Ontology of Earthquake Probability: Metaphor

Philip B. Stark

Department of Statistics
University of California, Berkeley

Dynamics of Seismicity, Earthquake Clustering and Patterns in Fault Networks
2013 Program on Mathematics of Planet Earth
SAMSI
Research Triangle Park, NC
9–11 October 2013

What are earthquake probabilities?

Where does earthquake probability come from? What does "the chance of an earthquake" mean? Stochastic models for seismicity have a peculiar ontological status that makes it difficult to interpret earthquake probabilities. The compelling appeal of probability models for seismicity seems to derive in part from confusion in the literature about the crucial distinction between empirical rates and probabilities. The underlying physics of earthquakes is hardly understood, but does not appear to be intrinsically stochastic, merely unpredictable. Earthquake probabilities reflect assumptions and metaphors rather than knowledge: They amount to saying that earthquakes occur as if according to a casino game–a thesis for which there is little evidence. Many stochastic models have been invented to produce features similar to features of real seismicity. Those models contradict each other; none is a great match to what is believed about the underlying physics; none seems to hold up statistically when there enough data for a reasonably powerful hypothesis test; and none has been demonstrated to predict better than a very simplistic "automatic alarm" strategy. I contend that probabilistic models for earthquakes lack adequate scientific basis to justify using them for high-consequence policy decisions, that such models obfuscate and confuse more than they illuminate and edify, and that for the purpose of protecting the public they should be abandoned in favor of common sense.

Earthquake Probability as Metaphor

Shall I compare thee to a game of cards?

Thou art less predictable and yet not random.

–Wm. ShakesEarth

Earthquake Poker

Earthquake probabilities are based on a metaphor

- Earthquakes occur "as if" in a casino game whose rules are embodied in some mathematical model known to the seismologist.
- Like saying that there is a deck of seismology cards.
  - The deck contains some blank cards and some numbered cards.
  - In a given region, in every time interval, a card is dealt from the deck.
  - If the card is blank, there is no earthquake.
  - If the card has a number on it, an event with that magnitude occurs.

What's the game?

- Different models make different assumptions about how many cards of each type there are, the shuffling, whether drawn cards are returned to the deck, etc.
- One extreme: earthquake cards are distributed fairly evenly (the characteristic earthquake model).
- Another extreme: cards are thoroughly shuffled, and after each draw the card is replaced and the deck is re-shuffled (tantamount to the Poisson model).
- In between: deck is shuffled less than thoroughly (e.g., high cards tend to be followed by low cards—aftershocks), cards are not replaced (modeling stress accumulation or stress release), deck not re-shuffled between draws.

Also “Earthquake Urns”

Stein & Stein, 2013. Shallow Versus Deep Uncertainties in Natural Hazard Assessments, EOS http://onlinelibrary.wiley.com/doi/10.1002/2013EO140001/abstract

Gushing commentary by Mohi Kumar: http://blogs.agu.org/sciencecommunication/2013/04/02/simple-math-gives-readers-x-ray-vision/

“ a few weeks ago, a gem came across my desk. I barely needed to touch it, and after reading it I experienced a stillness of the mind. You know those magicians black boxes we build in our heads, where complicated stuff goes in, hands are waved, and—poof!—useable information comes out? I knew that one of those floating in my mind was just rendered transparent. . . . ‘Imagine an urn containing balls . . . in which e balls are labeled “E” for event and n balls are labeled “N” for no event,’ the authors write. ‘The probability of an event is that of drawing an E ball, which is the ratio of the number of E balls to the total number of balls.’ ”

I will explain why this is deep confusion, not shallow uncertainty.

Gambling and Terror

- *Why* should the occurrence of earthquakes be like a card game (or like drawing marbles)?
- It's only a metaphor.
- Why not like terrorist bombings?
- We might know that a terrorist plans to detonate a bomb, but not where or when or how big
- Ignorance of place, time, and magnitude of the threat does not make them random.

Weather prediction; Signal-to-noise

- Common to say earthquake prediction is like weather prediction. It isn't.
- Earthquakes account for only a small fraction of the energy budget of plate tectonics.
- Trying to predict large earthquakes is not like trying to predict rain.
- More like like trying to predict where the lightning will be in a storm: Tiny part of the energy budget of weather.
- Lightning more common some places than others, but trying to predict precisely where and when it will strike is impossible.
- Lightning might be frequent enough in some places that one could test a stochastic model for lightning statistically, with a test that had good power.
- Not true for stochastic models of large, local events: Too rare for meaningful statistical tests

Earthquake probabilities

Why do we think earthquakes have probabilities?

