Elements of Seismic Data Processing:
No Removing of a DC Component

Björn E. Rommel

version: 2.1.1

Abstract

Instruments are calibrated by matching their null-line – that is, zero amplitude – to a long-time average of noise samples. Indeed, noise, if generated by a multitude of independent events, tends to be normally distributed, and its mean becomes repeatable. That property, however, holds only for a very large dataset, and not for typically 8 to 12s long seismic records cut out of that dataset. In the following, I will demonstrate how much the means of seismic records will fluctuate, no matter how perfectly calibrated the entire dataset might be.

1 Introduction

Seismic instruments don't have a built-in null-line, that is they typically do not show a perfect zero amplitude in a sound-less environment. Instead, the null-line is determined from a seismic dataset. More precisely, since no environment is free of noise, the null-line is defined to be the average of infinitely many data samples of noise, without any seismic event.

Furthermore, noise, as an infinite number of independent and random events, is typically assumed to be normally distributed. In fact, the central limit theorem states that any distribution of random numbers, not necessarily normally distributed random numbers, approach a normal distribution with increasing sampling number. The point being, with increasing sampling number. Consequently, the null-line of a large dataset can well be calibrated as the mean of that normal distribution.

However, sub-datasets – our seismic records cut out of the entire dataset – contain only a small number of samples. And even if the entire dataset is calibrated, the means of sub-datasets may not; there simply is no mechanism forcing each subset to be balanced. In fact, the means of such sub-datasets form themselves a normal distribution around the mean of the dataset (or population in the language of statisticians).

Consequently, we might well calibrate a seismic instruments from noise samples recorded over weeks and months, with at least a 'large' sampling number. However, once we have extracted seismic records over typically 8 to 10 seconds, their respective means will most likely not be zero.

2 Mean of an Infinite Dataset

Let's do a numerical experiment assuming ideal noise. That is, the noise samples follow a normal distribution with zero mean and one standard deviation. Then, let me generate a large dataset of such noise samples. Large, but nonetheless limited in size. Therefore, its mean will likely deviate from zero (see Fig. 1); and repeatedly generating large datasets will likely yield slightly varying, but nonzero means. The larger the dataset, the less variation there will be.

Although not likely a perfect zero, we define that mean as the desired null-line of the instrument's ideal response. That is, we calibrate the instrument response accordingly.

Now, let's cut the dataset into subsets called (seismic) records, each of which typically contains some 1000's samples, or typically covers some 8 to 12s long. Although the entire dataset has been calibrated, the mean of each record is not (see Fig. 2); indeed, the means fluctuate around the null-line. In fact, the means form themselves a normal distribution (see Fig. 3).

The means of any realization of a normally distributed set of random numbers – our records in the language of statistics – are themselves normally distributed: the mean of means remains 0.0; and its standard deviation decreases with $\sigma/\sqrt{n_{\text{record}}}$, where σ is the standard deviation of random numbers individually. Strictly speaking, the entire dataset is, in fact, just one single realization of random numbers, albeit a large one; and its mean is only close to the ideal 0.0 within a standard deviation of $\sigma/\sqrt{n_{\text{dataset}}}$.

