Computing Joint Density Functions of Random Variable with Spatial Dependencies

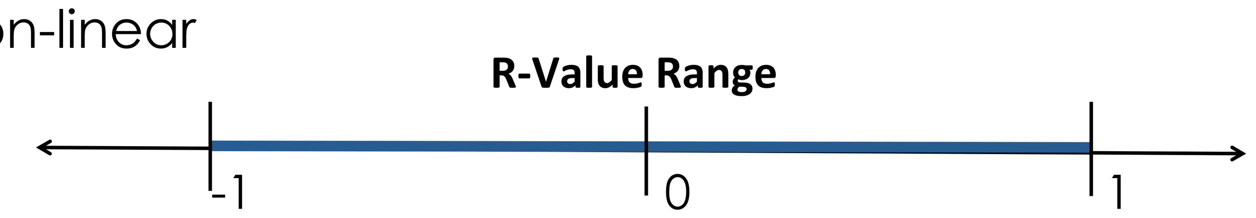
Alli Behr[1] • Advised by Isaac Kumar Isukapati, PhD
[1]Department of Physics and Astronomy, University of Pittsburgh and Department of Mathematics, University of Pittsburgh

Introduction

- The dependence structure among random variables (RVs) plays an instrumental role on how we compute joint density functions. Studying these underlying dependencies has gained popularity in finance, insurance, hydrology and medical studies. The boundary conditions for these dependencies can be classified as:
  - **Comonotonic** {if RVs have strong positive dependencies
  - **Countermonotonic** {if RVs have strong negative dependencies}
  - **Independent** {if RVs follow iid}
- Existing statistical methods provide guidance on how to compute joint for the boundary conditions. In this research we show the effectiveness of computational statistical methods in synthesizing joint density functions.
  - Synthesizing travel time density functions in urban transportation networks is used as a specific example.

Methodology

- A network of 43 sensors in Pittsburgh was used to record trip times of automobiles, whose individual segment times had correlations that ranged from completely negatively correlated, to completely positively correlated.
- Statistical methods allow one to find the joint density function of two random variables if the random variables have a correlation coefficient, or R value, of 0 (uncorrelated), 1 (positively correlated), or -1 (negatively correlated). However, R-values of the times of two links on a route showed correlation coefficients ranging from -1 to 1, and were very rarely exactly -1, 0, or 1.
- It is also important to note that the correlations can be non-linear



- The boundary conditions (R-values of 1, 0, and -1) can be used to create a composite CDF (Cumulative Density Function).
- **Boundary Conditions**: The following classifications can be made of our system, assuming respective correlations:
  - If the R-value of two links is **1,** the pair can be classified as **Comonotonic**, and hence the percentiles of two CDFs can be added to find the composite CDF.
  - If the R-value of two links is **0,** the system can be classified as **Independent**, and hence a composite CDF can be found using Monte-Carlo sampling (Convolution) of both link CDFs.
  - If the R-value of the two links is **-1,** the system can be classified as **Countermonotonic**. Finding a way to combine two CDFs of a Countermonotonic system was an aim of this project.

Methodology Continued

- Three CDFs are created assuming each boundary condition, and then sampled using proportionality constants α, β, and γ to create the composite CDF.

