Maximum-Likelihood-Estimation Stacking

Stewart Trickett*
Kelman Technologies, Calgary, AB, Canada
stewart@kelman.com

Summary

Common-midpoint (CMP) stacking is often done using the arithmetic mean. There are good reasons for this: it's simple, linear, and is optimal if the noise has a Gaussian distribution. Where the noise is not Gaussian, however, the mean does a poor job, resulting in a stacked section contaminated by erratic noise. By estimating the probability distribution of the noise as it varies with time and CMP, and stacking using a maximum-likelihood estimator for that distribution, we get a result which is identical to a normal stack where the noise is Gaussian, but is far cleaner where the noise is erratic.

Introduction

One of the first examples of statistical estimation in seismic processing was stacking (Mayne, 1962), which summed or took the arithmetic mean of the sample values of the traces at each time point of a common-midpoint (CMP) gather. I shall refer to stacking with the arithmetic mean as "mean stacking". It is still commonly used today since it's simple, intuitive, linear, and optimal when the noise across each time point behaves like a Gaussian (that is, normal) random variable.

Still, mean stacking often produces poor results when the noise is erratic, such as is the case with shot noise, environmental noise at the geophone, and recording noise (spikes and clipping). Even background noise which is Gaussian in time may be non-Gaussian when viewed at a constant time within a gather when the amount of noise varies with trace. These problems are aggravated by the demands of AVO-friendly processing (Chopra, 2005), where trace-by-trace scaling is discouraged.

Statistical estimators that behave well in the presence of erratic noise are called "robust estimators". Some robust estimators that have been used in stacking are α-trimmed mean, median, Huber norm, and M-estimators (Elston, 1990). Despite their advantages, people tend to avoid these robust estimators. First, they are non-linear, so they generate high-frequency "jitter" even when the prestack traces are band limited. Second, they do not attenuate noise as well as mean stacking when the noise truly is Gaussian, as it often is.

Our goal is to find a stacking method that is near-optimum for the characteristics of the noise everywhere in the stack, behaving like the mean where the noise is Gaussian, and like a robust estimator where it is not. In other words, to have our cake and eat it too.

Method

The strategy is this: For each CMP gather, estimate the probability distribution of the noise as a function of time, and then determine the stacked trace by calculating the maximum-likelihood estimators given these distributions.

First, we need a family of distributions to draw from, preferably one that covers a wide variety of statistical predicaments. A favourite among statisticians for this purpose is Student's t-distribution (Lange, Little, and Taylor, 1989), a family of symmetric distributions described by parameters:

- μ Expected value, or location.
- σ Standard deviation, or scale.
- ν Degrees of freedom, or shape.

We are interested in parameter ν, the shape of the distribution, defined over the interval $[1,\infty]$. An infinite interval is awkward, so I introduce parameter $\lambda = 1/\sqrt{\nu}$ defined over the interval [0,1]. The Gaussian distribution is $\lambda = 0$ and the Cauchy distribution is $\lambda = 1$. The central value $\lambda = .5$ is a favourite among statisticians when they have little information to go on (Figure 1).

If we know λ at a particular CMP and time, we can estimate the stack value there by calculating the maximum-likelihood estimation (MLE) of the location μ of the sample values (Lange, Little, and Taylor, 1989). When the distribution is Gaussian $(\lambda = 0)$, the MLE is the arithmetic mean of the sample values. Otherwise we calculate the MLE using the Expectation-Maximization, or EM, algorithm (McLachlan and Krishnan, 1997).

The behaviour of MLE depends critically on the shape of the probability distribution that is assumed. Fat-tailed distributions (large λ values) give MLEs which are robust - that is, insensitive to wild or outlying values. Thin-tailed distributions (small λ values) give MLEs which are sensitive to outliers, and thus are ill-suited for handling erratic noise.

Thus we need a good estimate of λ as it varies with CMP and time. Ideally we want to maximize the likelihood of the data samples over μ, σ, and λ simultaneously. This can be done using the EM algorithm, but is expensive and unstable. Another approach is to measure the kurtosis (Joanes and Gill, 1998) of the sample values $x_i, i = 1,\ldots,n,$ about the location μ at each time point, defined by

$$n\sum_{i=1}^{n}(x_i-\mu)^4 \;\; / \;\; (\sum_{i=1}^{n}(x_i-\mu)^2)^2$$

Kurtosis is a useful indicator of distribution because thin-tailed distributions have low kurtosis and fat-tailed distributions have high kurtosis. Although the expected mean of the kurtosis is undefined for $\lambda \geq .5$, the expected *median* is defined for all λ, shown in Figure 2 for various stacking folds.

The final algorithm for MLE stacking of a trace gather is as follows:

1. Estimate the kurtosis of the noise at each time point in the gather.
2. Calculate λ at each time point based on the fold and the kurtosis.
3. Median filter the λ values with time (this is critical).
4. Stack the trace gather at each time point by calculating the maximum-likelihood estimation of the location μ using the value of λ at that time point.

