Institute and Faculty of Actuaries

Signal vs. Noise

Management Challenge Accepted

by Chris Smerald

The Better Sensing and Responding to Change Working Party

24 April 2017

Signal vs. Noise: Management Challenge Accepted

As claims actuary I can be asked to do just about anything. Like this:

> "We should define rules that help spot signal indicators vs noise indicators. Chris, Can we get a draft in short order together with a plan to finalize before end of q1?"

I was anxiously pleased. The deadline was short, but it was a great question born of data science and posed to an actuary.

I had to accept: I am proud of the heuristic modelling capabilities of actuarial science which in complex situations are superior to purely statistical approaches[i]; I am also responsible for claims diagnostics and structural driver analytics at work; and am chair of The Sensing and Responding to Change working party.

Sometimes close deadlines are your friend. They force you to ask for help from anyone nearby, and to look hungrily at things otherwise unappetising, like signal processing (the original signal vs. noise area). This led me on to quality control theory where the easier pickings were. The whole hunting experience, however, gave me an appreciation of how these two areas can help make actuarial science more robust. By the end of this article, I hope you will agree.

The person nearby was Heidi on my Sensing working party. She happened to have analysed signals vs noise in her PHD and told me a little about it. Our conversation gave me enough confidence to push further. Surely there would be some good analogies? I soon found a signal processing related paper by several actuaries and an electrical engineer: "A row-wise stacking of the runoff triangle: state space alternatives for IBNR reserve prediction"[ii]

This shows that triangles can be directly translated into the sine like waves dealt with in signal processing. The authors feel that this format makes applications from "linear state space statistics" more intuitive. I like being able to appreciate the data in a new way visually and am intrigued by what possibilities are afforded by a simple transformation. Though I did not have time to fully digest this paper, it seemed to be onto something.

From This: To This!

Table 1: AFG runoff triangle (thousands of dollars).

Accident year w	Development d 0	1	2	3	4	5	6	7	8	9
1	5012	3257	2638	898	1734	2642	1828	599	54	172
2	106	4179	1111	5270	3116	1817	-103	673	535	
3	3410	5582	4881	2268	2594	3479	649	603		
4	5655	5900	4211	5500	2159	2658	984			
5	1092	8473	6271	6333	3786	225				
6	1513	4932	5257	1233	2917					
7	557	3463	6926	1368						
8	1351	5596	6165							
9	3133	2262								
10	2063									

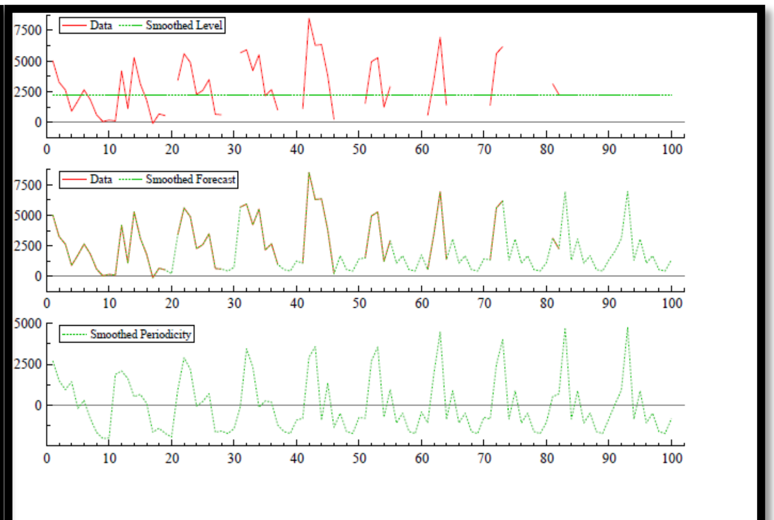


Figure 5: Kalman smoothing estimation results from model I-c for the original scale.

The red lines in the above right chart are incremental reported losses by accident year. The green are IBNR.

Looking more broadly and for something a little simpler, I happened upon a paper by a former and renewed colleague; Uri Korn : "An Extension to the Cape-Cod Method with Credibility Weighted Smoothing"[iii] His paper tells one how to squeeze more information out of accident year triangles using practical assumptions with Kalman filtering.

