Earthquake Inter-Event Time Modeling: Theoretical Explanation

Contents

1 Introduction

Understanding the temporal patterns of earthquakes is crucial for seismic hazard assessment and risk mitigation. One approach to analyzing these patterns is by studying the time intervals between consecutive earthquakes, known as **inter-event times**. This document provides a theoretical overview of statistical models used to analyze earthquake inter-event times, focusing on both standard probability distributions and advanced modeling approaches.

2 Statistical Tests for Model Evaluation

Statistical tests are essential for evaluating how well a proposed model fits the observed data. In the context of earthquake inter-event time modeling, goodness-of-fit tests help determine whether a particular probability distribution is an appropriate model for the inter-event times.

2.1 Kolmogorov-Smirnov Test

Overview:

- Non-parametric test that compares the empirical distribution function of the data with the cumulative distribution function (CDF) of the proposed model.
- Assesses the hypothesis that the data follow a specified distribution.

Test Statistic:
The Kolmogorov-Smirnov (K-S) test statistic is defined as:

$$D = \sup_t |F_n(t) - F(t)|$$

where:

- $F_n(t)$ is the empirical distribution function of the sample data.
- $F(t)$ is the cumulative distribution function of the hypothesized distribution.
- $\sup_t$ denotes the supremum (maximum) over all values of t.

Interpretation:

- A smaller value of D indicates a better fit between the data and the model.
- The p-value associated with D helps determine the statistical significance.
- If the p-value is less than a chosen significance level (e.g., $\alpha = 0.05$), the null hypothesis that the data follow the specified distribution is rejected.

Applicability to Earthquake Data:

- Used to test whether inter-event times follow a proposed probability distribution (e.g., exponential, Weibull).
- Helps in selecting the most appropriate model among several candidates.

2.2 Other Goodness-of-Fit Tests

Anderson-Darling Test:

- A modification of the K-S test that gives more weight to the tails of the distribution.
- Often more sensitive than the K-S test for detecting deviations at the extremes.

Chi-Square Goodness-of-Fit Test:

- Compares the observed frequencies of data falling into predefined intervals with the expected frequencies under the model.
- Requires sufficient sample size and appropriate binning.

2.3 Information Criteria

Akaike Information Criterion (AIC):

- Provides a measure of the relative quality of statistical models for a given dataset.
- Balances model fit with model complexity.

Formula:

$$\text{AIC} = 2k - 2\ln(L)$$

where:

- k is the number of parameters in the model.
- L is the maximized value of the likelihood function for the model.

Interpretation:

- Lower AIC values indicate a better balance of goodness-of-fit and simplicity.
- Used for model selection among a set of candidate models.

Bayesian Information Criterion (BIC):

- Similar to AIC but includes a stronger penalty for models with more parameters.

Formula:

$$\text{BIC} = k\ln(n) - 2\ln(L)$$

where:

- n is the number of observations.

Applicability to Earthquake Data:

- Helps in comparing models with different numbers of parameters.
- Supports the selection of a model that provides an adequate fit without unnecessary complexity.

3 Standard Probability Distributions

Standard probability distributions are often used as initial models to describe the statistical properties of inter-event times. These distributions have well-understood properties and provide a baseline for more complex models.

3.1 Exponential Distribution

Overview:

- Models the time between independent events occurring at a constant average rate.
- Suitable for a **Poisson process** with a constant hazard rate and memoryless property.

Probability Density Function (PDF):

$$f(t; \lambda) = \lambda e^{-\lambda t}, \quad t \geq 0$$

where $\lambda > 0$ is the rate parameter (events per unit time).
Properties:

- **Memoryless Property:** The future probability is independent of the past.
- **Constant Hazard Rate:** The event rate λ does not change over time.

Applicability to Earthquakes:

- Serves as a baseline model for random, independent earthquake occurrences.
- **Limitations:** Does not account for aftershock clustering or varying seismic rates.

3.2 Weibull Distribution

Overview:

- A flexible distribution that can model increasing, decreasing, or constant hazard rates.
- Captures different behaviors in the occurrence of events over time.

PDF:

$$f(t; k, \lambda) = \frac{k}{\lambda} \left(\frac{t}{\lambda} \right)^{k-1} e^{-(t/\lambda)^k}, \quad t \geq 0$$

where $k > 0$ is the shape parameter and $\lambda > 0$ is the scale parameter.
Hazard Function:

$$h(t) = \frac{k}{\lambda} \left(\frac{t}{\lambda} \right)^{k-1}$$

- $k < 1$: Decreasing hazard rate.
- $k = 1$: Constant hazard rate (reduces to exponential distribution).
- $k > 1$: Increasing hazard rate.

