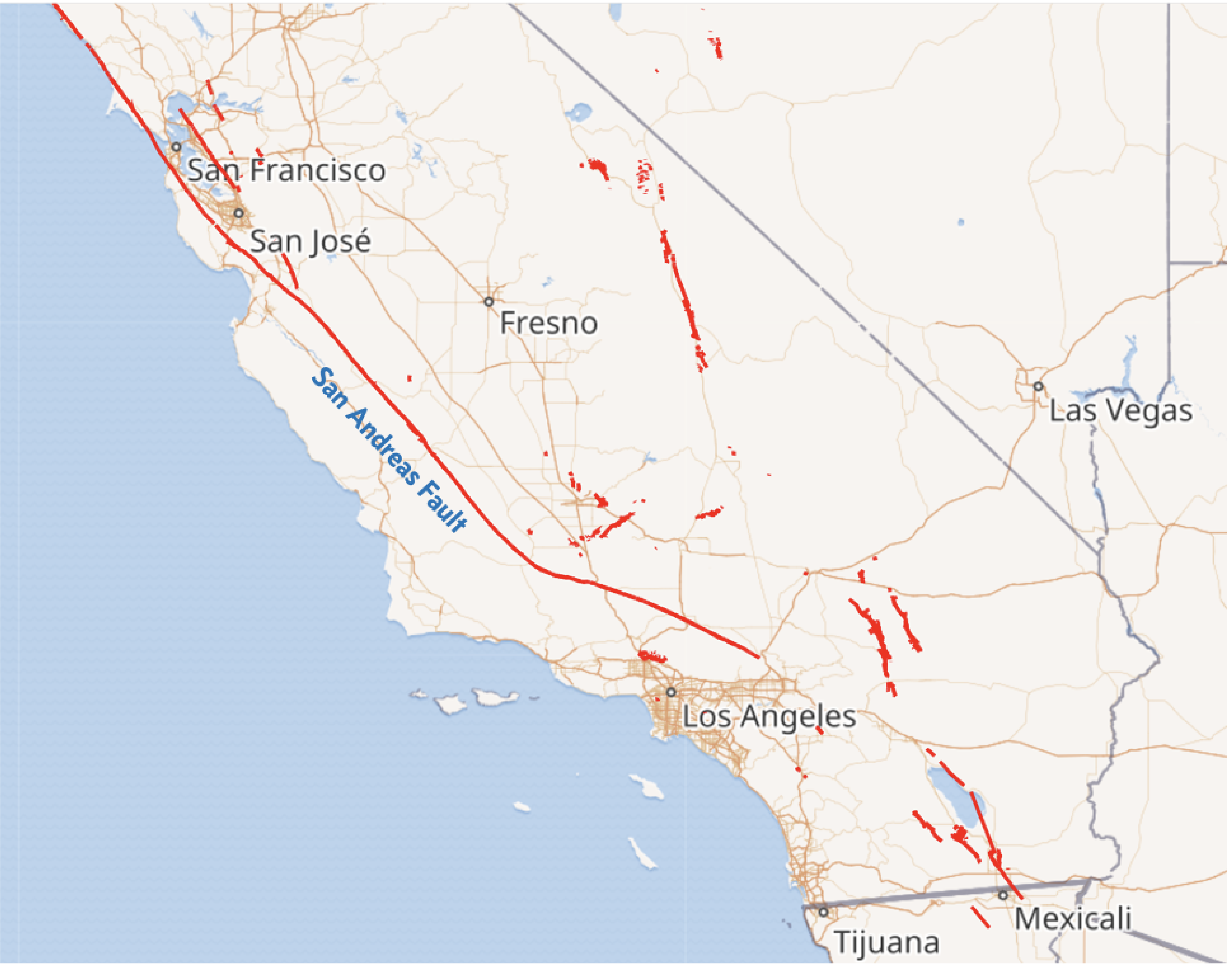
Introduction to statistical seismology

CASE STUDIES IN STATISTICAL THINKING

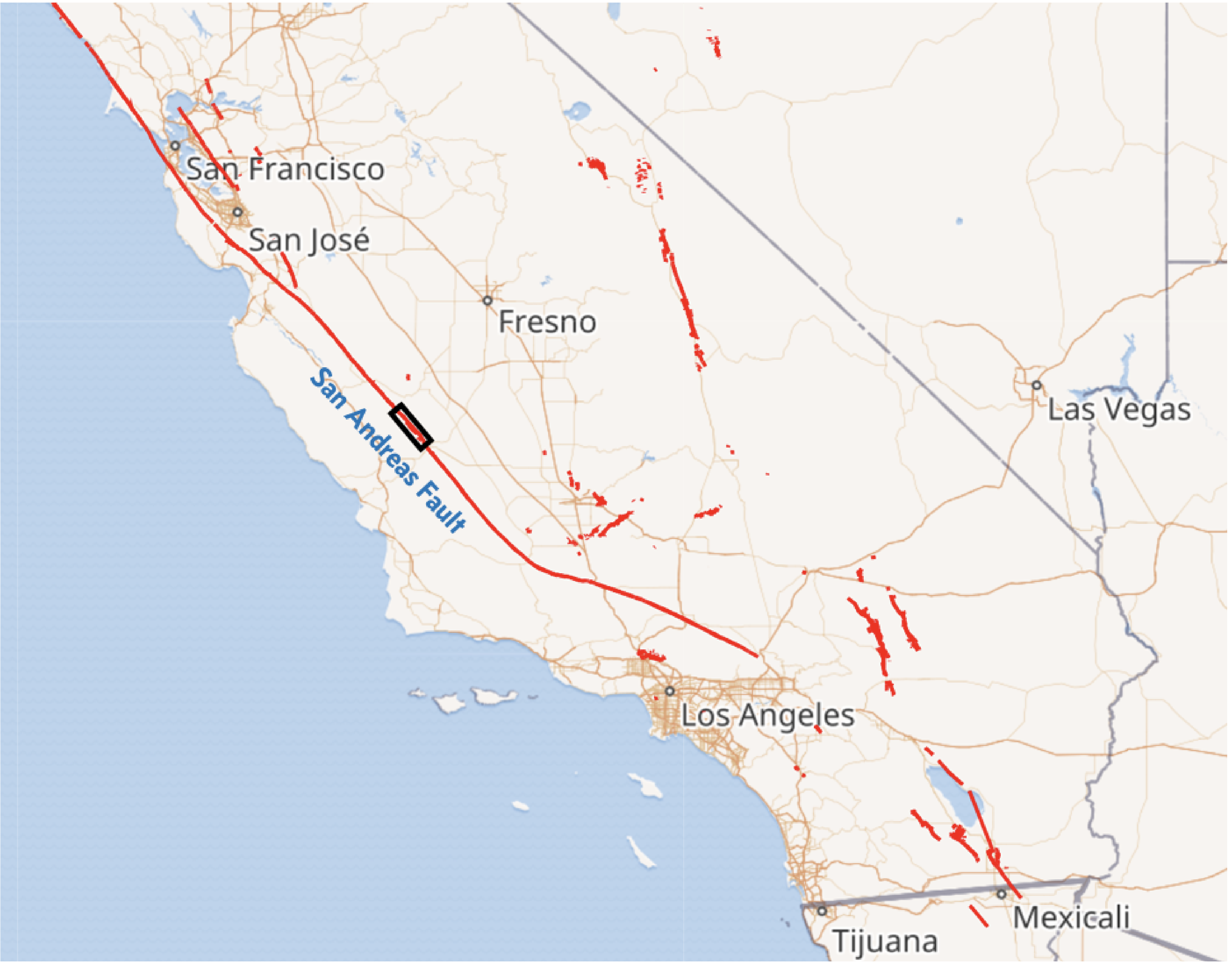
Justin Bois
Lecturer, Caltech

California moves and shakes



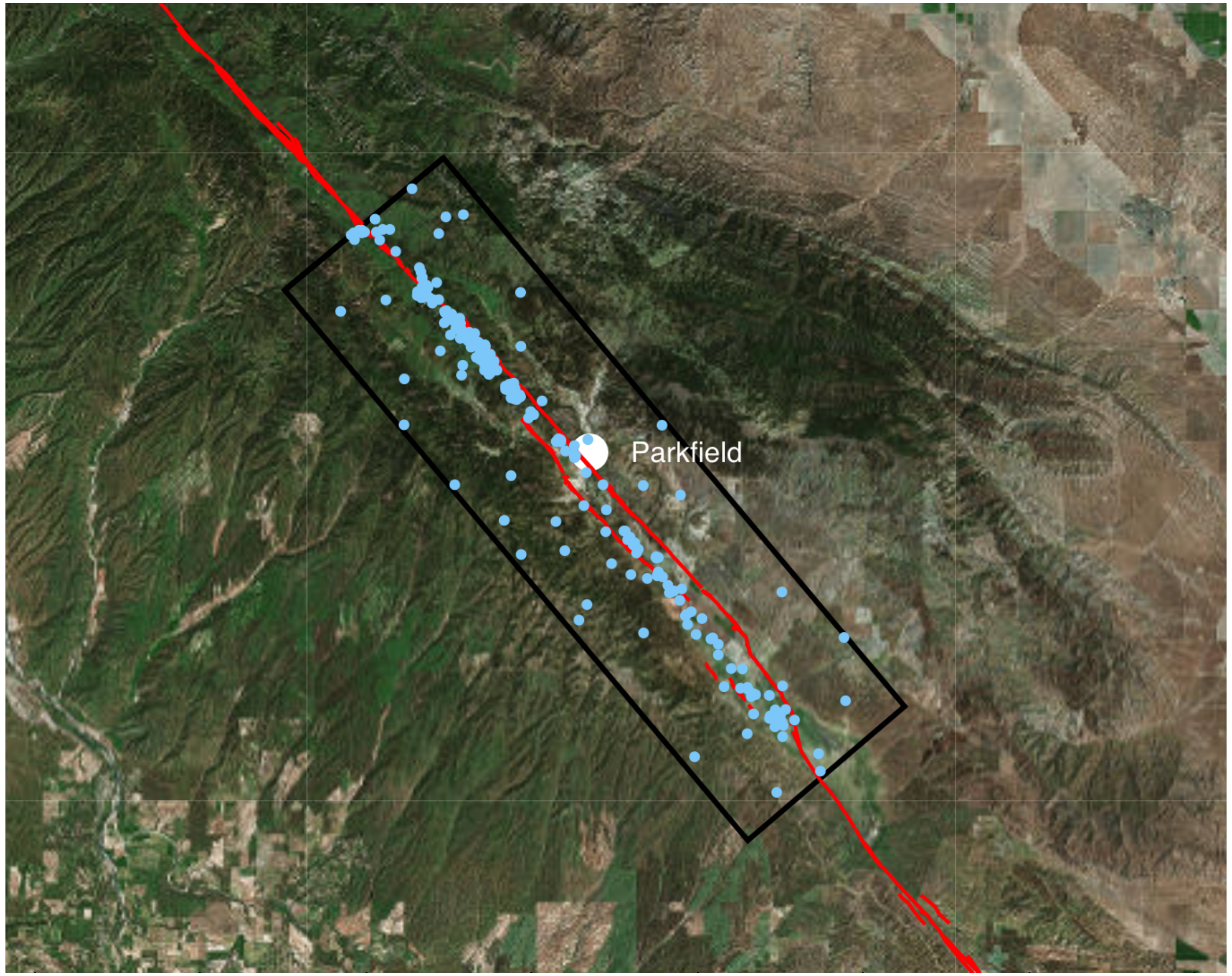
[1] Fault data: USGS Quaternary Fault and Fold Database of the United States