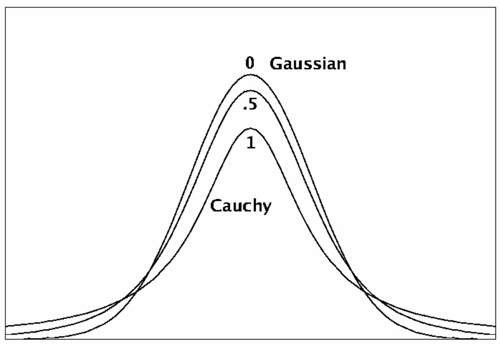

Figure 1

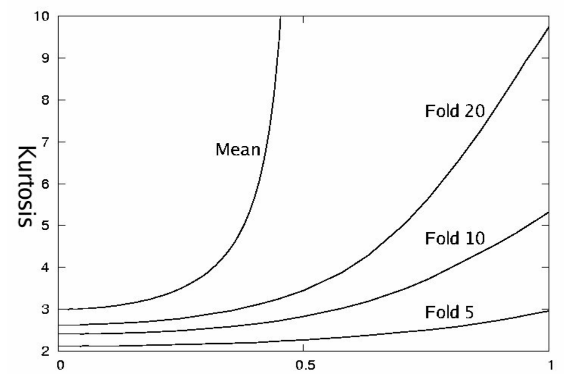

Figure 2

Figure 1. Student's t-distribution for λ = 0 (Gaussian), .5, and 1 (Cauchy). Note that the tails get fatter as λ increases.

Figure 2. Kurtosis as a function of λ. "Mean" is the expected mean of the Kurtosis. The rest are the expected median of the kurtosis for various folds

Data Examples & Future Work

Figure 3 shows a mean stack contaminated with air blast. The MLE stack has cleaned up almost all of this noise. The estimation of λ tracks the air blast well, even into parts of the section where the noise is nearly invisible to the naked eye in the mean stack.

Figure 4 shows a mean stack contaminated by erratic noise bursts. Once again MLE stacking has cleaned up the section well. Notice that MLE stacking has made little or no improvement for most of the background noise, which the λ estimate shows is near Gaussian.

A by-product of this method is a new attribute – namely λ, the estimated shape of the probability distribution of the noise as it varies with time and CMP. An intriguing question is whether λ is useful for other things. For example, might λ, or some similar statistic, indicate the presence of multiples? This will be the subject of future research.

Acknowledgements

Thanks to Compton Petroleum Corp. and Pulse Data Inc. for permission to show their data, and thanks to Nanna Eliuk and Neal Coleman at these companies for their kind assistance.

References

Chopra, S., *et al*, 2005, Expert Answers, CSEG Recorder **30** (2), 14-17.

Elston, S. F., 1990, Use of robust estimators in multichannel stacking: 60th Annual International Meeting, SEG, Expanded Abstracts , 1693-1696.

Joanes, D. N. & Gill, C. A.,1998, Comparing measures of sample skewness and kurtosis, Journal of the Royal Statistics Society, (Series D): The Statistician **47** (1), 183–189.

Lange, K. L., Little, R. J. A., and Taylor, J. M. G., 1989, Robust statistical modeling using the t distribution, J. Amer.Statist. Assoc., **84**, 881-896.

Mayne, W. H., 1962, Common reflection point horizontal data stacking techniques: Geophysics, **27** , no.6, 927-938.

McLachlan, G. & Krishnan, T., 1997, The EM Algorithm and Extensions. Wiley, New York, USA.