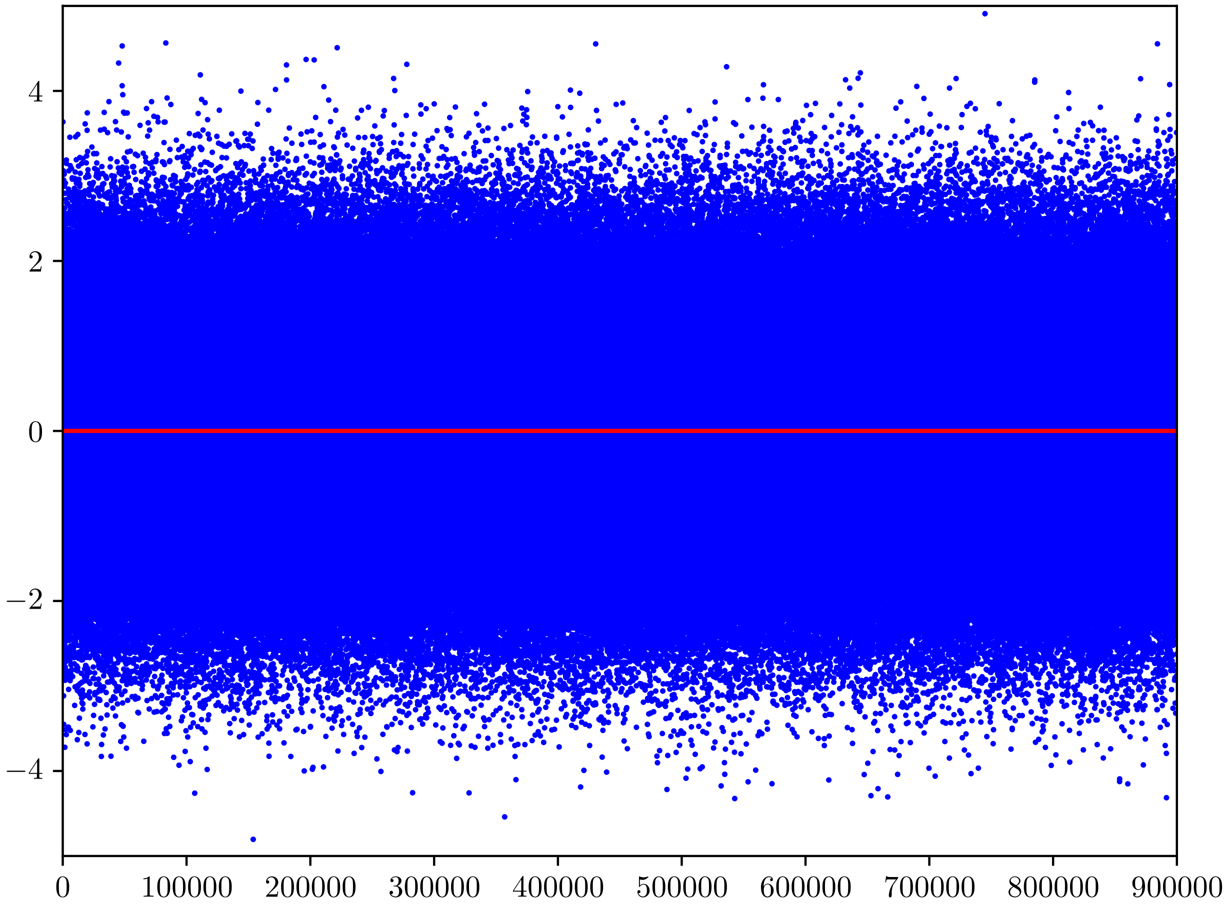

Figure 1: Noise dataset with mean: The noise samples, which are marked by blue dots, are normally distributed around a mean μ_{theory} of 0.0 with a standard deviation σ_{theory} of 1.0. The red line marks the actual mean μ_{dataset} of 0.000503. The dataset will be calibrated accordingly.