Applicability to Earthquakes:

- Models the decreasing rate of aftershocks after a main event ($k < 1$).
- Captures the variability in inter-event times better than the exponential distribution.

3.3 Gamma Distribution

Overview:

- Generalizes the exponential distribution with an additional shape parameter.
- Suitable for modeling waiting times for multiple events.

PDF:

$$f(t; \alpha, \beta) = \frac{\beta^\alpha t^{\alpha-1} e^{-\beta t}}{\Gamma(\alpha)}, \quad t \geq 0$$

where $\alpha > 0$ is the shape parameter, $\beta > 0$ is the rate parameter, and $\Gamma(\alpha)$ is the gamma function.
Properties:

- **Flexibility:** Can model various shapes of data distributions.
- **Overdispersion:** Suitable when data variance exceeds the mean.

Applicability to Earthquakes:

- Captures more variability in inter-event times.
- Useful when inter-event times are not memoryless and exhibit overdispersion.

3.4 Log-Normal Distribution

Overview:

- Models a random variable whose logarithm is normally distributed.
- Appropriate for modeling positive-valued data with right skewness.

PDF:

$$f(t; \mu, \sigma) = \frac{1}{t\sigma\sqrt{2\pi}} e^{-\frac{(\ln t - \mu)^2}{2\sigma^2}}, \quad t > 0$$

where μ is the mean and $\sigma > 0$ is the standard deviation of the natural logarithm of the variable.
Properties:

- **Positive Skewness:** Captures long-tail behavior in data.
- **Multiplicative Processes:** Suitable when events result from multiplicative factors.

Applicability to Earthquakes:

- Models inter-event times with occasional large intervals.
- Reflects the variability in seismic activity over time.

4 Advanced Modeling Approaches

While standard distributions provide insights into the statistical properties of inter-event times, they often fail to capture the complex temporal dependencies observed in earthquake occurrences, such as aftershock clustering and non-stationarity. Advanced models address these limitations.

4.1 Hawkes Process

Overview:

- A **self-exciting point process** where each event increases the likelihood of subsequent events for some period.
- Captures the clustering behavior observed in aftershock sequences.

Mathematical Formulation:
Conditional Intensity Function ($\lambda(t)$):

$$\lambda(t) = \mu + \sum_{t_i < t} \phi(t - t_i)$$

where:

- μ is the background rate of events.
- $\phi(t - t_i)$ is the triggering kernel function representing the influence of past events.

Exponential Kernel Example:

$$\phi(t - t_i) = \alpha e^{-\beta(t - t_i)}$$

where:

- α is the excitation parameter (influence magnitude).
- β is the decay rate of the excitation effect.

Properties:

- **Clustering:** Events cluster in time due to self-excitation.
- **Non-Stationarity:** The intensity function evolves over time based on past events.

Applicability to Earthquake Data:

- **Aftershock Modeling:** Effectively models the increased seismicity following a mainshock.
- **Background and Triggered Events:** Distinguishes between spontaneous events and those triggered by previous earthquakes.

Advantages:

- Captures temporal dependencies and clustering.
- Provides a framework for probabilistic forecasting of aftershocks.

Challenges:

- Requires estimation of multiple parameters.
- Computationally intensive for large datasets.

4.2 Epidemic-Type Aftershock Sequence (ETAS) Model

Overview:

- An extension of the Hawkes process tailored for seismicity.
- Considers both temporal and spatial components of earthquake occurrences.

Mathematical Formulation:
Conditional Intensity Function:

$$\lambda(t, x, y) = \mu(x, y) + \sum_{t_i < t} K_0 e^{\alpha(M_i - M_0)} \left(\frac{1}{(t - t_i + c)^p} \right) f_s(x - x_i, y - y_i)$$

where:

- $\mu(x, y)$ is the spatially varying background rate.
- K_0, α, c, p are model parameters.
- M_i is the magnitude of the i-th event.
- M_0 is the minimum magnitude threshold.
- f_s is the spatial distribution function.

Properties:

- **Magnitude Dependence:** Larger events have a higher potential to trigger aftershocks.
- **Spatial and Temporal Clustering:** Models how aftershocks are distributed in both space and time.