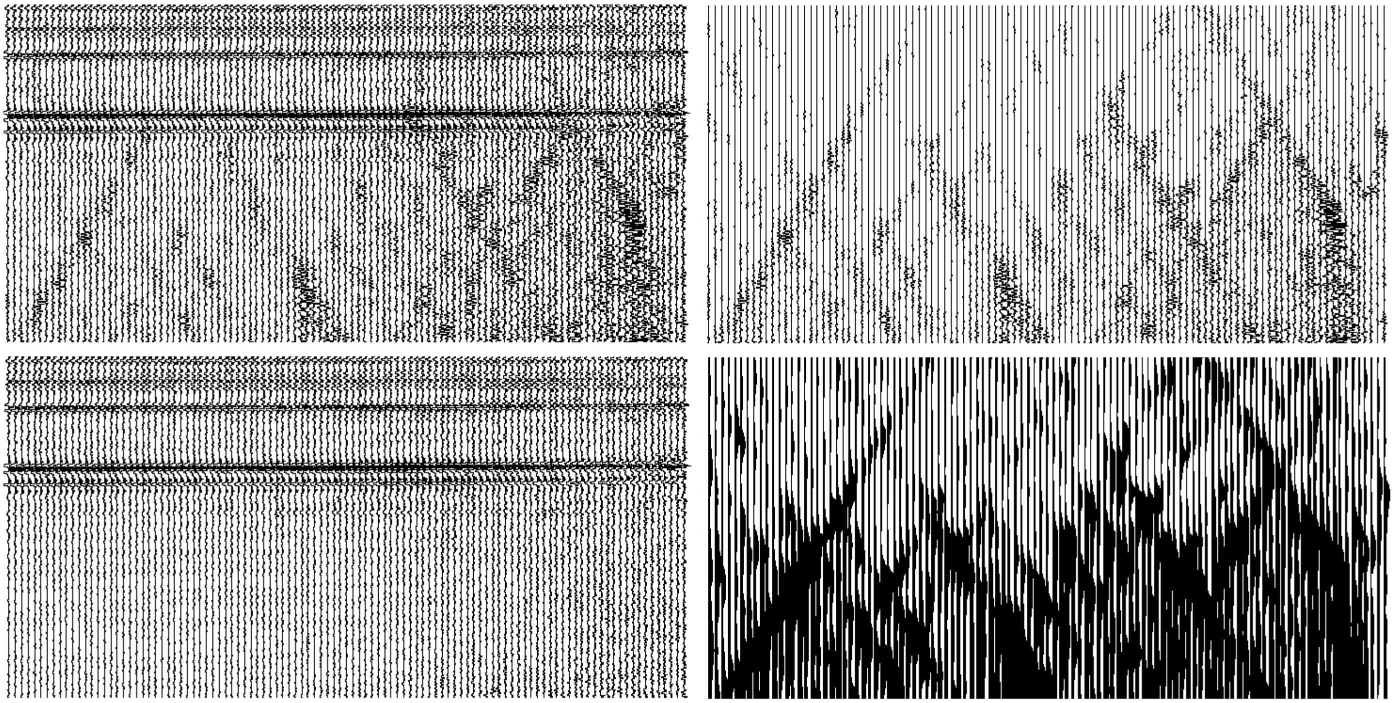

Figure 3. Mean stack contaminated with air blast (top left), MLE stack (bottom left), difference between the two (top right), and λ (white = 0 = Gaussian, black = 1 = Cauchy) (bottom right).

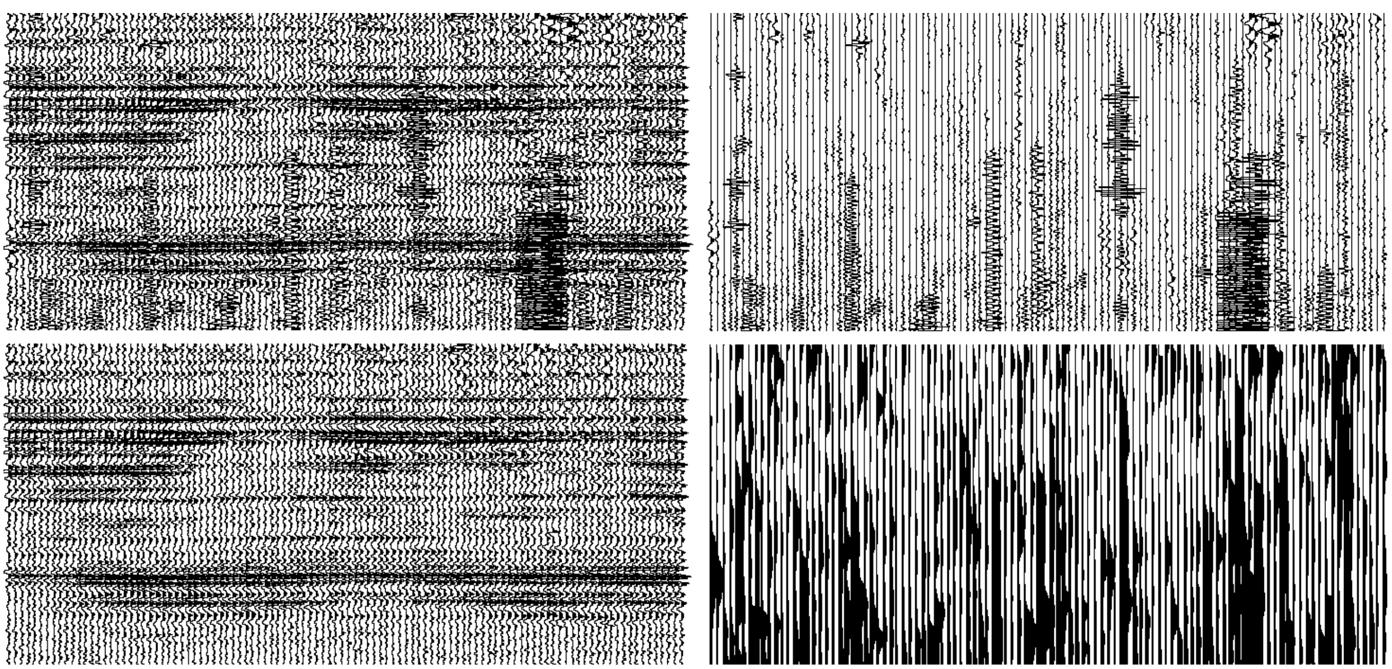

Figure 4. Mean stack contaminated with erratic noise (top left), MLE stack (bottom left), difference between the two (top right), and λ (white = 0 = Gaussian, black = 1 = Cauchy) (bottom right).