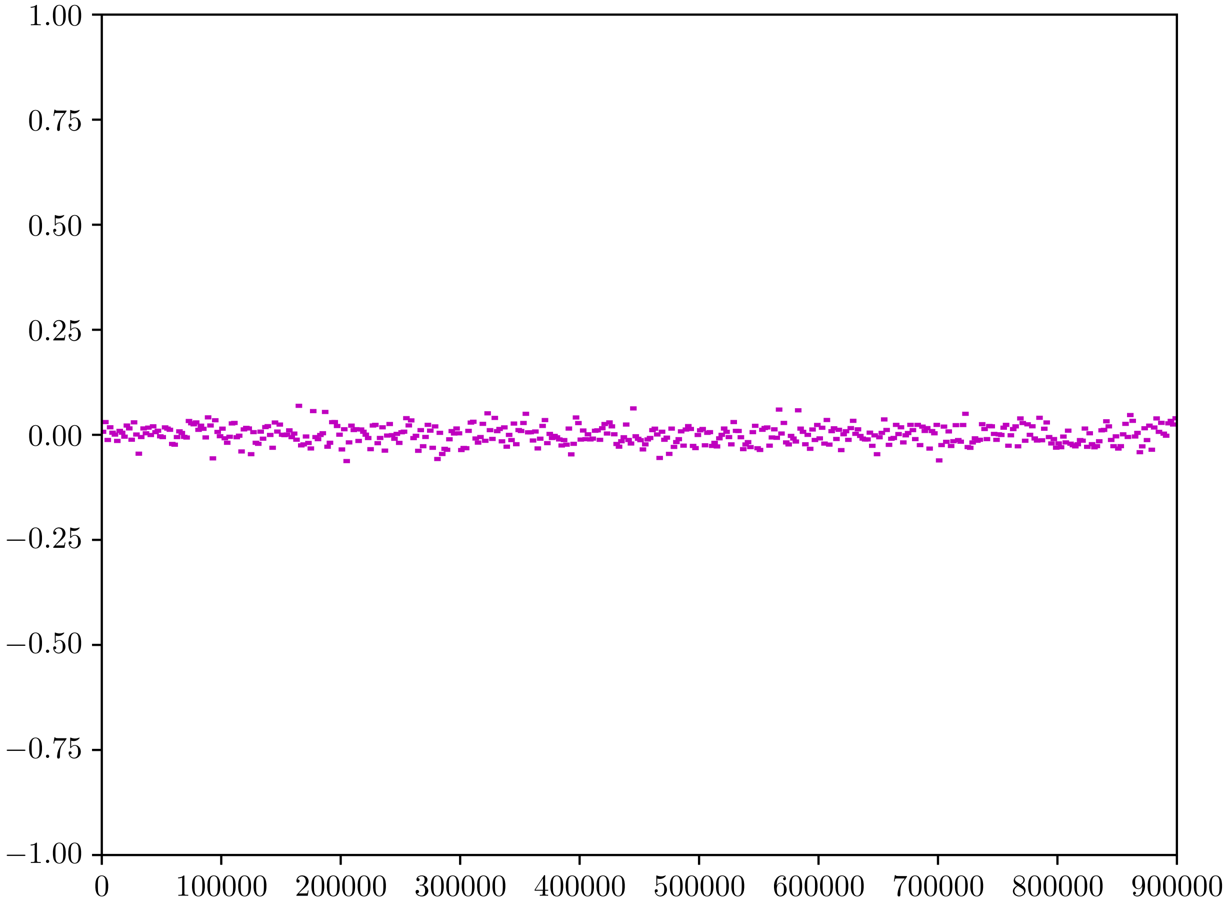

Figure 2: Record means: Each record is a consecutive subset of the noise dataset (Fig. 1) after calibration. Here, only the means of each such record are shown. These means μ_{record} fluctuate with a standard deviation σ_{record} of 0.022129, in theory $\sigma/\sqrt{n_{\text{record}}} = 0.022361$.