Applicability to Earthquake Data:

- **Comprehensive Seismicity Modeling:** Accounts for both mainshocks and aftershocks.
- **Risk Assessment:** Helps in understanding the spread and decay of aftershock sequences.

Advantages:

- Incorporates important physical aspects of seismicity.
- Improves accuracy in forecasting aftershock probabilities.

Challenges:

- Complex parameter estimation.
- Requires high-quality data, including accurate event locations and magnitudes.

4.3 Time-Varying Models

Overview:

- Models that allow parameters to change over time, capturing non-stationary behaviors in seismic activity.
- Reflects changes due to geological processes or external influences.

Examples:

- **Time-Varying Poisson Process:** The rate parameter $\lambda(t)$ is a function of time.

- **State-Space Models:** Uses latent variables to model the evolution of seismicity rates over time.

Properties:

- **Adaptability:** Parameters adjust in response to new data.
- **Non-Stationarity:** Accounts for changes in the underlying process.

Applicability to Earthquake Data:

- **Seismic Rate Changes:** Models periods of increased or decreased seismicity.
- **Response to External Factors:** Can incorporate effects such as fluid injection or tidal forces.

Advantages:

- Provides a dynamic view of seismic hazard.
- Can detect and model trends or cycles in seismicity.

Challenges:

- Requires complex statistical methods.
- Parameter estimation may be unstable with limited data.

5 Model Selection and Application

Selecting an appropriate model depends on the characteristics of the earthquake data and the specific goals of the analysis.

5.1 Factors to Consider

1. **Data Availability and Quality:**
    - Quantity and accuracy of event times, magnitudes, and locations.
2. **Seismicity Characteristics:**
    - Presence of aftershocks and clustering.
    - Variations in seismicity rates over time.
3. **Model Complexity vs. Interpretability:**
    - Simpler models are easier to implement but may not capture complex behaviors.
    - Complex models provide better fit but require more data and computational resources.

5.2 Steps in Model Application

1. **Exploratory Data Analysis:**
    - Examine inter-event time distributions.
    - Identify patterns such as clustering or periodicity.
2. **Model Fitting:**
    - Estimate model parameters using methods like Maximum Likelihood Estimation (MLE).

- For advanced models, specialized techniques or software may be required.

3. **Model Evaluation:**

   - Use goodness-of-fit tests (e.g., Kolmogorov-Smirnov test) for standard distributions.
   - Compare models using information criteria like AIC or BIC.
   - Assess predictive performance through cross-validation or out-of-sample testing.

4. **Interpretation and Forecasting:**

   - Interpret model parameters in the context of seismicity.
   - Use the model to forecast future earthquake probabilities or rates.

6 Limitations and Considerations

- **Independence Assumption:**
  - Standard distributions assume events are independent, which may not hold in the presence of aftershocks.
- **Stationarity Assumption:**
  - Many models assume a constant rate over time, which may not reflect reality.
- **Data Limitations:**
  - Incomplete or biased data can lead to inaccurate parameter estimates.
  - Detection thresholds can result in missing small events.
- **Model Uncertainty:**
  - All models are simplifications and may not capture all aspects of seismicity.
  - Uncertainty in parameter estimates should be considered in interpretations.

7 Practical Implications

- **Seismic Hazard Assessment:**
  - Models help estimate the likelihood of future earthquakes, informing building codes and risk mitigation strategies.
- **Emergency Preparedness:**
  - Understanding aftershock probabilities aids in planning for emergency response and resource allocation.
- **Scientific Understanding:**
  - Modeling contributes to the understanding of earthquake processes and stress interactions in the Earth's crust.

8 Conclusion

Modeling earthquake inter-event times is a vital component of seismology and risk assessment. While standard probability distributions offer a foundation for understanding seismicity patterns, advanced models like the Hawkes process and ETAS model provide deeper insights into the temporal dependencies and clustering behavior of earthquakes.

The choice of model should be guided by the data characteristics, the specific objectives of the analysis, and an understanding of the model assumptions and limitations. Combining different modeling approaches can enhance the robustness of seismic hazard assessments and contribute to more effective risk mitigation strategies.

9 References

- Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, 83(401), 9-27.
- Daley, D. J., & Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes.* Springer.
- Utsu, T., Ogata, Y., & Matsu'ura, R. S. (1995). The centenary of the Omori formula for a decay law of aftershock activity. *Journal of Physics of the Earth*, 43(1), 1-33.
- Reasenberg, P. A., & Jones, L. M. (1989). Earthquake hazard after a mainshock in California. *Science*, 243(4895), 1173-1176.