The Parkfield region



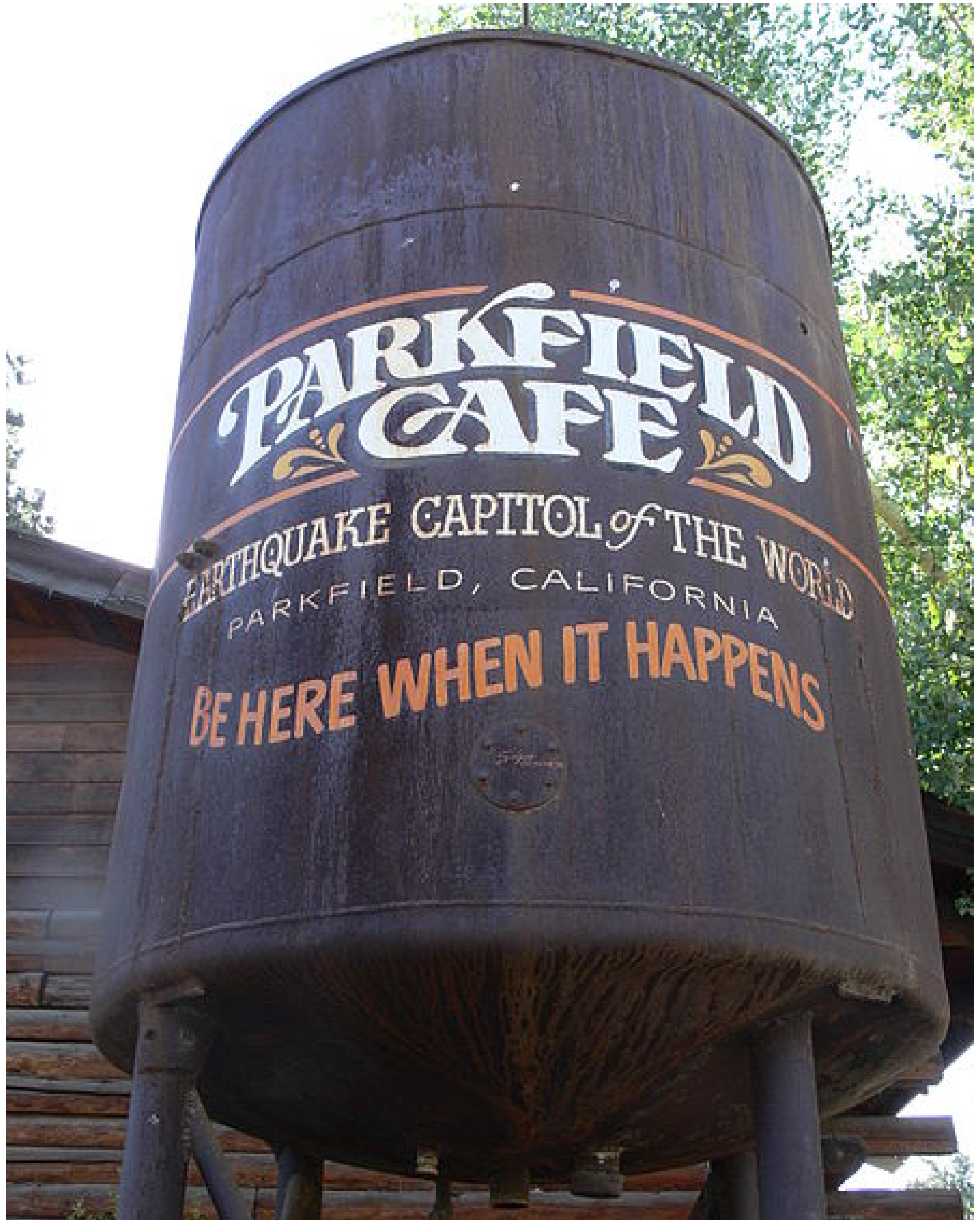
[1] Fault data: USGS Quaternary Fault and Fold Database of the United States

The Parkfield region



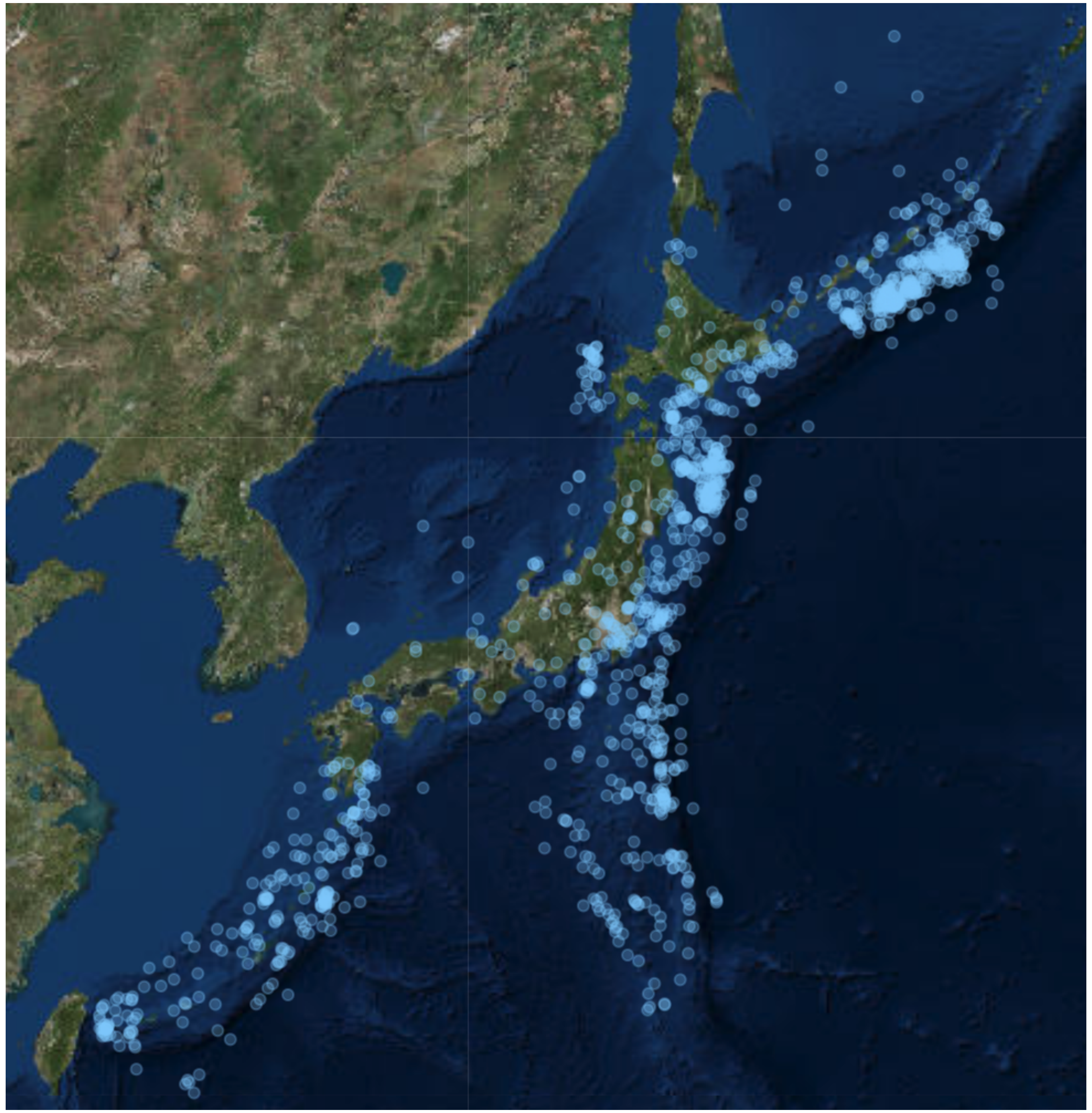
[1] Fault data: USGS Quaternary Fault Fault and Fold Database of the United States [2] Earthquake data: USGS ANSS Comprehensive Earthquake Catalog

