha)\,. \tag{46}$$

The final bootstrap estimate for $f(X)$ is the average of these, *i.e.*

$$\boxed{\overline{f^B} = \frac{1}{N_{\rm boot}}\sum_{\alpha=1}^{N_{\rm boot}} f^B_\alpha\,.} \tag{47}$$

Following the methods in the jackknife section, one can show that the uncertainty in this estimate is

$$\boxed{\sqrt{\frac{N}{N-1}}\;\sigma_{f^B}}\,, \tag{48}$$

where

$$\boxed{\sigma^2_{f^B} = \overline{(f^B)^2} - \left(\overline{f^B}\right)^2 ,} \tag{49}$$

in which

$$\overline{(f^B)^2} \equiv \frac{1}{N_{\rm boot}} \sum_{\alpha=1}^{N_{\rm boot}} \left(f^B_\alpha\right)^2 . \tag{50}$$

Equation (49) corresponds to the result Eq. (45) which we derived for the special case of $f(x) = x$. Again following the same path as in the jackknife section, it is straightforward to show that Eqs. (47) and (49) are unbiased estimates for N large. However, if N is not too large it may be useful to eliminate the leading contribution to the bias in the mean, as we did for jackknife in Eq. (23). The result is

$$\boxed{2f(\overline{x}) - \overline{f^B} .} \tag{51}$$

The bias in Eq. (51) is of order $1/N^2$, whereas $f(\overline{x})$ and $\overline{f^B}$ each have a bias of order $1/N$.

I have not systematically studied the values of $N_{\rm boot}$ that are needed in practice to get accurate estimates for the error. It seems that $N_{\rm boot}$ in the range 100 to 500 is typically chosen, and this seems to be adequate irrespective of how large N is.

We now show three examples. The first two are the same as the first two in the jackknife section, with the *same* data set x_i.

A. Example 1

With $N_{\rm boot} = 100$ data sets I found 0.4927 ± 0.0279. This is consistent with the exact value of 0.5 and very close to the jackknife result.

B. Example 2

The kurtosis is found for each of the $N_{\rm boot} = 100$ bootstrap samples, and the mean and error bar obtained from Eqs. (47) and (48)–(49). The result is

$$g = 3.072 \pm 0.1320 , \tag{52}$$

which agrees well with the jackknife estimate and is consistent with the exact value of 3.

C. Example 3

For most problems, the jackknife and bootstrap methods give similar results. However, there is at least one class of problems where the jackknife approach is unsatisfactory, because the data sets are too similar to each other, while the bookstrap method works. This is the calculation of the median and related quantities. I am very grateful to Mathias Körner for helpful discussions on this point.

The median is the value of x such that half the data points are greater than x and half less than x. When there is an odd number of data points, the median is simply the middle one. When there is an even number of data points, the median is the average of the two middle ones. Suppose that N is even and we have sorted the x_i, then the median of the whole data set is $(x_{N/2} + x_{(N/2)+1})/2$, while the median of half the jackknife sets is $x_{N/2}$ and that of the other half is $x_{(N/2)+1}$. Since the jackknife data sets only give one of two possible results for the median, the method cannot give a reliable estimate for the error in it. A similar argument shows that also if N is odd only two values for the median are obtained from the jackknife approach.

However, in bootstrap, the $N_{\rm boot}$ data samples are significantly different from each other, so the error in the median can be estimated. As an example, I took $N = 1001$ data points generated from the positive half of a Lorentzian:

$$P(x) = \begin{cases} \dfrac{2}{\pi}\dfrac{1}{1+x^2} & (x > 0) , \\ \\ 0 , & (x < 0) . \end{cases} \tag{53}$$

Note that this is a very broad distribution for which even the mean is not defined. However, the median *is* defined and a standard integral gives the value 1.

Including all 1001 values of x_i, I found the median to be 0.9613. Using bootstrap with $N_{\rm boot} = 5000$ data sets I found 0.9603 ± 0.0507. We see that the overall average and the mean of the bootstraps are very close, and the result agrees with the exact value of 1 within the error bar.