Extreme Value Theory for Influenza Peak Prediction

Problem: Predicting Seasonal Flu Peaks

Research Question: Can we predict the magnitude of seasonal influenza peaks across regions using extreme value theory?

Challenge:

- Seasonal flu peaks vary widely (2-13% ILI)
- Rare extreme seasons (e.g., 2017-18)

Data: CDC FluView ILINet

- 14 seasons (2010-2024), 10 HHS regions
- 4,620 weekly observations
- 140 seasonal peak measurements

Approach: Extreme Value Theory

Why EVT? Specifically designed to model rare extreme events

Generalized Extreme Value (GEV) Distribution:

$$F(x) = \exp\left\{-\left[1+\xi\frac{x-\mu}{\sigma}\right]^{-1/\xi}\right\}$$

- μ: location parameter (center of distribution)
- $\sigma > 0$: scale parameter (spread)
- ξ: shape parameter (tail behavior)
  - $\xi > 0$: Heavy tail (rare extreme events possible)
  - $\xi = 0$: Exponential tail (Gumbel distribution)
  - $\xi < 0$: Bounded tail (upper limit exists)

Application: Model distribution of seasonal peak maxima
Training: 80/20 split (112 train, 28 test region-seasons)

Baseline: SIR Epidemiological Model

Susceptible-Infected-Recovered (SIR) Model:

$$\frac{dS}{dt} = -\beta \frac{SI}{N}, \quad \frac{dI}{dt} = \beta \frac{SI}{N} - \gamma I, \quad \frac{dR}{dt} = \gamma I$$

- S: susceptible population
- I: infected population
- R: recovered population
- β: transmission rate
- γ: recovery rate
- $R_0 = \beta/\gamma$: basic reproduction number

Why SIR as baseline?

- Standard mechanistic model for epidemic dynamics
- Predicts peak infected counts based on transmission dynamics
- Tests whether mechanistic modeling can compete with statistical approach

Evaluation Methods

Train/Test Split:

- Training: 112 region-season peaks (first 80%)
- Testing: 28 region-season peaks (last 20%)

Model Fitting:

- GEV: Maximum likelihood estimation on training peaks
- SIR: Fitted separately for each test season's weekly data

Prediction Task:

- GEV: Predict test peaks using 2-year return level
- SIR: Predict peak from fitted epidemic curve

Performance Metrics:

- MAE: Mean Absolute Error
- RMSE: Root Mean Square Error
- MAPE: Mean Absolute Percentage Error

Results

Model	MAE	RMSE	MAPE (%)
GEV	**2.19**	**2.72**	**29.4**
SIR	6.28	7.09	95.3

Fitted GEV Parameters:

- $\mu = 3.70$, $\sigma = 1.92$, $\xi = 0.045$
- Goodness-of-fit: KS test $p = 0.888$

Return Level Estimates: 10-year: 8.2% ILI — 100-year: 13.5% ILI

2017-18 observed max (13.4% ILI) aligns with 100-year estimate