The Parkfield region



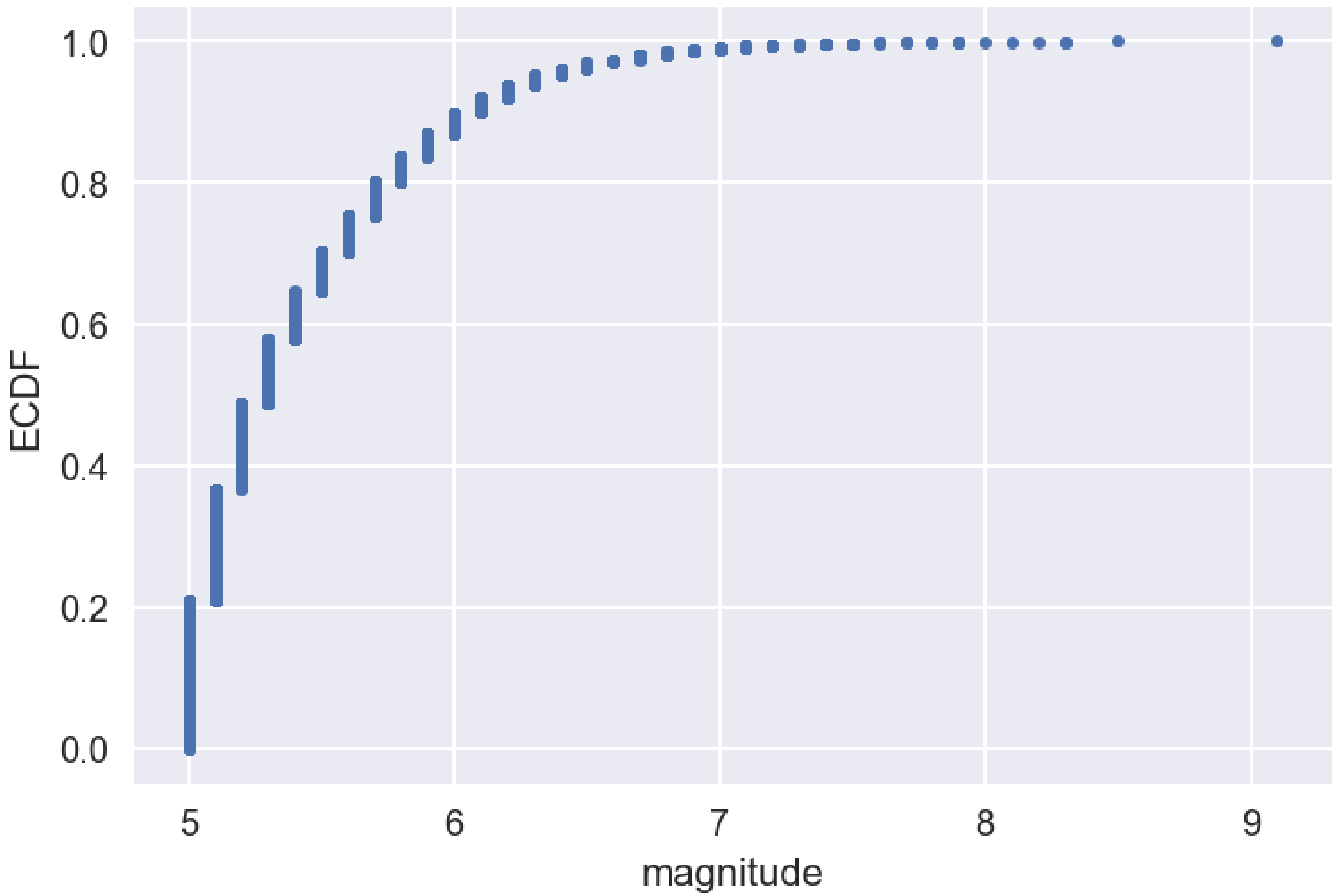
[1] Image: Linda Tanner, CC-BY-2.0

Seismic Japan



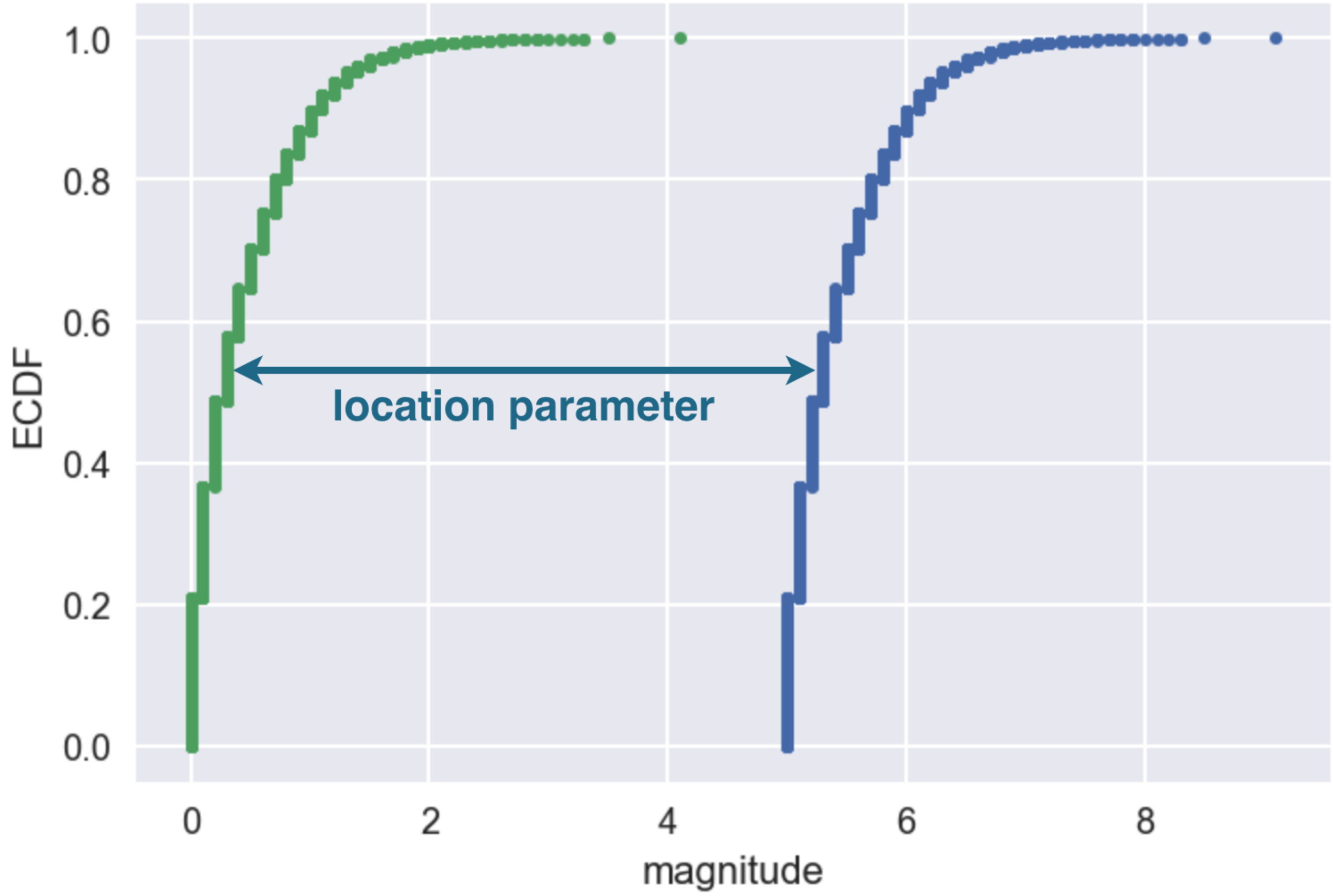
[1] Data source: USGS ANSS Comprehensive Earthquake Catalog (ComCat)

ECDF of magnitudes, Japan, 1990-1999



[1] Data source: USGS ANSS Comprehensive Earthquake Catalog (ComCat)

Location parameters



$$m' \equiv m - 5 \sim \text{Exponential}$$

$$m' \equiv m - m_t \sim \text{Exponential}$$

The Gutenberg-Richter Law

The magnitudes of earthquakes in a given region over a given time period are Exponentially distributed

One parameter, given by $\overline{m} - m_t$, describes earthquake magnitudes for a region

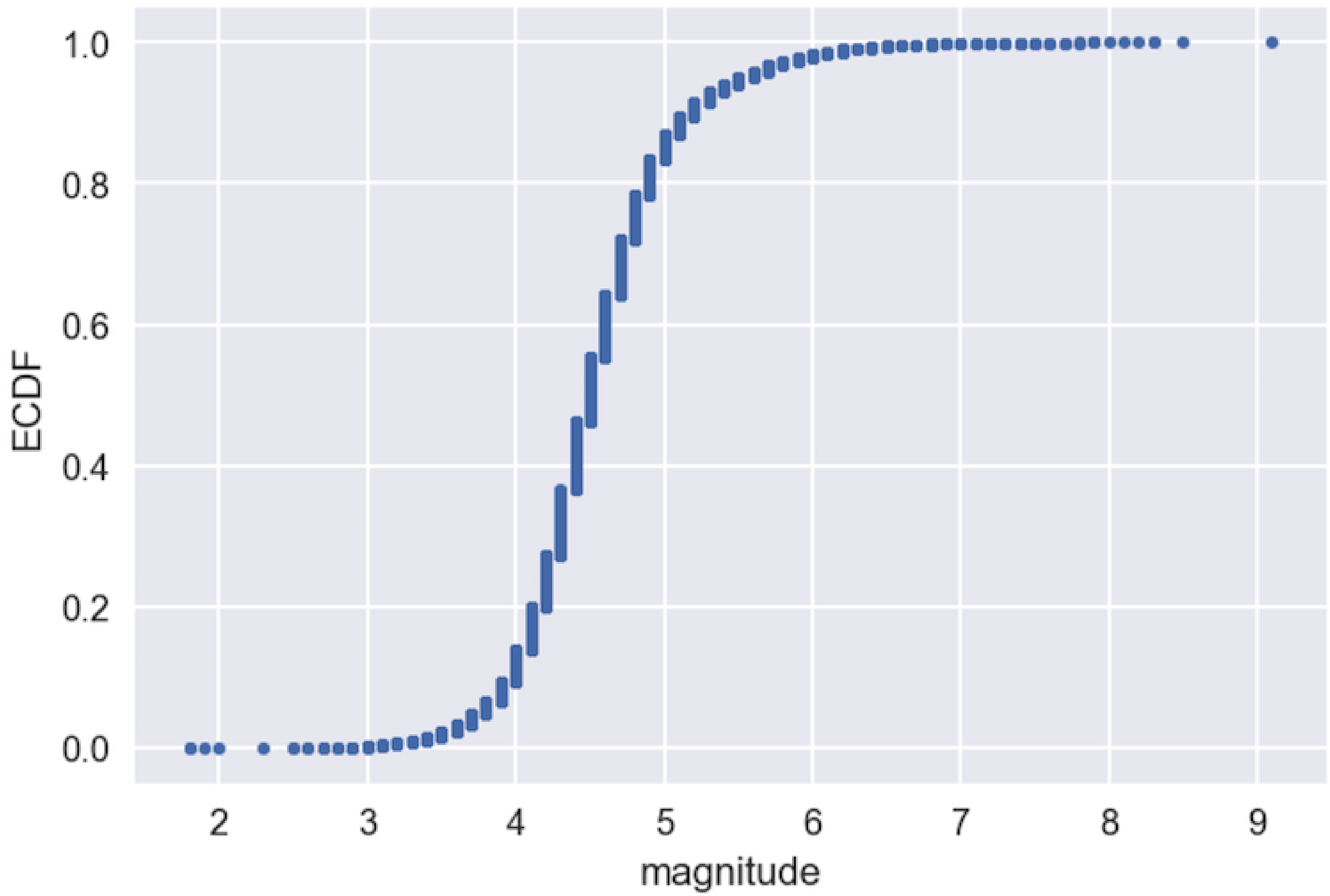
The b-value

$$b = (\overline{m} - m_t) \cdot \ln 10$$

```
# Completeness threshold
mt = 5

# b-value
b = (np.mean(magnitudes) - mt) * np.log(10)

print(b)
```