- Standard argument:
  $M = 8$ events happen about once a century.
  Therefore, the chance is about 1% per year.
- But rates are not probabilities.
- Probabilities imply rates in repeated random trials; can estimate the probability from the rate.
- Rates need not be the result of anything random.
- Having an empirical rate doesn't make something random.

Mortality Tables

'What are my chances, doc?'

The US Social Security Actuarial Life Table says that 6,837 of 1,000,000 men my age are expected to die in the next year. It reports that as a one-year "death probability" of 0.006837.

Is that the probability that I (or any other particular man my age) dies in the next year?

Thought experiment 1

Chance of death

You are in a group of 100 people. Two people in the group will die in a year. What's the chance you will die in the next year?

Thought experiment 2

Chance of name

You are in a group of 100 people. Two people in the group are named "Philip." What's the chance your name is "Philip?"

What's the difference?

Ignorance does not create chance

- For some reason, the first scenario invites answering with a probability but the second does not.
- If the mechanism for deciding which two people would die were to pick two at random and shoot them, then the chance would indeed be 2%.
- But the mechanism were to shoot the two tallest people, there's no "chance" about it: You are one of the two tallest, or not.

Rates are not (necessarily) related to chances

- Every list of numbers has a mean, but not every list of numbers is random.
- About 1 in 8 people in the US lives in California. Is the probability you live in California $\sim 12\%$?

  No: nothing random.

  But, chance a person selected at random from the US population lives in California is $\sim 12\%$.

  Chance comes from selection mechanism, not from rate.
- About 60% of births are in Asia.
  You are about to have a baby.
  Is the chance your baby will be Asian about 60%?

Confusing measured and simulated rates: Musson (2012)

In seismology, a similar solution could be applied to answering the question "What is the probability that tomorrow there will be an earthquake larger than 6 M_w somewhere in the world?" It would be sufficient to collate the data for the past 1,000 days and observe on how many days an earthquake above 6 Mw was recorded.

To answer the question "What is the annual probability of 0.2 g PGA at my site?" is more intractable, as the data are insufficient, and there may actually be no past observations of the target condition. The test of a statement about earthquake ground motion would be to make observations over a long enough number of years (say, 100,000) and count the number of times an acceleration over 0.2 g is recorded.

Musson (2012) contd.

The model is essentially a conceptualization of the seismic process expressed in numerical form, describing 1) where earthquakes occur, 2) how often they occur, both in terms of inter-event time and magnitude-frequency, and 3) what effects they have. With these three elements, one describes everything that determines the statistical properties of the seismic effects that will occur at a given site in the future. This is, therefore, all that one needs to simulate that future. One does not know precisely what will happen in the future, one only knows the aggregate properties of the seismicity and the ground motion propagation. Therefore simulations need to be stochastic. For each simulation, the probability density functions for earthquake occurrence in the model are randomly sampled to produce one possible outcome compatible with the model.

Musson (2012) contd.

One 50-year catalog, with ground motions assessed at site for each event, represents one possible outcome of the seismicity around the site in the next 50 years that is compatible with 1) what is known about the properties of the regional seismicity and 2) what is known about the relationship between ground motion, magnitude, and distance. Obviously, the content of this single catalog owes much to chance, and reality may be quite different. But when one repeats the process a very large number of times, say 200,000 times, the result is 10,000,000 years' worth of pseudo-observational data, from which computing the probability of any result is as simple as counting. In fact, this is as close to a purely frequentist approach to probabilistic hazard as one can get, as the simulated observations form a collective in the manner of von Mises (1957).

Musson critique

- explicitly claims that simulation is tantamount to a frequency theory measurement.
- false: all simulation does is approximate a distribution that is built into the assumptions.
- does not measure anything, just substitutes floating-point arithmetic for calculations that might be difficult to perform in closed form
- amounts to Monte Carlo integration of the assumed density.
- the density itself has not been measured, nor even established to exist. It's an input.
- like claiming you can tell whether a coin is fair by guessing its chance of heads, and having a computer simulate tosses of an ideal coin with that chance—without tossing the actual coin or measuring its chance of heads.