Usually we use the information from older accident years to inform current ones, but this iterative method cycles between using earlier years to help project later ones, then use later year's information to help project earlier ones, then repeat. There is even a handy excel spreadsheet illustrating the ideas. This did not exactly fit management's question, but led me to more discovery.

Searching Kalman filtering further, I came across eminent statistician Donald J. Wheeler in the field of quality control. From there I found two directly relevant articles: "What Makes the XmR Chart Work? How does it separate the signals from the noise?"[iv] and "Separating the Signals From the Noise"[v]. These are practical introductions to answering questions like my management's. In the meanwhile I ordered his book: *Advanced Topics in Statistical Process Control*, *Second Edition* © 2004 SPC Press and so far it is very appetising.

In his XmR chart article, Wheeler explains *"two basic ideas or principles that need to be respected when creating a chart for individual values and a moving range (an XmR Chart)."* I will leave digesting this to the reader, but it gives many practical tips in displaying and communicating change data.

In the second article Wheeler discusses how to best filter out noise by looking at incremental data against a backdrop of mean and variability measures –and the way these are calculated matters.

> *"while some data contain signals, all data contain noise, therefore, before you can detect the signals you will have to filter out the noise. This act of filtration is the essence of all data analysis techniques. … it is common to begin with the computation of some summary statistics for location and dispersion. … the structures within our data will often create alternate ways of computing these measures of dispersion. Understanding the roles of these different methods of computation is essential for anyone who wishes to analyze data."*

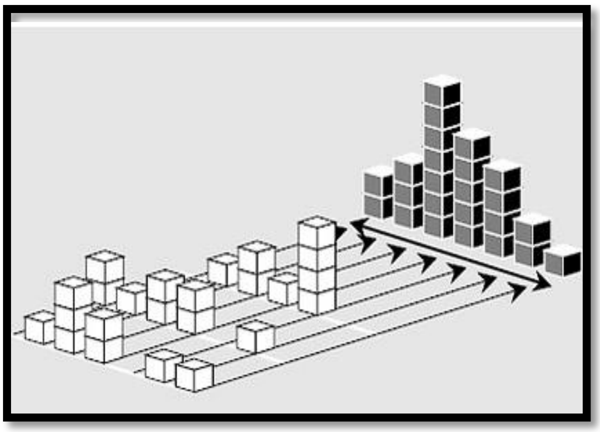
He goes on to show us an interesting use of range statistics to estimate standard deviation plus his calculations remind us that the sample standard deviation needs a "C4" correction factor to adjust for small sample bias.

He considers three methods of assessing the noise in ones data and concludes:

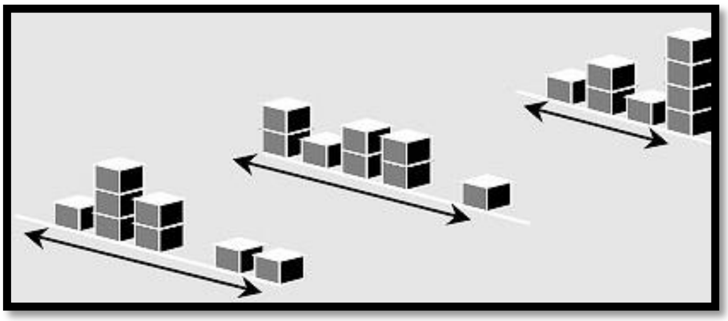
> *So when it comes to filtering out the noise you have a choice between method two, method two, or method two. Any method is right as long as it is method two!*

By noise he means standard deviation. He wants to calculate standard deviation in such a way that it is smallest. The way in which data are grouped and calculations performed affect the estimate of the standard deviation. Method two happens to minimize the standard deviation when the mean in changing.