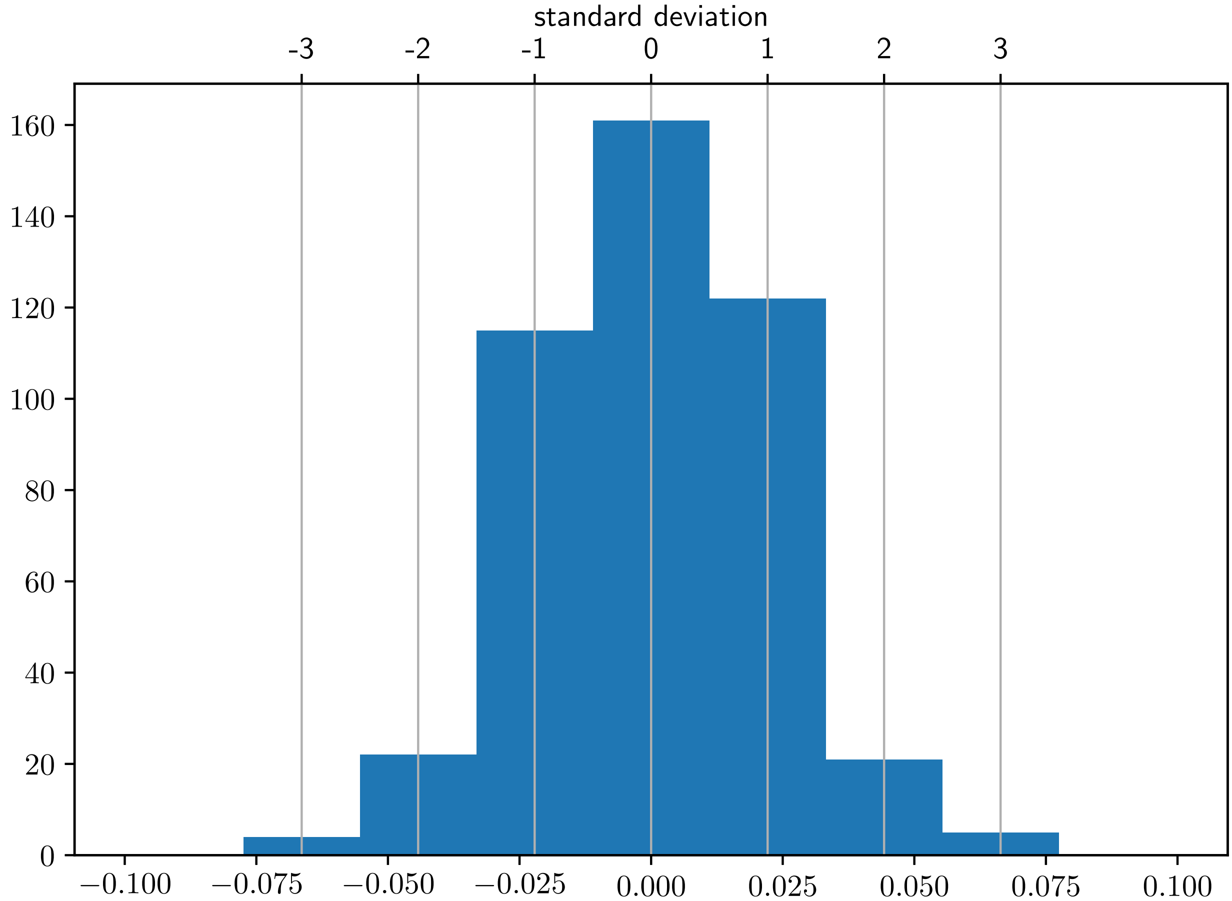

Figure 3: Histogram of record means: The record means (from Fig. 2) are binned. With increasing number of records, and a larger dataset from which the records are derived, the resultant histogram approaches a normal distribution with a mean μ_{record} of 0.0 due to calibration and a standard deviation σ_{record} of 0.022129. Recall, about 99.7% of all samples are within $\pm$ 3 standard deviations of the mean.

3 Conclusion

The mean of a large dataset approaches the ideal null-line within an error margin of $\sigma/\sqrt{n_{\mathrm{dataset}}}$. Not knowing the null-line any better, we calibrate seismic instruments accordingly. However, assuming that any data subsets show the same mean is a logical fallacy: a calibrated dataset does not imply likewise calibrated records. Instead, the means of such subsets fluctuate around the ideal null-line with a standard deviation of $\sigma/\sqrt{n_{\mathrm{record}}}$.