On the Probability Distribution of the Durations of Heatwaves

Sohini Raha and Sujit Ghosh

Department of Statistics, North Carolina State University

The Problem and Objective

The Problem:

- There are hundreds of definitions for heatwaves based on temperature and humidity measures
- BUT...almost all are at best ad hoc (i.e., not based on probabilistic framework)

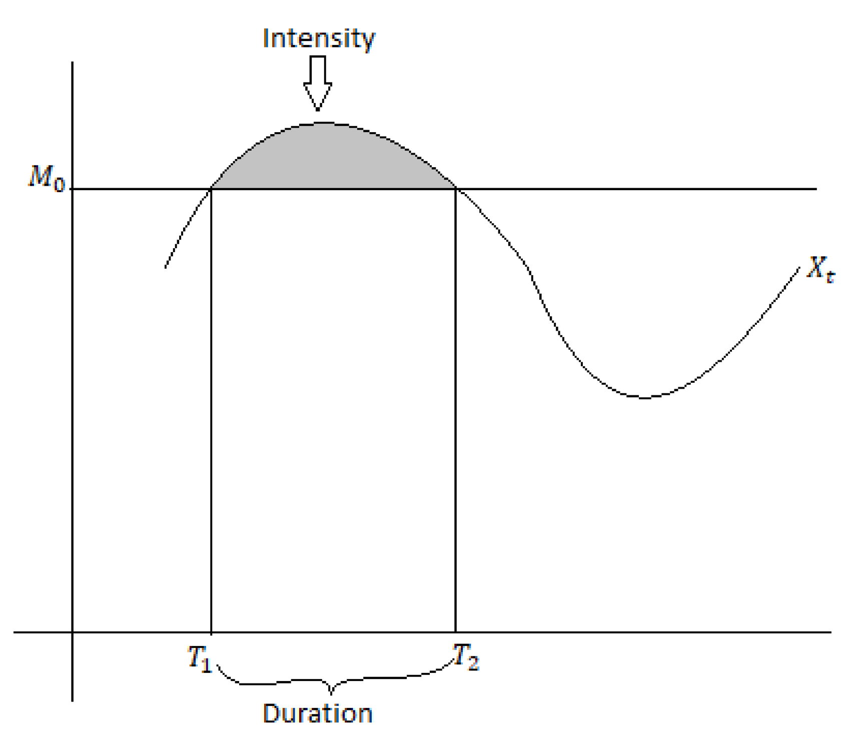
Objective: Build a definition of Heatwave based on a probabilistic inferential frameowrk.

What are Heatwaves?

Some examples of existing definitions of heatwaves (out of >90 available in literature):

- **Frich et al. 2002:** *At least 5 consecutive days, the maximum temperature exceeds the normal temperature by 5°C (9°F) based on a period of 1961-1990 .*
- **Heat Index (HI):** A nonlinear function of *Relative Humidity* and *Temperature*.
- *Caution* if HI is 80-90°F
  Extreme Caution if HI is 90-103°F
  Danger if HI is 103-125°F
  Extreme Danger if HI over 125°F
- Xu et al.(2018) evaluated 29 different definitions of Heatwave
- Vaidyanathan et al.(2016) explored 92 different definitions

Figure 1: Duration and Intensity



Probabilistic Framework

$X_1, X_2, ...$: A strcitly stationary series
M : Threshold (location specific)
$T_1, T_3, ...$: Times of the up-crossings
$T_2, T_4, ...$: Times of the down-crossings

- **Duration** : $D_k = T_{2k} - T_{2k-1}$, k= 1,2,...
- **Intensity** : $X_{T_{2k-1}} + ... + X_{T_{2k}-1} - M(T_{2k} - T_{2k-1})$

Distribution of Duration (Approx)

Let $B_t = \mathbb{I}(X_t > M)$. Under a set of regularity conditions [1], the distribution of D_k can be approximated by

$$D_k \approx Geo(e^{-P(B_1=1)})$$

Distribution of Duration (Exact)

Assume that $X_1, X_2, ...$ follows $AR_p(\theta)$, p-th order stationary AR process with parameter θ. Then

$$P(D_k = l) = \begin{cases} \pi_l(\theta) & l < p \\ A(\theta) \times B(\theta)^{l-p} & l \geq p \end{cases}$$

for l= 1,2,...