Rabbits

The Rabbit Axioms

1. For the number of rabbits in a closed system to increase, the system must contain at least two rabbits.
2. No negative rabbits.

Rabbits contd.

Freedman's Rabbit-Hat Theorem

You cannot pull a rabbit from a hat unless at least one rabbit has previously been placed in the hat.

Corollary

You cannot "borrow" a rabbit from an empty hat, even with a binding promise to return the rabbit later.

Applications of the Rabbit-Hat Theorem

- Can't turn a rate into a probability without assuming the phenomenon is random in the first place.
- Cannot conclude that a process is random without making assumptions that amount to assuming that the process is random. (Something has to put the randomness rabbit into the hat.)
- Testing whether the process appears to be random using the *assumption* that it is random cannot prove that it is random. (You can't borrow a rabbit from an empty hat.)
- Can't conclude a process is stationary and random without assumptions strong enough to imply the process is random and stationary. Observing the process isn't enough.

Rabbits and Earthquake Casinos

What would make the casino metaphor apt?

1. the physics of earthquakes might be stochastic
2. stochastic models might provide a compact, accurate description of earthquake phenomenology
3. stochastic models might be useful for predicting future seismicity

- Unless you believe (1), Rabbit Theorem says you can't conclude process is random.
- Might still be useful to treat it as random for reason (2) or (3).
- We will look at (2) and (3) for several common models.

Seismicity models

Common Stochastic Models for Seismicity

- Poisson. Clearly doesn't fit: too little clustering
- Poisson for "declustered" catalogs.
  Will pass test if you remove enough events, but standard algorithms don't (Luen & Stark)
- Gamma renewal. Doesn't fit (Luen, 2012)
- ETAS. Doesn't fit (Luen, 2012)

Are reclustered catalogs approximately Poisson?

Examine several declustering methods on SCEC data; test for temporal and spatiotemporal Poisson behavior.

Years 1932–	Mag (events)	Meth	n	MC		CC	KS	Romano	Reject?	
				χ^2	Sim			P	Time	Space-time
1971	3.8 (1,556)	GKl	437	0.087	0.089	0.069	0.011	0.005	Yes	Yes
		GKlb	424	0.636	0.656	0.064	0.006	0.000	Yes	Yes
		GKm	544	0	0	0	0.021	0.069	Yes	No
		Rl	985	0	0	0	0.003	0	Yes	Yes
		dT	608	0.351	0.353	0.482	0.054	0.001	No	Yes
	4.0 (1,047)	GKl	296	0.809	0.824	0.304	0.562	0.348	No	No
		GKlb	286	0.903	0.927	0.364	0.470	0.452	No	No
		GKm	369	<0.001	<0.001	0	0.540	0.504	Yes	No
		Rl	659	0	0	0	0	0.001	Yes	Yes
		dT	417	0.138	0.134	0.248	0.051	0	No	Yes
2010	3.8 (3,368)	GKl	913	0.815	0.817	0.080	0.011	0.214	Yes	No
		GKlb	892	0.855	0.855	0.141	0.005	0.256	Yes	No
		GKm	1120	0	0	0	0.032	0.006	Yes	Yes
		Rl	2046	0	0	0	0	0	Yes	Yes
		dT	1615	0.999	1.000	0.463	0.439	0	No	Yes
	4.0 (2,169)	GKl	606	0.419	0.421	0.347	0.138	0.247	No	No
		GKlb	592	0.758	0.768	0.442	0.137	0.251	No	No
		GKm	739	0	0	0	0.252	0.023	Yes	Yes
		Rl	1333	0	0	0	0	0	Yes	Yes
		dT	1049	0.995	0.999	0.463	0.340	0.001	No	Yes

P-values for tests for Poisson behavior of declustered SCEC catalog. χ^2: multinomial chi-square test using χ^2 approximation. Sim: multinomial chi-square test conditional on n. CC: conditional chi-square test. Sim, CC estimated w/ 10^5 simulated catalogs ($\pm \sim 0.16\%$). KS: Kolmogorov-Smirnov test that event times are iid uniform given n. Romano: permutation test for conditional exchangeability of times given locations. Time reject if simulation P for any of 4 temporal tests is < 0.0125. Space-time reject if P for Romano test < 0.05. From Luen & Stark, 2012.