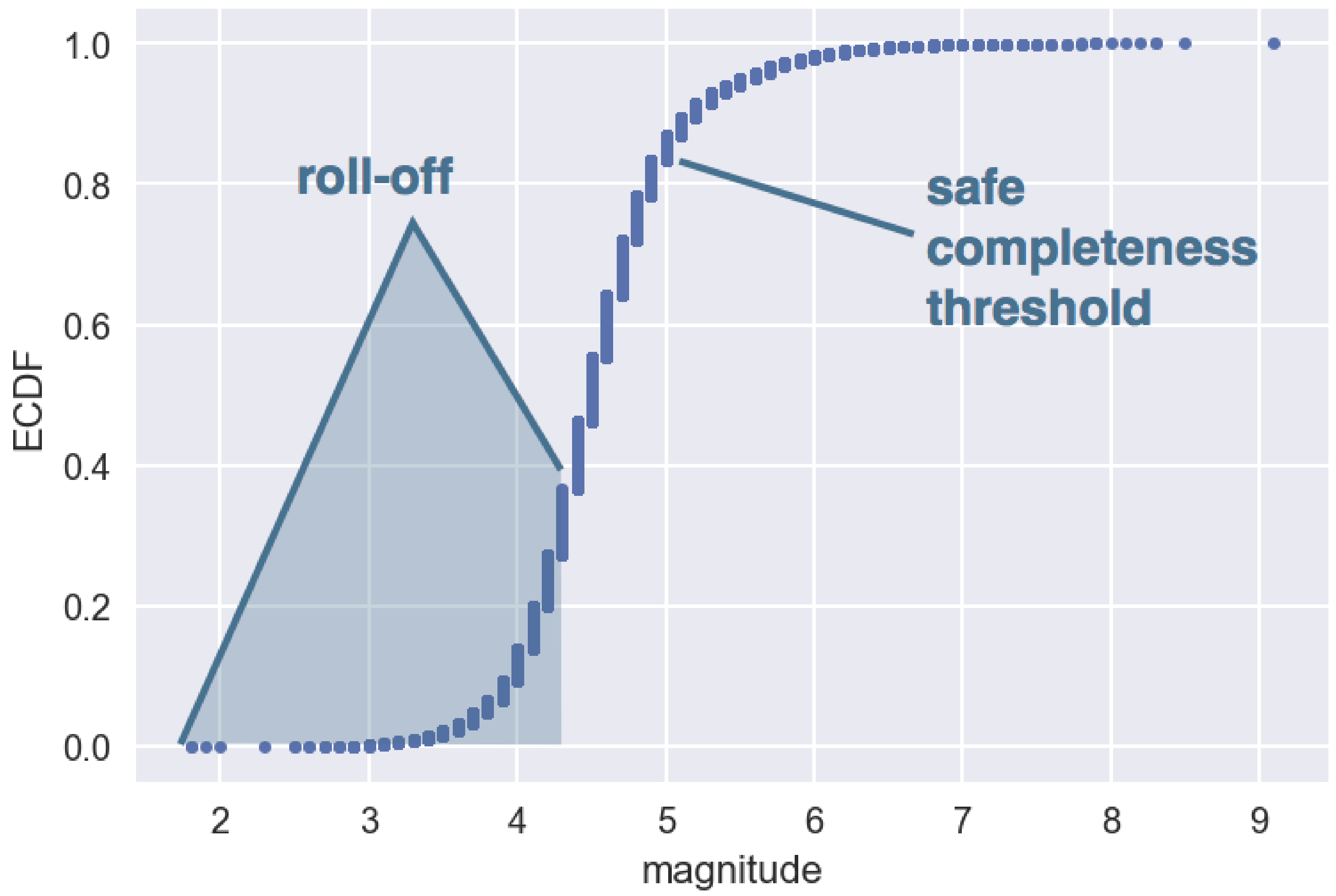
```
0.9729214742632566
```

ECDF of all magnitudes



[1] Data source: USGS ANSS Comprehensive Earthquake Catalog (ComCat)

ECDF of all magnitudes



[1] Data source: USGS ANSS Comprehensive Earthquake Catalog (ComCat)

Completeness threshold

The magnitude, m_t, above which all earthquakes in a region can be detected

Let's practice!

CASE STUDIES IN STATISTICAL THINKING

Timing of major earthquakes

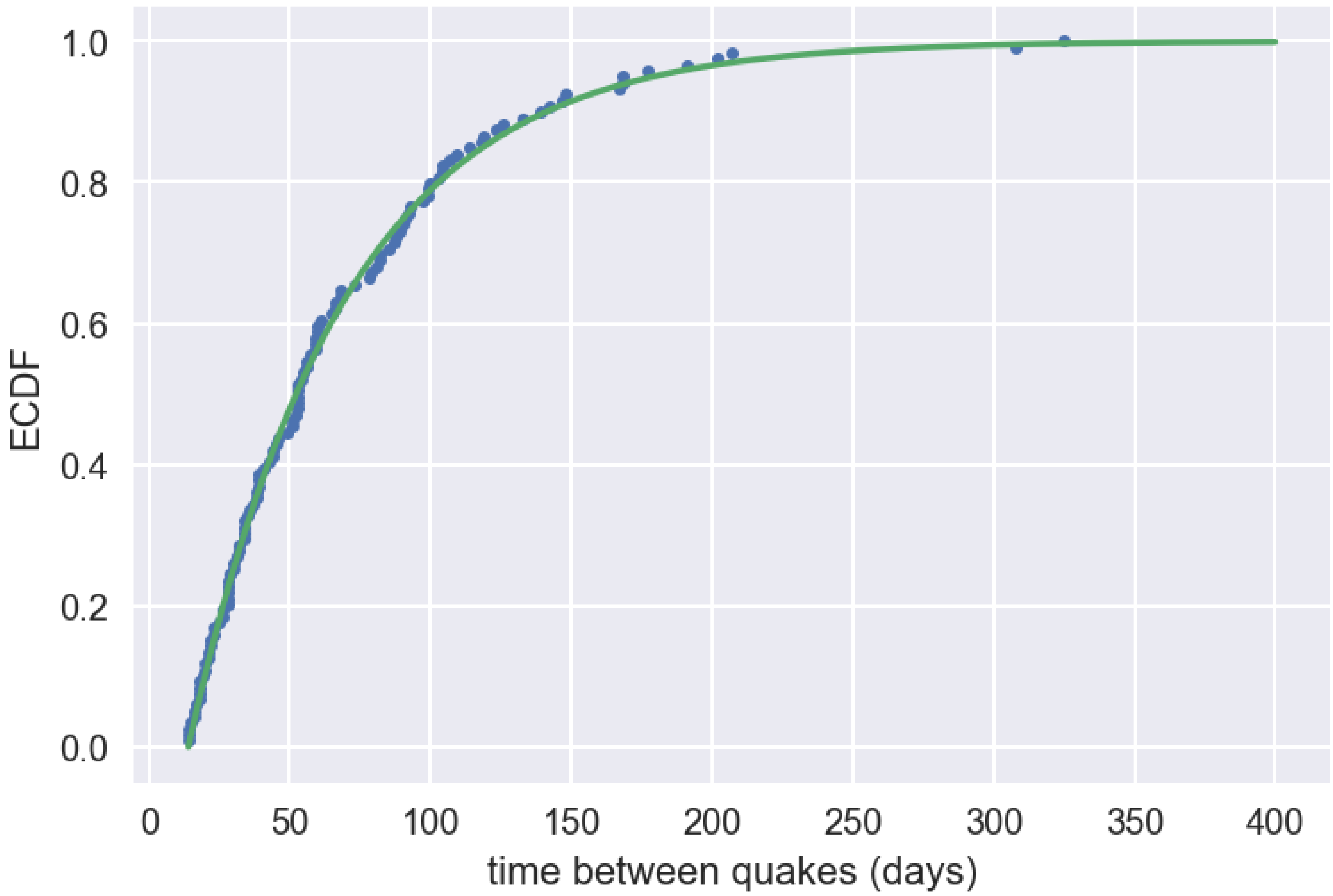
CASE STUDIES IN STATISTICAL THINKING

Justin Bois
Lecturer, Caltech

Models for earthquake timing

- Exponential: Earthquakes happen like a Poisson process
- Gaussian: Earthquakes happen with a well-defined period

Stable continental region earthquakes



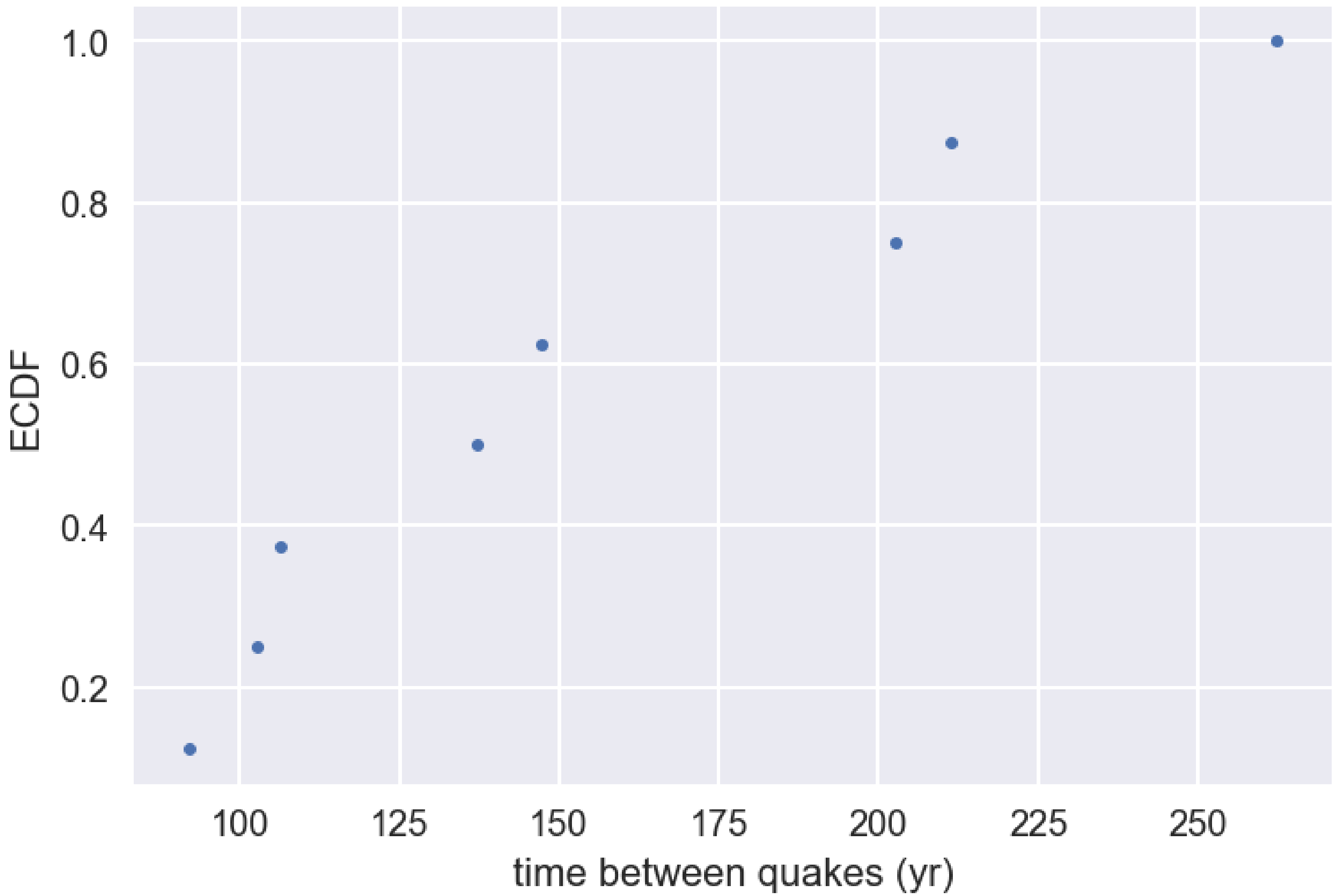
[1] Data source: USGS Earthquake Catalog for Stable Continental Regions