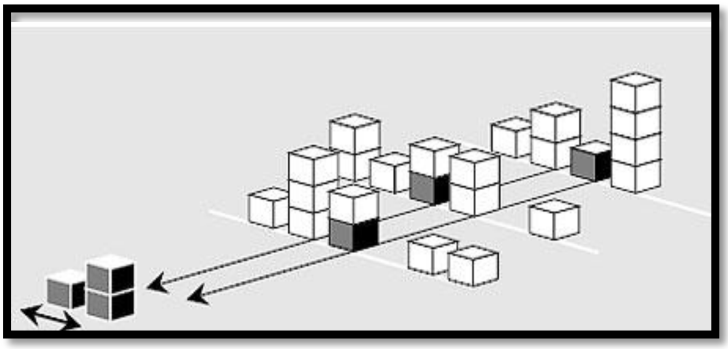
Let's say you have observations from different time periods. You could analyse them as a single group, then track how the mean and variability change as new time periods are added. This is method one. It's proneness to distortion via a changing mean is known as Quetelet's Fallacy.



Method 1: Group all the data together (black) then look at the mean + dispersion



Method 2: Estimate the mean + dispersion of each subgroup.



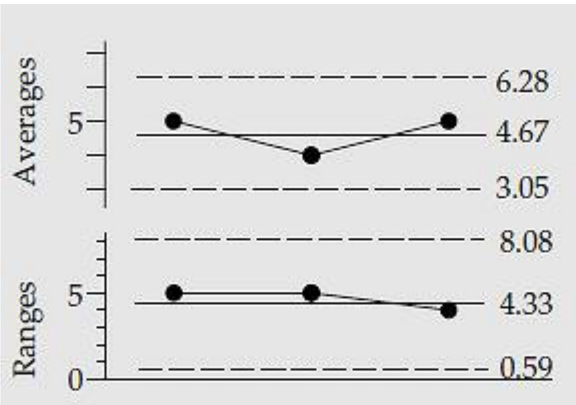
Method 3: Est. the mean of each subgroup and then the dispersion between means

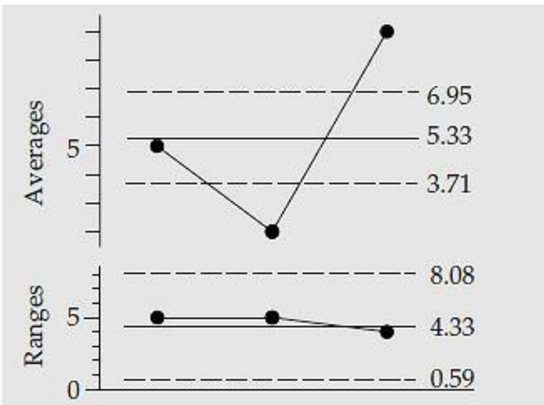
Or, in method two, method 2, method 2, (Within-Subgroup Variation) you could estimate the mean and standard deviation of each time period, then average these to estimate the overall mean and standard deviation.

For method three (Between-Subgroup Variation) you can calculate the average for each time period and then the standard deviation from the averages. This may respond more to signals in the data, than method two, but the estimate of the standard deviation may be poor.

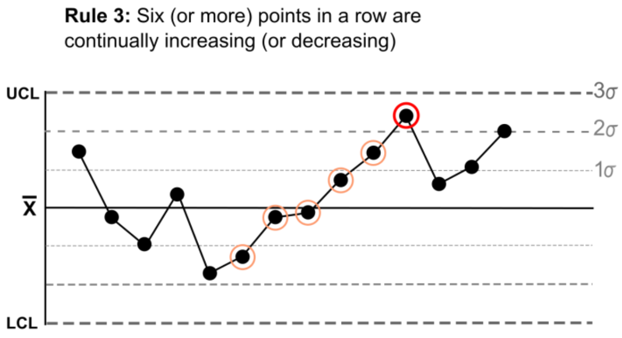
In times of changing mean, but stable standard deviation, method one moderately overstates the standard deviation, so the signal may be mistakenly taken for noise. Method two is more optimal and he feels is more likely to detect the signal. Method three will sometimes wildly overestimates the noise and is more likely to miss the signal than the other methods.