Poisson doesn't fit, even after declustering using standard approaches.

Do Gamma renewal or ETAS fit?

B. Luen, 2010. PhD Dissertation, UC Berkeley

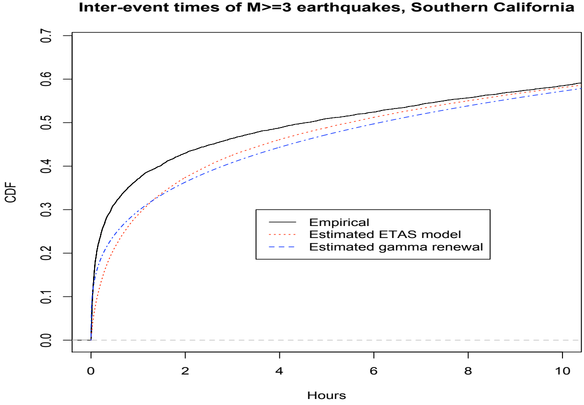

Figure 4.1: Cumulative distribution functions of inter-events times attached. The empirical inter-event distribution (SCEC catalog of Southern Californian $M \geq 3$ earthquakes, 1984-2004, $n = 6958$) is significantly different from both the fitted ETAS and gamma renewal models (in both cases, the P-value is less than 0.00001 for a test using the Kolmogorov-Smirnov test statistic). Empirically, there are more inter-event times under 2 hours than either fitted model would suggest. Beyond 12 hours, the difference in empirical distributions is small (not pictured).

Prediction

Recap

- No physical basis for any of the stochastic models
  Rabbits all the way down
- Poisson doesn't fit raw or declustered catalogs
- Gamma renewal and ETAS don't fit raw catalog
- Poisson obviously useless for prediction
- Does ETAS help for prediction?

Automatic Alarms and MDAs

- Automatic alarm: after every event with $M > \mu$, start an alarm of duration τ

  No free parameters.
- Magnitude-dependent automatic alarm (MDA): after every event with $M > \mu$, start an alarm of duration τu^M

  1 free parameter (u)

For both, adjust fraction of time covered by alarms through τ.

- Optimal ETAS predictor: level set of conditional intensity.

ETAS has 4 free parameters: K, α, c, p.

ETAS v Auto (Luen, 2010. PhD Dissertation, Berkeley)

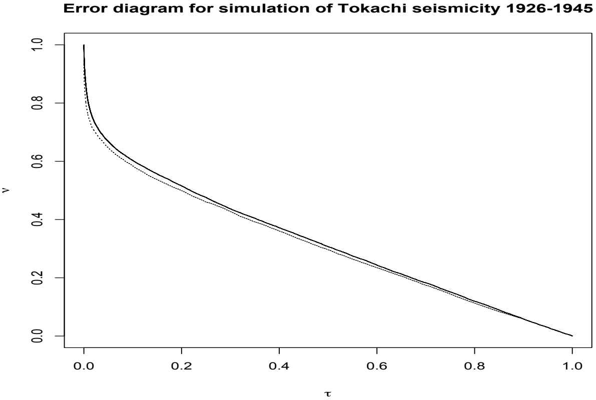

Figure 5.6: Error diagram for a simple automatic alarm strategy (solid line) and conditional intensity predictor (dotted line) for a 200,000 (with 10,000 day burn-in) day simulation of Tokachi seismicity based on parameters estimated by Ogata [1] from the catalog from 1926-1945. The simulation parameters were $m_0 = 5, m_1 = 9, b = 1, \mu = 0.047, K = 0.013, c = 0.065, \alpha = 0.83, p = 1.32$. On the x-axis, τ gives the fraction of time covered by alarms; on the y-axis, ν gives the fraction of earthquakes of magnitude 5 or greater not predicted. The 10th percentile of interarrival times is 40 minutes, the median is 4.3 days, and the 90th percentile is 34 days.