The Nankai Trough

Earthquakes in the Nankai Trough

Date	Magnitude
684-11-24	8.4
887-08-22	8.6
1099-02-16	8.0
1361-07-26	8.4
1498-09-11	8.6
1605-02-03	7.9
1707-10-18	8.6
1854-12-23	8.4
1946-12-24	8.1

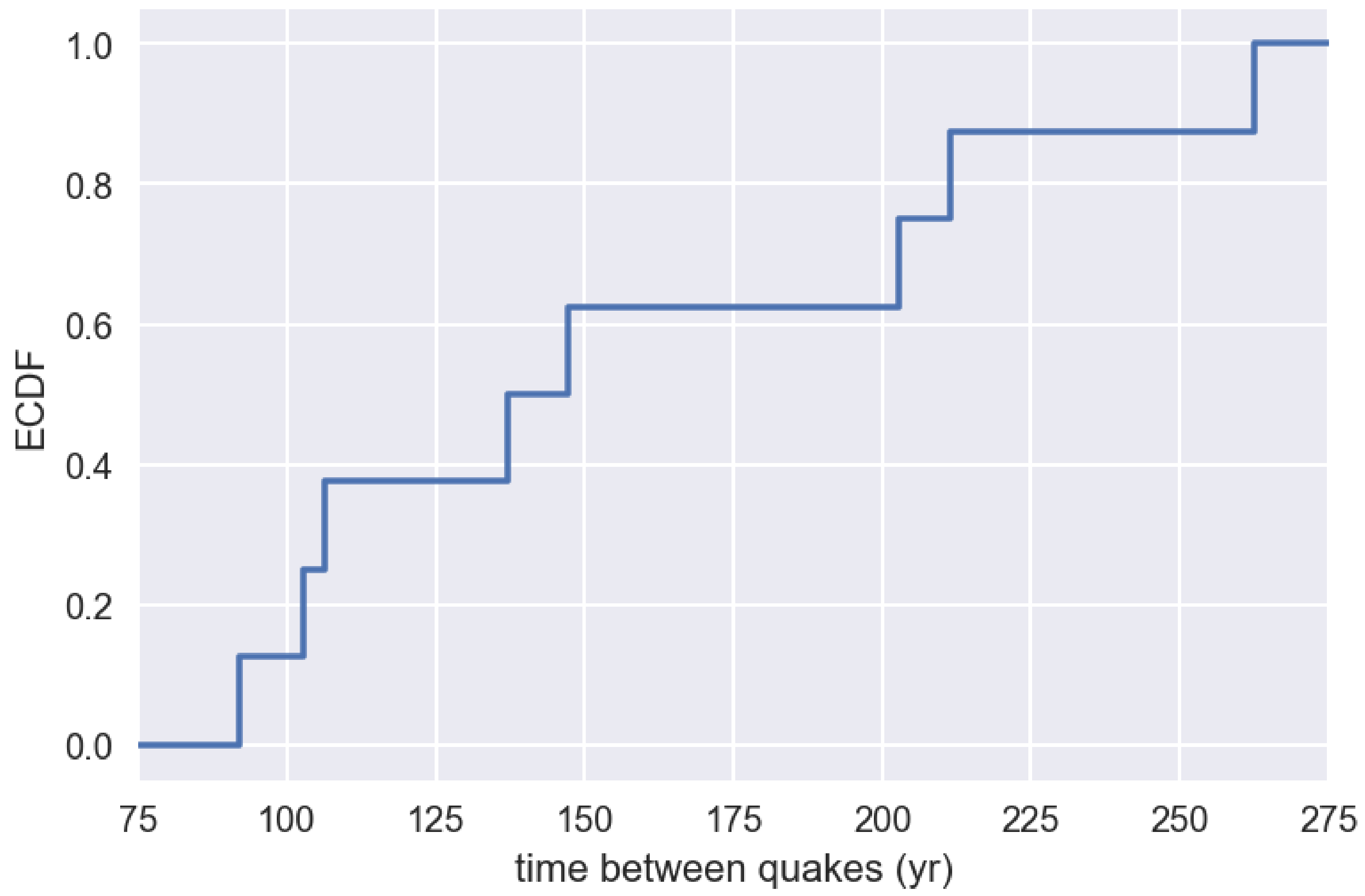
ECDF of time between Nankai quakes



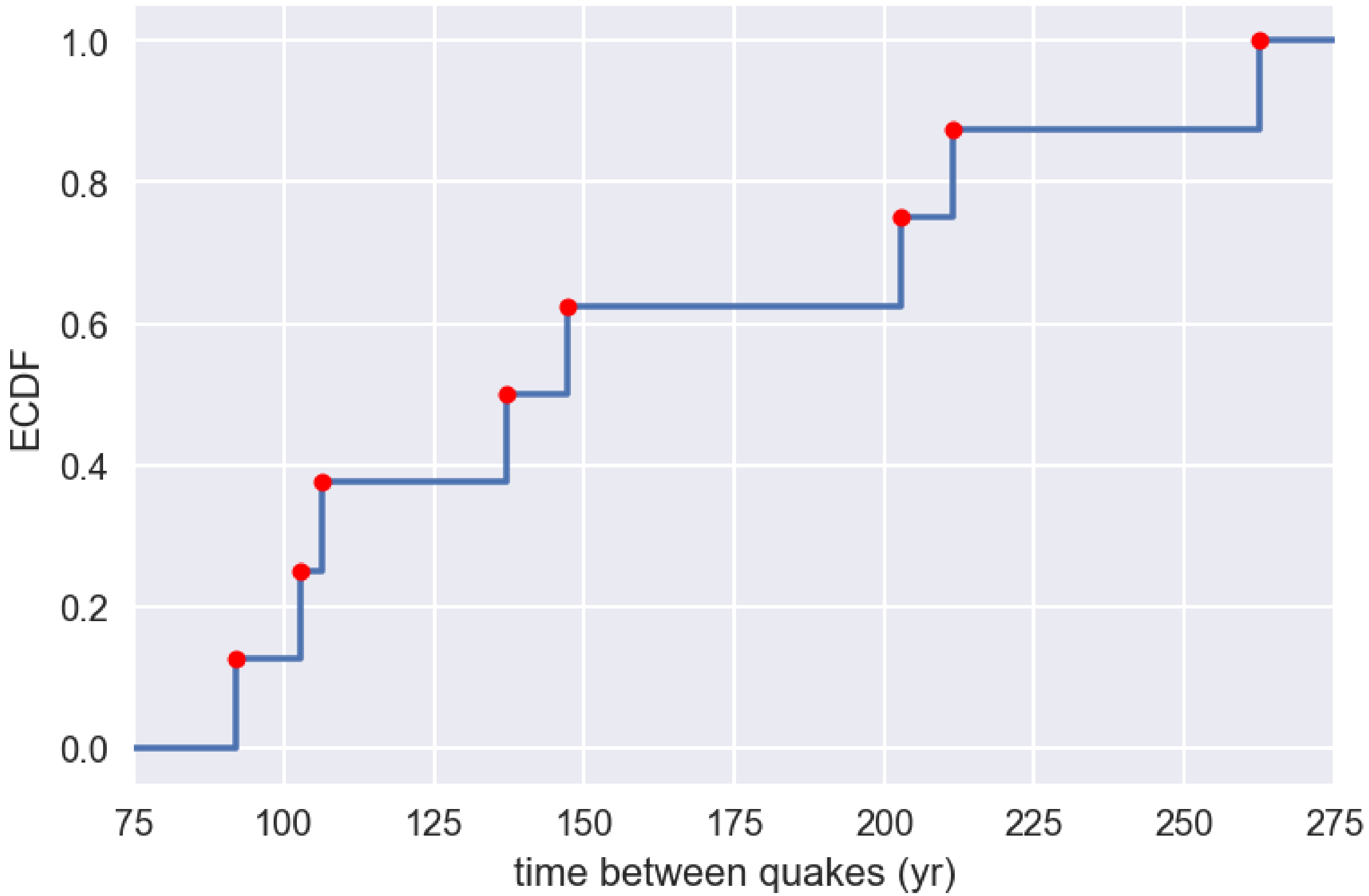
Formal ECDFs

$\text{ECDF}(x)$ = fraction of data points $\leq x$

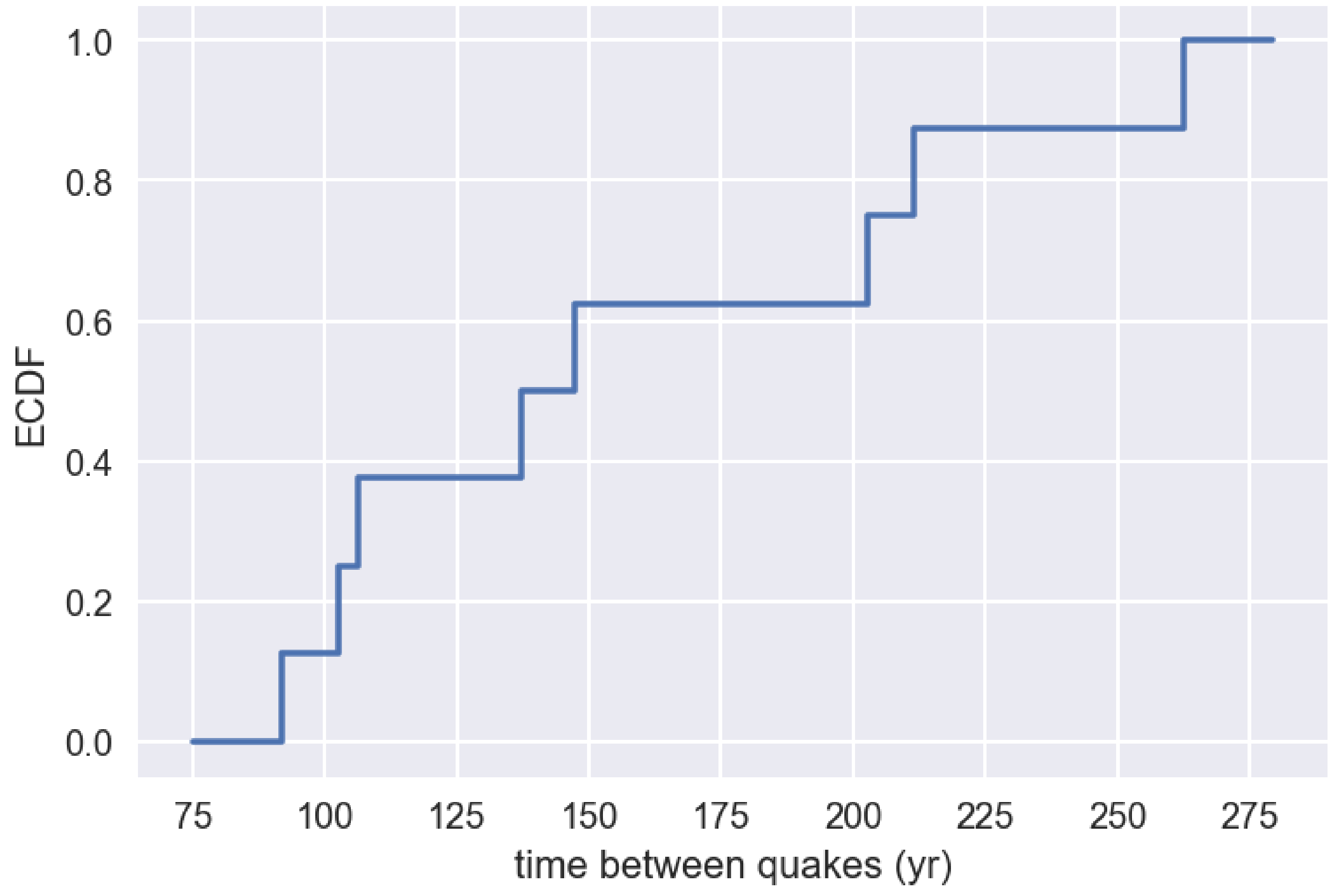
Formal ECDFs



Formal ECDFs



```python
# time_gap is an array of interearthquake times
_ = plt.plot(*dcst.ecdf(time_gap, formal=True))
_ = plt.xlabel('time between quakes (yr)')
_ = plt.ylabel('ECDF')
```

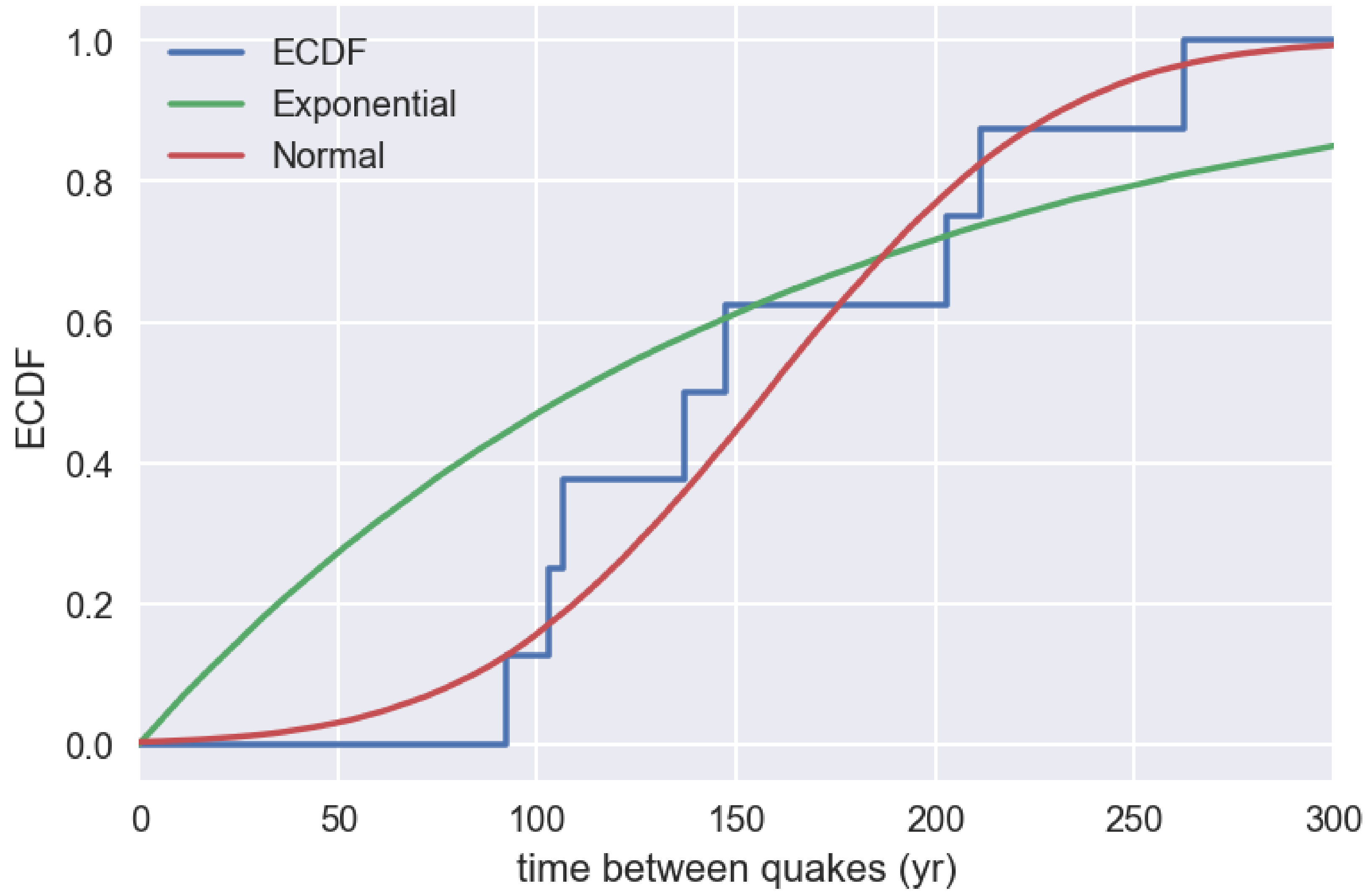


```
