2014 1st International Congress on Computer, Electronics, Electrical, and Communication Engineering (ICCEECE2014)
IPCSIT vol. 59 (2014) © (2014) IACSIT Press, Singapore
DOI: 10.7763/IPCSIT.2014.V59.34

C (α) Method to Check Daunting Over/Under Variances to Understand Times to Aftershocks since a Major Earthquake

Ramalingam Shanmugam

School of Health Administration, Texas State University, San Marcos, USA

Abstract. Model is an *abstraction* of the reality. The first step in data analysis is to capture its most suitable model. This step receives a fatal hit when the data exhibits either *over or under variance* against the model. This article provides an approach, based on *Neyman's C (α) principle*, to decide whether the *over or under variance* is weak enough for the model to work. To illustrate the method, the times (in hours) between main earthquake and aftershocks during 1973-1995 are considered and explained.

Keywords: exponential model, continuous distribution, p-value, power, ratio statistic

1. Introduction

An earthquake and its follow-up aftershocks are too dreadful to be ignored. Especially, when they occur underwater, they trigger tsunami which cause additional devastations. An example was the 9.2 Mw earthquake in the Indian Ocean below 30 km of sea level at local time 00:58:53, 26th December 2004. It caused tsunami which killed approximately 280,000 humans, thousands of uncounted animals, and damaged a vast amount of properties in Indonesia, Thailand, Maldives, Somalia, Sri Lanka, India etc. Both the immediate and long-range health damages are nightmare to governments and service agencies as the aftershocks are known to follow any major earthquake. Do the times (in hours) to aftershocks follow any pattern? Do the times follow an exponential probability pattern? If so, both prediction and planning to protect the public from further devastations. This theme is discussed below.

This article has developed a methodology in Section 2, based on Neyman's C (α) principle, to decide whether existing over or under variance in the data is negligible for the exponential model to work. More variance is translated to portray higher volatility. For illustration, the times (in hours) between main earthquake and aftershocks during 1973-1995 are considered and explained in Section 3. Few conclusive comments are made in Section 4

2. Derivation of a methodology

2.1. Why a modification of exponential model is necessary?

An under-water earthquake or volcano is not a daily event but occurs occasionally. Hence, earthquake is a Poisson type event. Since a main earthquake or volcano, do the times (in hours) to aftershock(s), follow an exponential probability pattern? To be specific, let Y be the random variable denoting the time to an aftershock. A reparametrized logarithmic version of the exponential probability density function (pdf) for Y is

$$\ln f(y|\phi,\upsilon) = \frac{1}{2}[\ln(1+\phi) - \ln(1-\phi) - \ln\upsilon] - [\frac{(1+\phi)}{(1-\phi)\upsilon}]^{1/2} y; y \geq 0, \upsilon > 0; -1 \leq \phi \leq 1 \qquad (1.a)$$

where the *imbalance measure*, $\phi = \frac{\upsilon - g(\mu)}{\upsilon + g(\mu)}$ reflects a tilt between the variance, $\upsilon > 0$ and the required functional equivalence, $g(\mu) = \mu^2$ of the mean, μ according to the model (1.a). In other words, when $\phi = 0$, there is no tilt (that is, $\upsilon = \mu^2$) from the well-known exponential distribution (1.b) as it is known to have its variance equal to the square of the mean. That is

$$\ln f(y|\phi=0,\upsilon) = -\ln\mu - [\frac{y}{\mu}]; y \geq 0, \mu > 0 . \quad (1.b)$$

When $\phi < 0$, there is a tilt indicating an *under variance* (that is, $\upsilon < \mu^2$). When $\phi > 0$, there is a tilt reflecting an *over variance* (that is, $\upsilon > \mu^2$). The mean and variance of the pdf (1.a) are $E(y|\phi,\upsilon) = \sqrt{\frac{(1-\phi)\upsilon}{(1+\phi)}}$ and $Var(y|\phi,\upsilon) = [E(y|\phi,\upsilon)]^2 = \frac{(1-\phi)\upsilon}{(1+\phi)}$ respectively. The estimate of the imbalance measure is $\hat{\phi} = \frac{(s^2 - \bar{y}^2)}{(s^2 + \bar{y}^2)}$. The memory is $\Pr[Y > y + m|Y > y] - \Pr[Y > y]$ and it is zero for both the models (1.a) and (1.b) meaning that the *imbalance measure,* ϕ does not alter the memory pattern. However, is the model (1.a) or (1.b) appropriate for the data?