ETAS v MDA: Simulations (Luen, 2010)

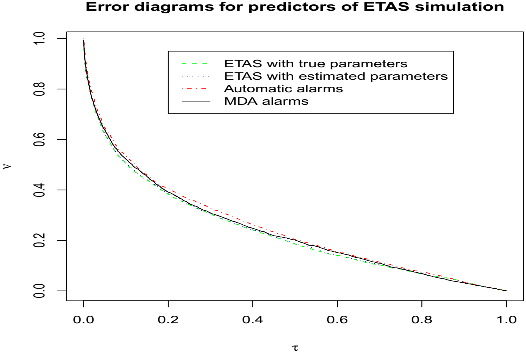

Figure 5.7: Error diagrams for predictors of a simulated temporal ETAS sequence. The parameters used in the simulation were those estimated for Southern Californian seismicity: $m_0 = 3, \mu = 0.1687, K = 0.04225, \alpha = 0.4491, c = 0.1922, p = 1.222$. Models were fitted to a 20-year training set and assessed on a 10-year test set. The ETAS conditional intensity predictor with the true parameters (green dashed line) performs very similarly to the ETAS conditional intensity predictor with estimated parameters (blue dotted line). The magnitude-dependent automatic alarms have parameter $u = 3.70$, chosen to minimise area under the error diagram in the training set. In the test set (solid black line), they perform slightly better than automatic alarms (red dotted-dashed line) and slightly worse than the ETAS conditional intensity predictors. No single strategy dominated any other single strategy.

ETAS v MDA: SCEC Data (Luen, 2010)

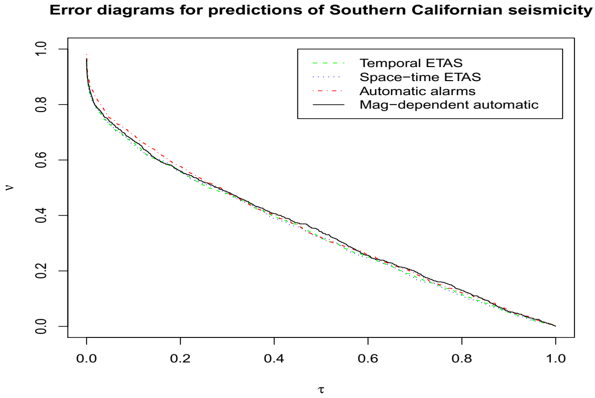

Figure 5.9: Error diagrams for predictors of Southern Californian seismicity. The predictors were fitted to the SCEC catalog from January 1st, 1984 to June 17th, 2004, and tested on the SCEC catalog from June 18th, 2004 to December 31st, 2009. For low values of $\hat{\tau}$, simple automatic alarms do not perform as well as the ETAS predictors. For high values of $\hat{\tau}$, MDA alarms do not perform as well as the ETAS predictors. Note that although success rates are determined for the test set only, predictors used both training and test data to determine times since past events (for simple automatic and MDA alarms) and conditional intensity (for ETAS predictors).

ETAS v MDA: SCEC Data (Luen, 2010)

Predictor	Training area	Test area	LQ test area
Space-time ETAS	0.234	0.340	0.161
Temporal ETAS	0.235	0.341	0.161
Typical ETAS	0.236	0.345	0.161
MDA, $u = 2$	0.253	0.348	0.165
MDA, $u = 5.8$	0.240	0.351	0.163
Simple auto	0.254	0.352	0.168

Table 5.5: Success of several predictors of Southern Californian earthquakes of magnitude $M \geq 3$. The predictors are fitted to a training set of data (the SCEC catalog from January 1st, 1984 to June 17th, 2004) and assessed on a test set of data (the catalog from June 18th, 2004 to December 31st, 2009). The predictors have parameters estimated on the training set, but may use times and magnitudes of training events in the test. The measures of success are area under the training set error diagram, area under the test set error diagram, and area under the left quarter of the test set error diagram. "Space-time ETAS" is a conditional intensity predictor using Veen and Schoenberg's space-time parameter estimates, given in the "VS spatial estimate" column of Table 4.2. "Temporal ETAS" uses parameters estimated using a temporal ETAS model, given in the "Temporal estimate" column of Table 4.2. "Typical ETAS" uses the parameters in Table 4.3. "MDA, $u = 2$" is a magnitude-dependent automatic alarm strategy with base 2. "MDA, $u = 5.8$" is an MDA alarm strategy alarm strategy with base determined by fitting alarms to a test set. "Simple auto" is a simple automatic alarm strategy.

Conclusions

- The probability models do not have a defensible basis in physics.
- They do not describe seismicity in a way that is probabilistically adequate, on the assumption that they are true (they fail goodness of fit tests)
- They do not appear to predict better than far simpler methods.
- Why are we so attached to these probability models?