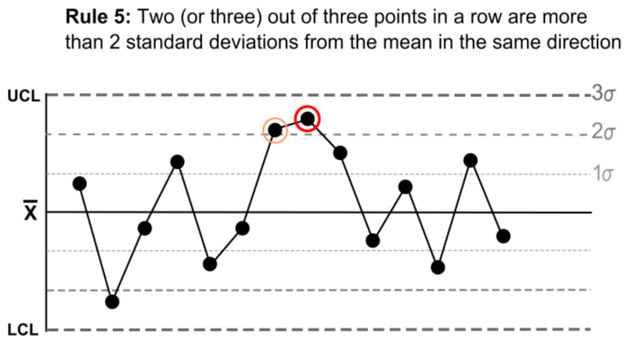
Data is displayed in Average and Range Charts, which look a good way to report possible change to management. Range is the difference of highest and lowest observations and Wheeler shows how they can be converted to standard deviation. The chart on the left shows the method two averages and ranges are stable, but correctly detects a change in the average while the range (comparable to standard deviation) stays stable.





My final stop in my moveable feast was to happen upon “Nelson’s Rules” https://en.m.wikipedia.org/wiki/Nelson_rules These are eight rules to help spot change in time series. Here are two examples, one for trend signals, another to detect extreme events.





These are very close to what I need. Practical tried and tested rules that can be easily applied to actual vs. expected statistics for paid, incurred, claim counts and etc., and may even suit actuarial ratios: average paid, claim disposal rates, on-level loss ratios and etc. I plan on applying these ideas to calendar year triangle statistics. The data needs some way of estimating its expectation, variance and covariance and away I can go. In particular I need to find out how to adjust for size and ratio effects: one accident year may be on a larger exposure base than other; loss development has more mass in the nose than in the tail; and expected values and variances of ratios may need special care in calculating. But these issues are readily solvable. We only lacked the motivation to develop them.

So what about the future? I will be coming back to Kalman filters and perhaps the sine wave analogy. I am also emboldened to take a survey of other fields where statistics are used to solve business problems. There is much to learn from our colleagues, including data scientists, and perhaps much they can learn from us.

Please feel free to contact me with comments or suggestions at chris.smerald@btinternet.com.

---

[i] Chris Smerald: Heuristics, Dammed Statistics and Complexities, LMAG presentation March 2016
[ii] Rodrigo Atherino, Adrian Pizzingay and Cristiano Fernandesz, Pontical Catholic University of Rio de Janeiro (PUC-Rio) Rua Marques de Sao Vicente, 225 – Gavea 22451-900 Rio de Janeiro, RJ - BRAZIL
[iii] The CAS E-Forum, Summer 2016, http://www.casact.org/pubs/forum/16sforum/
[iv] Donald J. Wheeler, Quality Digest Daily, December 3, 2012 https://www.spcpress.com/pdf/DJW250.pdf
[v] Donald J. Wheeler, Quality Digest Daily, October 3, 2013 https://www.spcpress.com/pdf/DJW260.pdf

Institute
and Faculty
of Actuaries

London

7th Floor · Holborn Gate · 326-330 High Holborn · London · WC1V 7PP
Tel: +44 (0) 20 7632 2100 · Fax: +44 (0) 20 7632 2111

Edinburgh

Level 2 · Exchange Crescent · 7 Conference Square · Edinburgh · EH3 8RA
Tel: +44 (0) 131 240 1300 · Fax: +44 (0) 131 240 1313

Oxford

1st Floor · Park Central · 40/41 Park End Street · Oxford · OX1 1JD
Tel: +44 (0) 1865 268 200 · Fax: +44 (0) 1865 268 211

Beijing

6/F · Tower 2 · Prosper Centre · 5 Guanghua Road · Chaoyang District · Beijing China 100020
Tel: +86 (10) 8573 1000 · Fax: +86 (10) 8573 1100

Hong Kong

2202 Tower Two · Lippo Centre · 89 Queensway · Hong Kong
Tel: +11 (0) 852 2147 9418 · Fax: +11 (0) 852 2147 2497

www.actuaries.org.uk