Hierarchical model

$d_{k,j}(M)$: k-th duration (days) within year j based on threshold M
$d_{k,j}(M) \sim \text{Geo}(p_j(M))$
$p_j(M) \sim \text{Beta}(\alpha(M)\tau(M), (1-\alpha(M))\tau(M))$

Proposed Definition of Heatwave

Expected length of duration for threshold M is given by

$$E(D_k(M)) = \frac{\tau(M) - 1}{(1-\alpha(M))\tau(M) - 1}$$

For each threshold, if the duration of any up-crossing exceeds the estimated expected length of duration for that threshold, we will call it a Heatwave.

Case Studies

- **Atlanta Data**: 22 years (1991-2012) of meteorology data (daily maximum temperature)

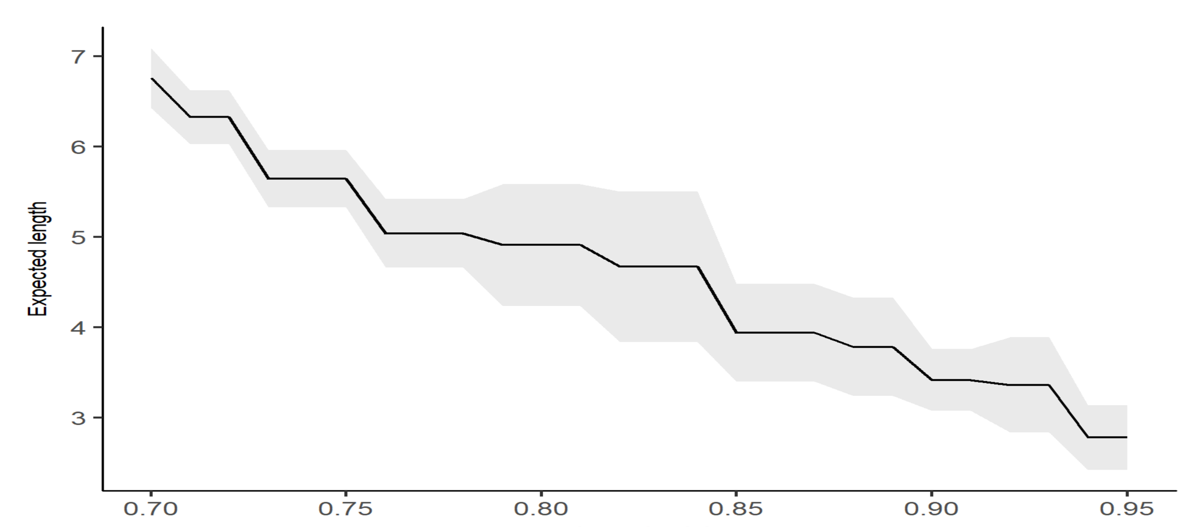


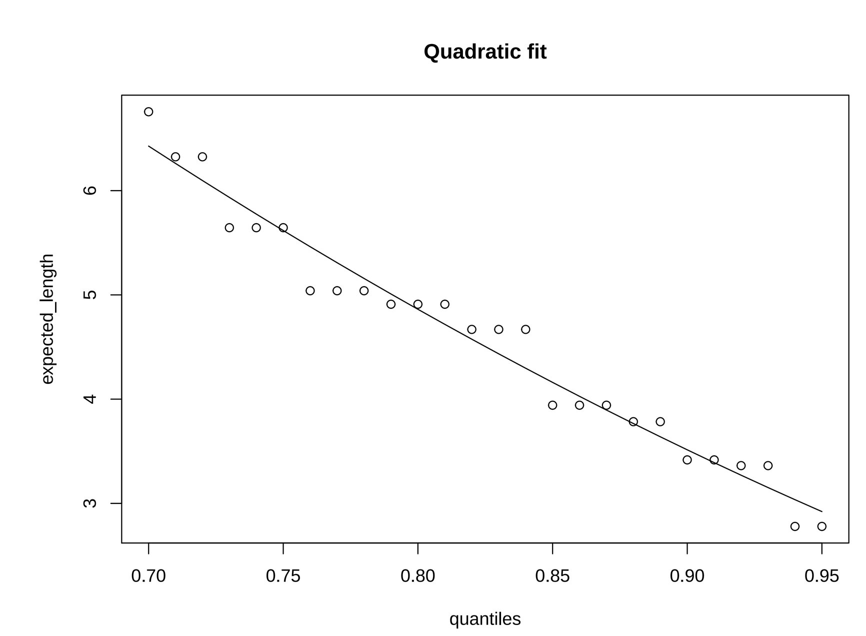
Figure 2: Expected lengths in Atlanta for quantiles (0.7 to 0.95)

- A quadratic curve provides a reasonable approximation to the above curve.