2.2. A methodology to test hypothesis on imbalance

The answer requires testing the *null hypothesis* $H_o : \phi = 0$ against an *alternative hypothesis* $H_1 : \phi \neq 0$. For this, we select the *Neyman's C (* α *)* principle as it is powerful. What is *Neyman's C (* α *)?* See Shanmugam (1992) for details. Following the principle, the statistic $T = \frac{\bar{y}}{\sqrt{s^2}}$ is considered. Using the formulas $E(\frac{U}{W}) \approx \frac{E(U)}{E(W)}\{1 + \frac{Var(W)}{[E(W)]^2} - \frac{Cov(U,W)}{E(U)E(W)}\}$, $E(h[W]) \approx h[E(W)]$, $Var(h[W]) \approx (\partial_w h[W])^2 Var(W)$ and $Var(\frac{U}{W}) \approx [\frac{E(U)}{E(W)}]^2\{\frac{Var(U)}{[E(U)]^2} + \frac{Var(W)}{[E(W)]^2} - 2\frac{Cov(U,W)}{E(U)E(W)}\}$ from Stuart and Ord (2009) with $U = \bar{y}\sqrt{n}$ and $W = \sqrt{s^2}$, we find after algebraic simplifications that $E(T|\phi,\upsilon) \approx \sqrt{\frac{n(1+\phi)}{(1-\phi)}}$ and $Var(T|\phi,\upsilon) \approx (\frac{1+\phi}{1-\phi})$. Hence, the standardized statistic Z follows the standard normal distribution. It means that the p-value of rejecting the null hypothesis $H_o : \phi = 0$ in favor of the research hypothesis $H_1 : \phi \neq 0$ is

$$p = 2\Pr[Z \geq \left|(\frac{\bar{y}}{\sqrt{s^2}} - \sqrt{n})\right|] . \quad (2)$$

When the null hypothesis is rejected at a α level and the true value of imbalance measure is known to be ϕ_1, the probability of accepting the true value ϕ_1 is called the *statistical power* and it is

$$power = \Pr[-\frac{z_{\alpha/2}\bar{y}}{|\bar{y} - \sqrt{ns^2}|}\sqrt{\frac{1-\phi_1}{1+\phi_1}} - \sqrt{n} \leq Z \leq \frac{z_{\alpha/2}\bar{y}}{|\bar{y} - \sqrt{ns^2}|}\sqrt{\frac{1-\phi_1}{1+\phi_1}} - \sqrt{n}] \quad (3)$$

Decades	**Tectonic Plates**	**n = # earth quakes**	**aftershocks magnitude (mg)**		**Hours to largest aftershock with power for $\phi_1 = 0.5$**				
			$\bar{y}$	s^2	$\bar{y}$	s^2	$\hat{\phi}$	p-value	power
1970	Inter	2	4.12	0.32	10.13	204	0.33	0.481	0.945
	intra	7	4.21	0.56	1.88	10.31	0.32	0.309	0.946
1980	Inter	3	5.57	0.56	8.37	137	0.44	0.169	0.995
	intra	10	5.33	0.70	11.43	322.50	0.49	0.039	0.999
1990	Inter	4	6.22	0.09	1.07	2.92	0.54	0.022	0.999
	intra	8	5.61	0.57	4.96	82.47	0.93	0.012	0.992

Table 1. Time (in hours) between earthquake and its major aftershocks during 1973-1995 as reported in Guo and Ogata (1997), Journal of Geophysical Research, 102, 2857-2873.

2.3. Illustration with aftershock times of earthquakes

In section, we examine whether the times to largest aftershock since a major earthquake follow an exponential probability pattern. The data in Table 1 are from Guo and Ogata (1997). The sample average, $\bar{y}$ and variance, s^2 are displayed in the table along with the number of earthquakes. Using (2) and (3), the p-value for the null hypothesis $H_o : \phi = 0$ to be true and the power of accepting the true value $\phi_1 = 0.5$ are calculated and displayed in Table 1. The p-values suggest that the imbalance measures are not negligible during the earthquakes on the intra tectonic plates of 1980 and both inter and intra tectonic plates of 1990. The power of accepting the true value $\phi_1 = 0.5$ is displayed for all the earthquakes on inter and intra tectonic plates during 1970, 1980 and 1990. The powers are excellent implying that the methodology is powerful.

In conclusion, the times to the largest aftershock, since a major earthquake, do follow an exponential probability pattern with memory less property. But, there is a significant imbalance at times in the functional relationship between the sample variance and mean as required by the exponential probability model. This kind of imbalance probably occurs not only in the earthquake data but also in other disciplines such as engineering, business, and science, marketing and communication networks. Does this imbalance signify a clue for the repeat of the incidence? What are the causing factors of this imbalance? These questions are worthy to pursue and it would be undertaken soon in the future.

3. Acknowledgements

The author appreciates and thanks the travel support by the Texas State University to present this research work in the 2014 International Conference on Computer Science and Engineering, Chennai, India.

4. References

[1] Guo, Z. and Ogata, Y. (1997). Statistical relations between the parameters of aftershocks in time, space, and magnitude, Journal of Geophysical Research, 102, 2857-2873

[2] Shanmugam, R. (1992). An inferential procedure for the Poisson intervention parameter, *Biometrics,* 48, 559-565.

[3] Stuart, A. and Ord, K. (2009). Kendall's Advanced Theory of Statistics. 6th Edn., Wiley, London, U.K., ISBN-10: 0340614307, pp: 700.

Dr. Ram Shanmugam is currently a professor in the School of Health Administration, Texas State University, San Marcos, TX 78666. He received his Ph. D. degree in applied statistics from the Fox Business School, Temple University. He has published more than one hundred refereed research articles in international journals, serves the healthcare related professions as a reviewer of grant proposals. He is an *elected fellow* of the International Statistical Institute. His research interests include *management consulting, informatics, clinical trials,* and *data mining methodologies.*