# Compute the mean time gap
mean_time_gap = np.mean(time_gap)

# Standard deviation of the time gap
std_time_gap = np.std(time_gap)
```

```
# Generate theoretical Exponential distribution of timings
time_gap_exp = np.random.exponential(mean_time_gap, size=100000)

# Generate theoretical Normal distribution of timings
time_gap_norm = np.random.normal(
  mean_time_gap, std_time_gap, size=100000
)
```

```
# Plot theoretical CDFs
_ = plt.plot(*dcst.ecdf(time_gap_exp))
_ = plt.plot(*dcst.ecdf(time_gap_norm))
```

Model for Nankai Trough



Let's practice!

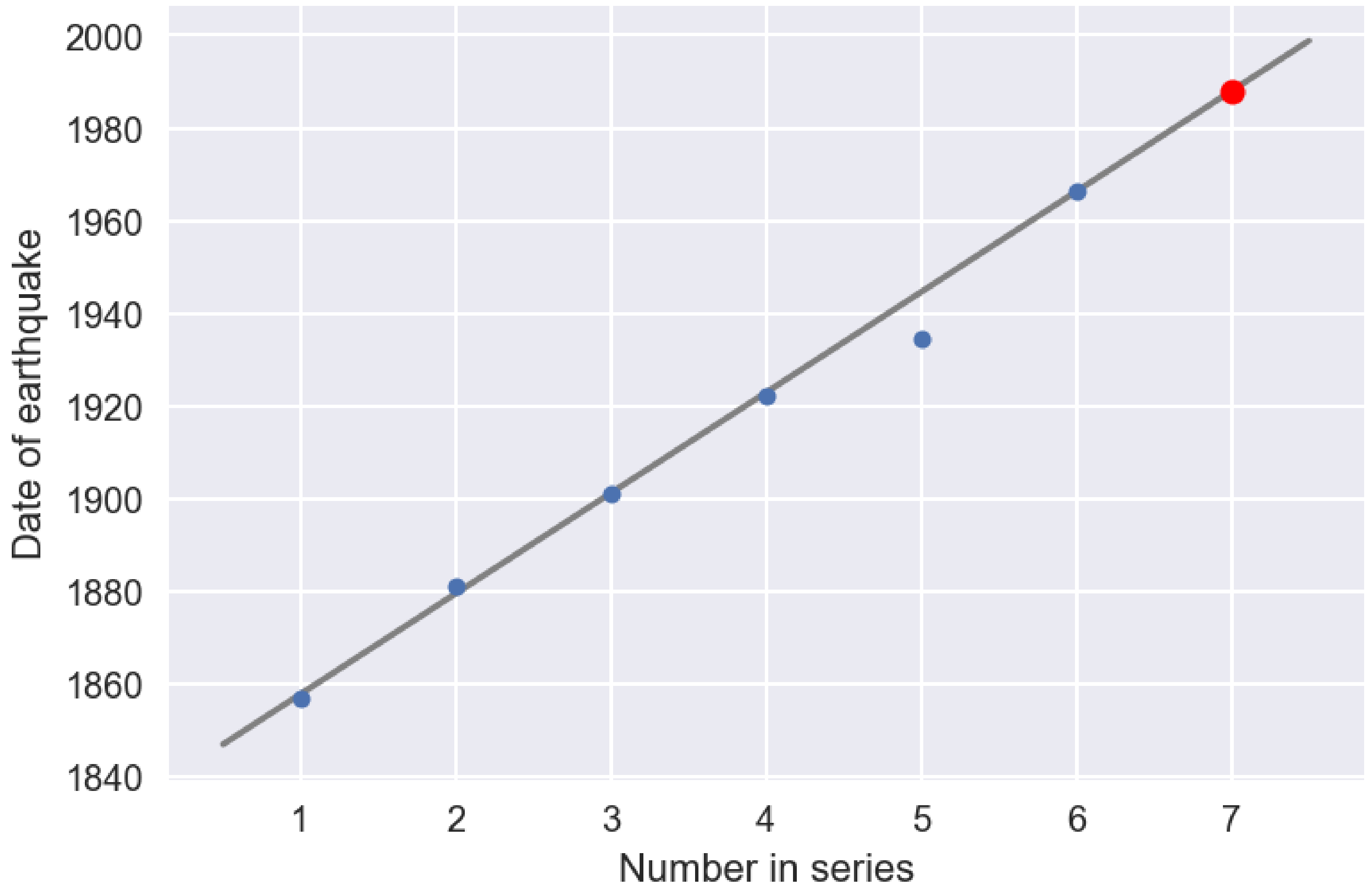
CASE STUDIES IN STATISTICAL THINKING

How are the Parkfield interearthquake times distributed?