Atlanta Data Analysis

- Expected length = 23.48 -31.96 q + 10.87 q^2 (where $q = P(X_t \leq M) \in (0.70, 0.95)$)

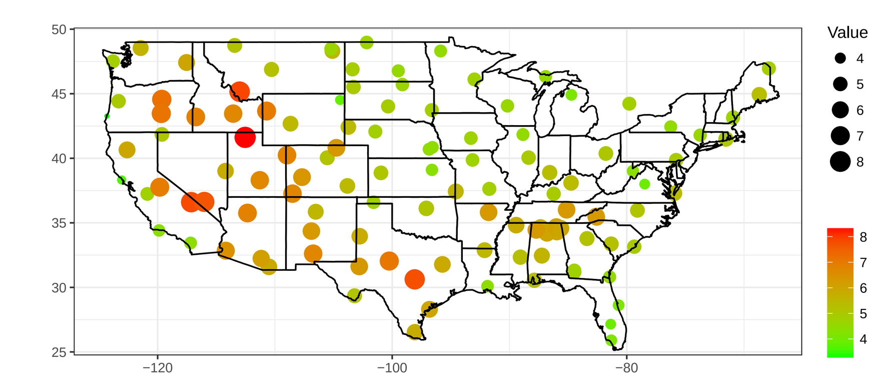
Figure 3: Quadratic fit to the expected lengths with Adjusted $R^2 = 0.96$



USCRN Data Analysis

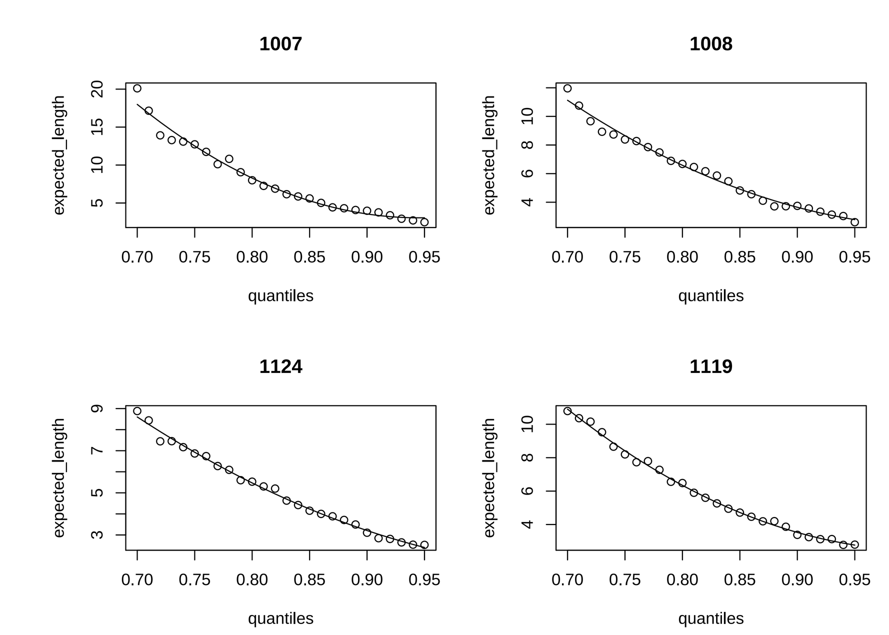
- **USCRN Data** 18 years (2000-2017) of meteorology data (daily average temperature) across 120 weather stations in USA and around.

Figure 4: Expected lengths at USCRN stations at 81 percentile



- Then we fit quadratic equations to the expected lengths with quantiles and for all the stations, the adjusted $R^2 > 0.92$.

Figure 5: Quadratic fit to four of the 120 stations



References

[1] Chen, L. H., & Rollin, A. (2013). Approximating dependent rare events. *Bernoulli*, 19(4), 1243-1267.

Acknowledgements

Thanks to Dr. Howard Chang and Dr. Jesse Bell for providing us with the dataset.