CASE STUDIES IN STATISTICAL THINKING

Justin Bois
Lecturer, Caltech

The Parkfield Prediction

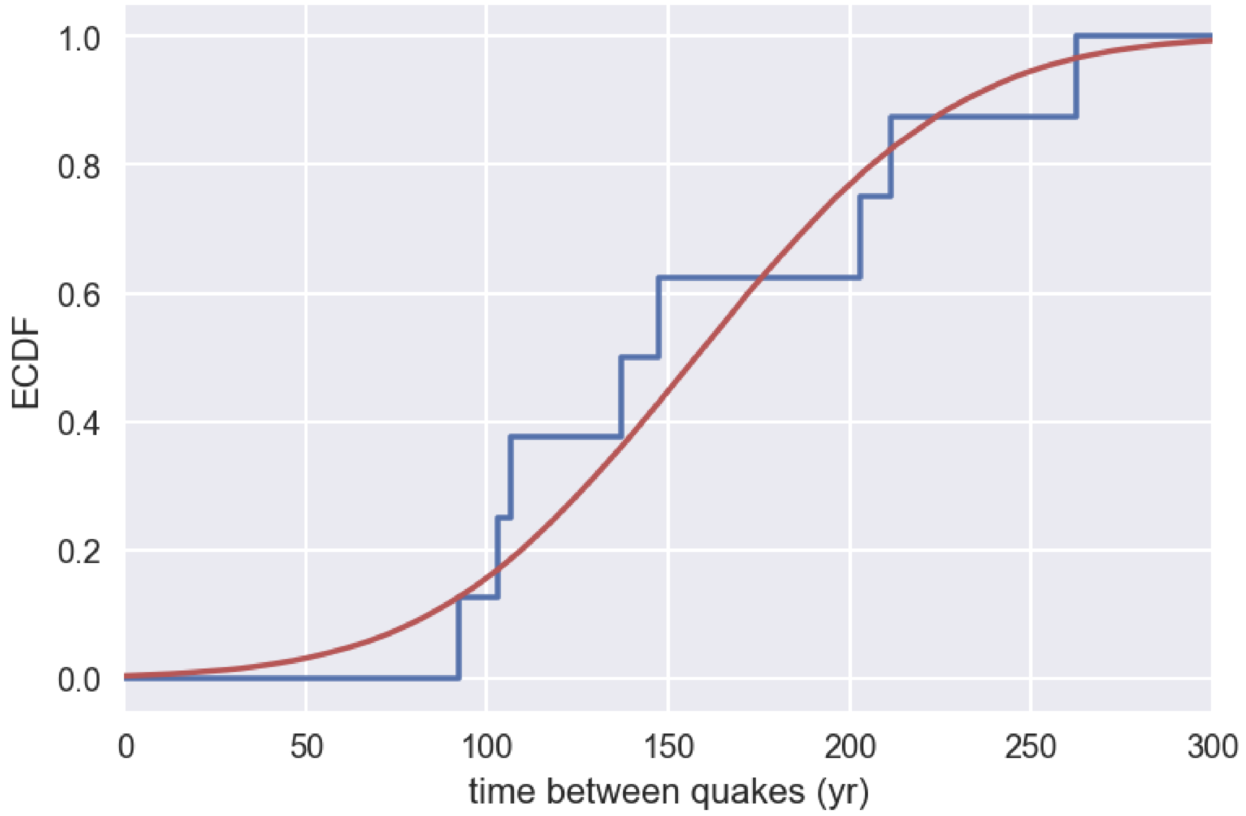


[1] Adapted from Barkun and Lindh, Science, 229, 619-624, 1985

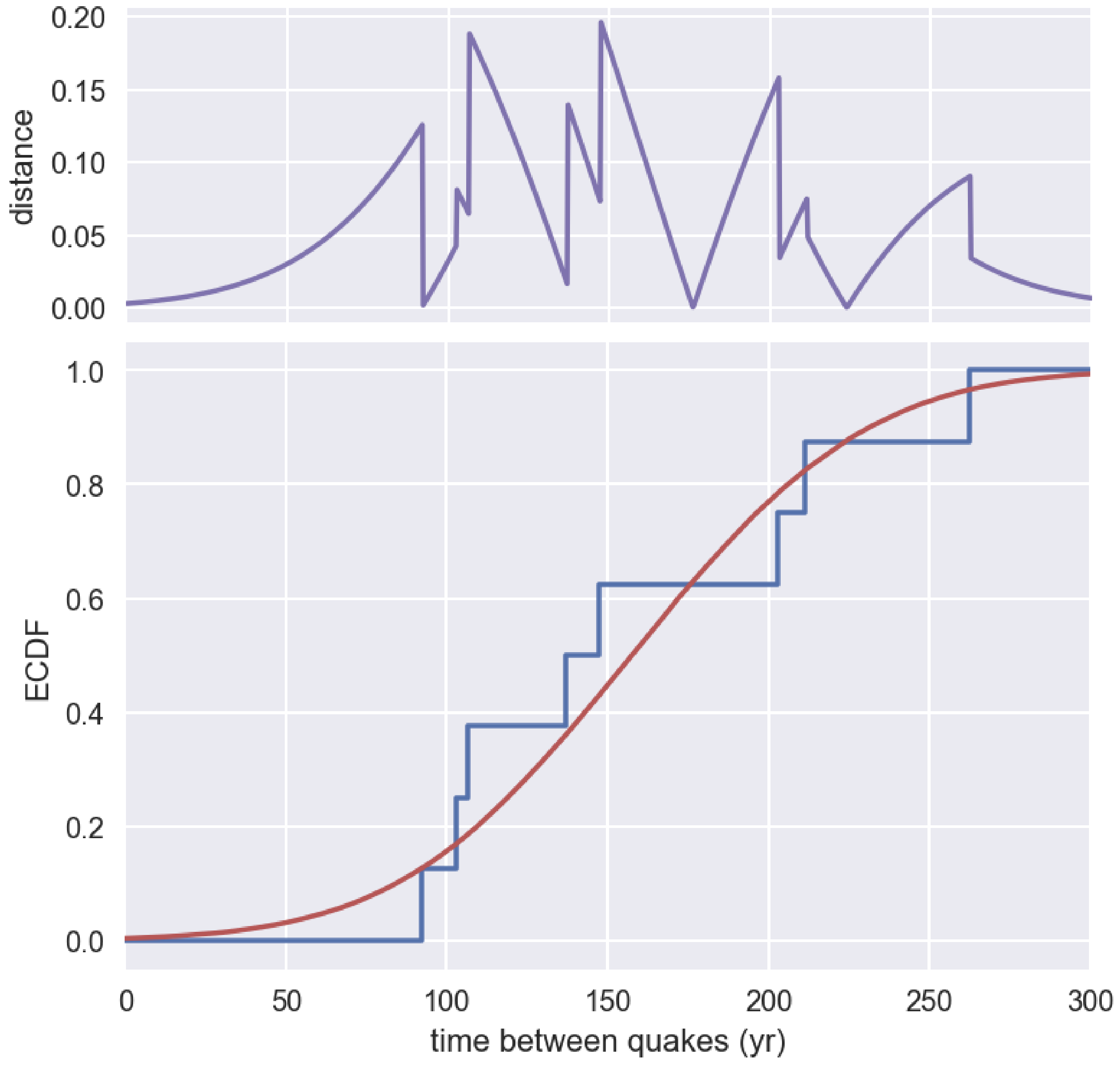
Hypothesis test on the Nankai megathrust earthquakes

- **Hypothesis:** The time between Nankai Trough earthquakes is Normally distributed with a mean and standard deviation as calculated from the data
- **Test statistic:** ??
- **At least as extreme as:** ??

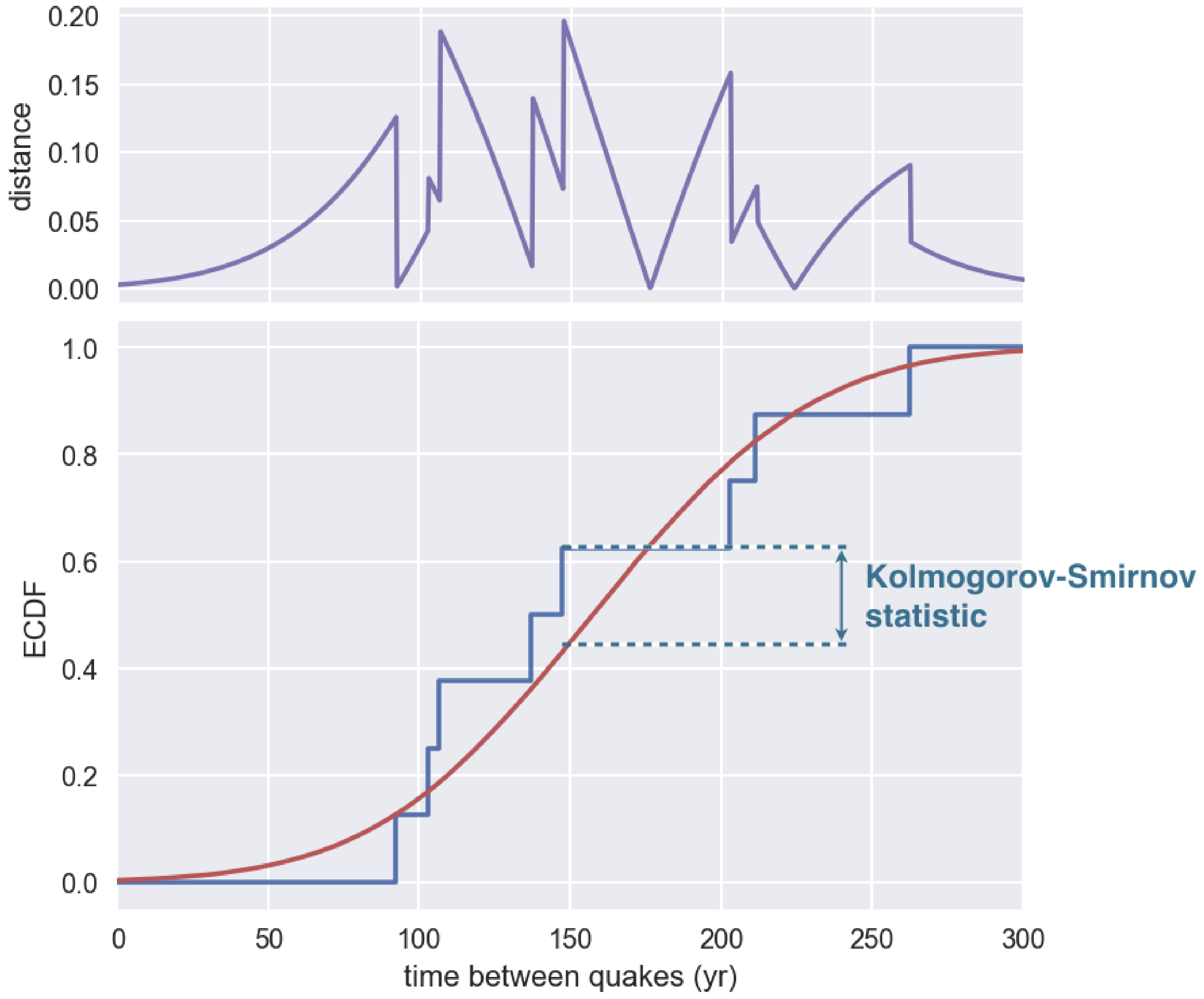
The Kolmogorov-Smirnov statistic



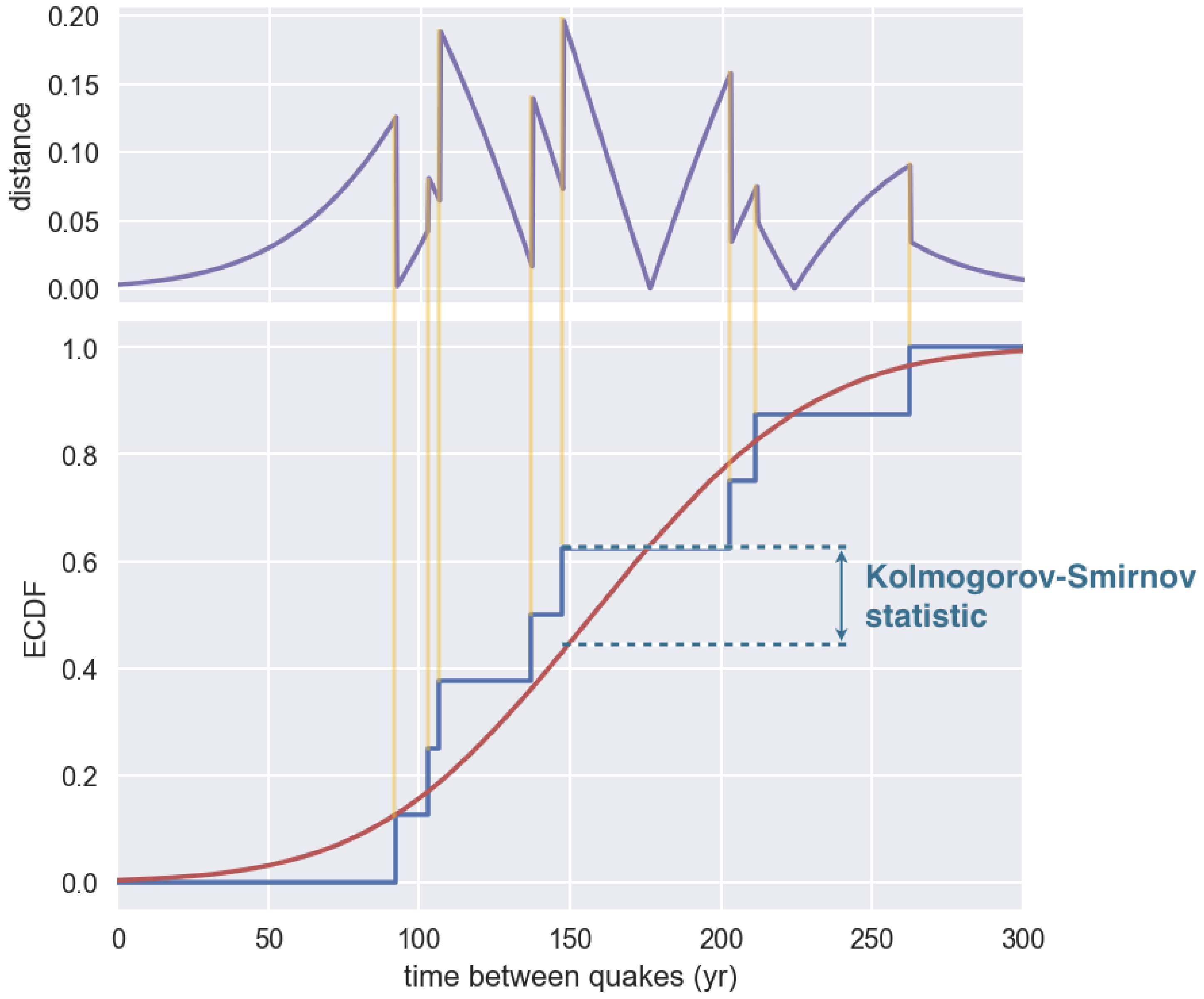
The Kolmogorov-Smirnov statistic



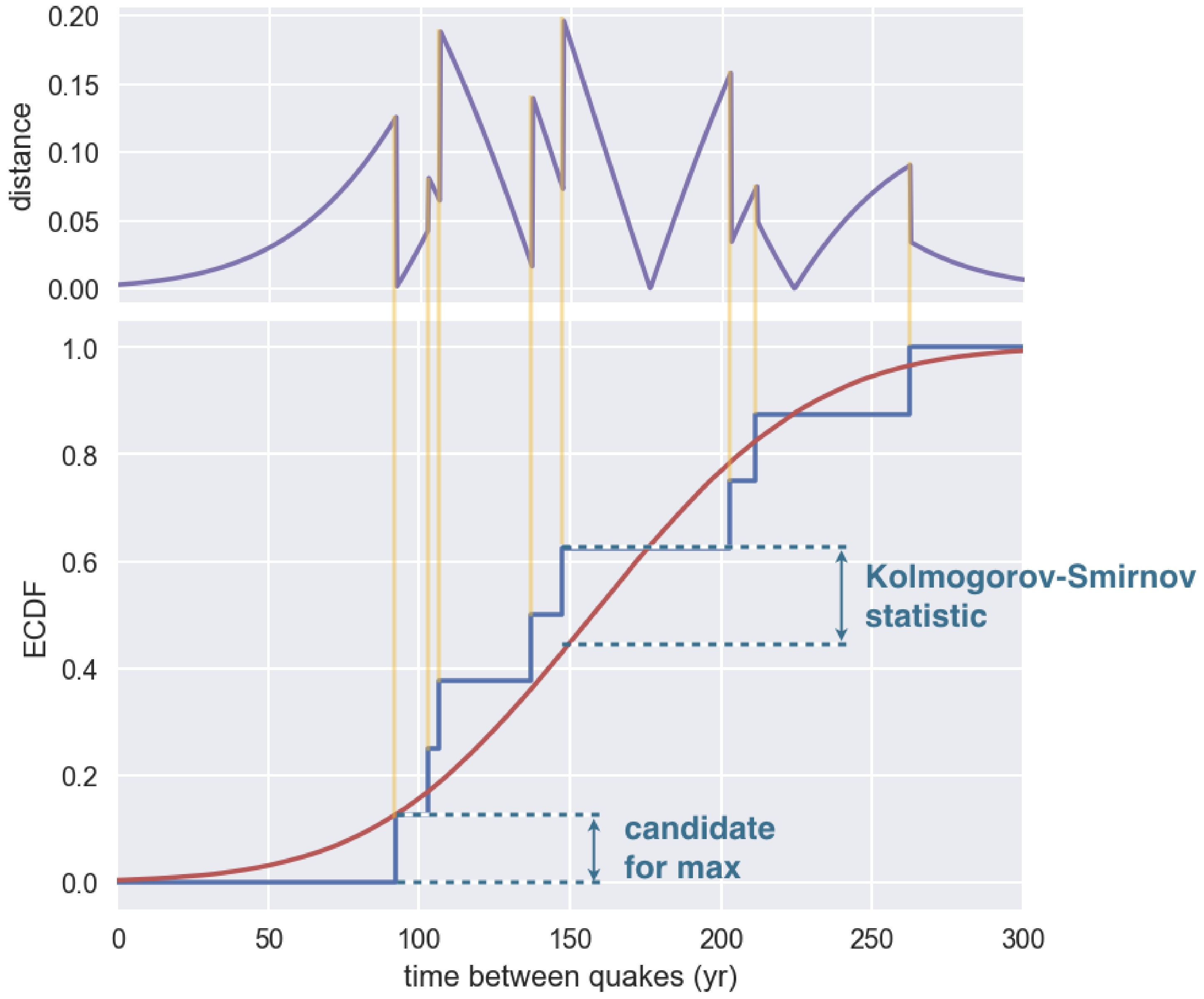
The Kolmogorov-Smirnov statistic



The Kolmogorov-Smirnov statistic



The Kolmogorov-Smirnov statistic



Kolmogorov-Smirnov test

- **Hypothesis:** The time between Nankai Trough earthquakes is Normally distributed with a mean and standard deviation as calculated from the data

- **Test statistic:** Kolmogorov-Smirnov statistic

- **At least as extreme as:** $\geq$ observed K-S statistic

Simulating the null hypothesis

- Draw and store lots of (say, 10,000) samples out of the theoretical distribution

- Draw *n* samples out of the theoretical distribution

- Compute the K-S statistic from the samples

```python
# Generate samples from theoretical distribution
x_f = np.random.normal(mean_time_gap, std_time_gap, size=10000)

# Initialize K-S replicates
reps = np.empty(1000)

# Draw replicates
for i in range(1000):
    # Draw samples for comparison
    x_samp = np.random.normal(
          mean_time_gap, std_time_gap, size=len(time_gap)
    )

    # Compute K-S statistic
    reps[i] = ks_stat(x_samp, x_f)

# Compute p-value
p_val = np.sum(reps >= ks_stat(time_gap, x_f)) / 1000
```

Let's practice!

CASE STUDIES IN STATISTICAL THINKING