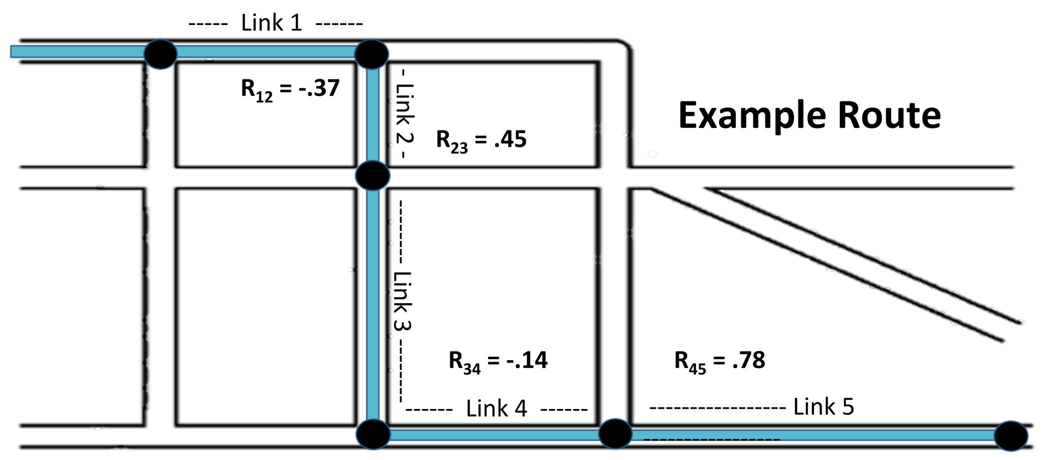


Figure 1: On one route (in blue), link-pairs can have very different R-values, and are rarely -1, 0 or 1. Therefore, traditional stochastic methods cannot be used

- It had already been discovered that for a road segment with no negatively correlated links, such as stretch of highway, the composite CDF of the Comontonic Boundary Condition and the Uncorrelated Boundary Condition produces a CDF in which around 80% of the percentiles are within 5% of the actual trip time CDF.
- In order to model all scenarios, a technique for generating composite CDFs for distributions with negative correlations had to be developed.
- To do this, we combine two link CDFs by adding the CDFs according to the relation:

$$(X_1,\ldots, X_N)= d(F_{x1}^{-1}(U), F_{x2}^{-1}(1-U), F_{x3}^{-1}(U),\ldots, F_{xN}^{-1}(V_N))$$

- After all three boundary condition CDFs were created, the composite CDF was created by combining n_i random percentiles from each CDF using constants α, β, and γ.
  - $n_1 = \alpha * N$ (Independent CDF)
  - $n_2 = \beta * N$ (Comonotonic CDF)
  - $n_3 = (1 - \alpha - \beta) * N$ (Countermonotonic CDF)
- Clustering of the data by the R-values of the link-pairs was used to analyze the effect of this method.
  - This allowed us to find more trips of similar dependencies than simply grouping by time, as the R-values can change very quickly with different conditions throughout the day.

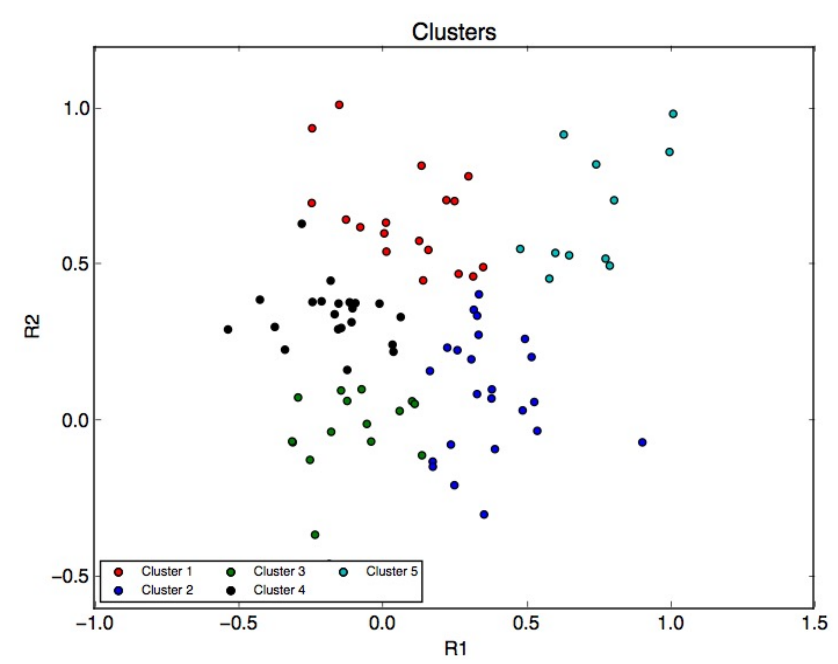


Figure 2: This diagram shows the clustering of a three segment trip on Penn Avenue. Each dot represents a group of 20 trips organized by time of day, with the R-values given by the X and Y axes

Results

- Without a method that accounts for negative correlations, we were still able to generate CDFs that display the results relatively well, but not nearly as well as routes that did not have negative correlations present.
- Adding in the boundary condition for negative correlations, and using Monte-Carlo simulation to sample from the boundary conditions CDFs (Countermonotonic, Comonotonic, and Convolution), the percentiles of the composite CDF within 5% of the actual travel time CDF went from around 60% to 80%.

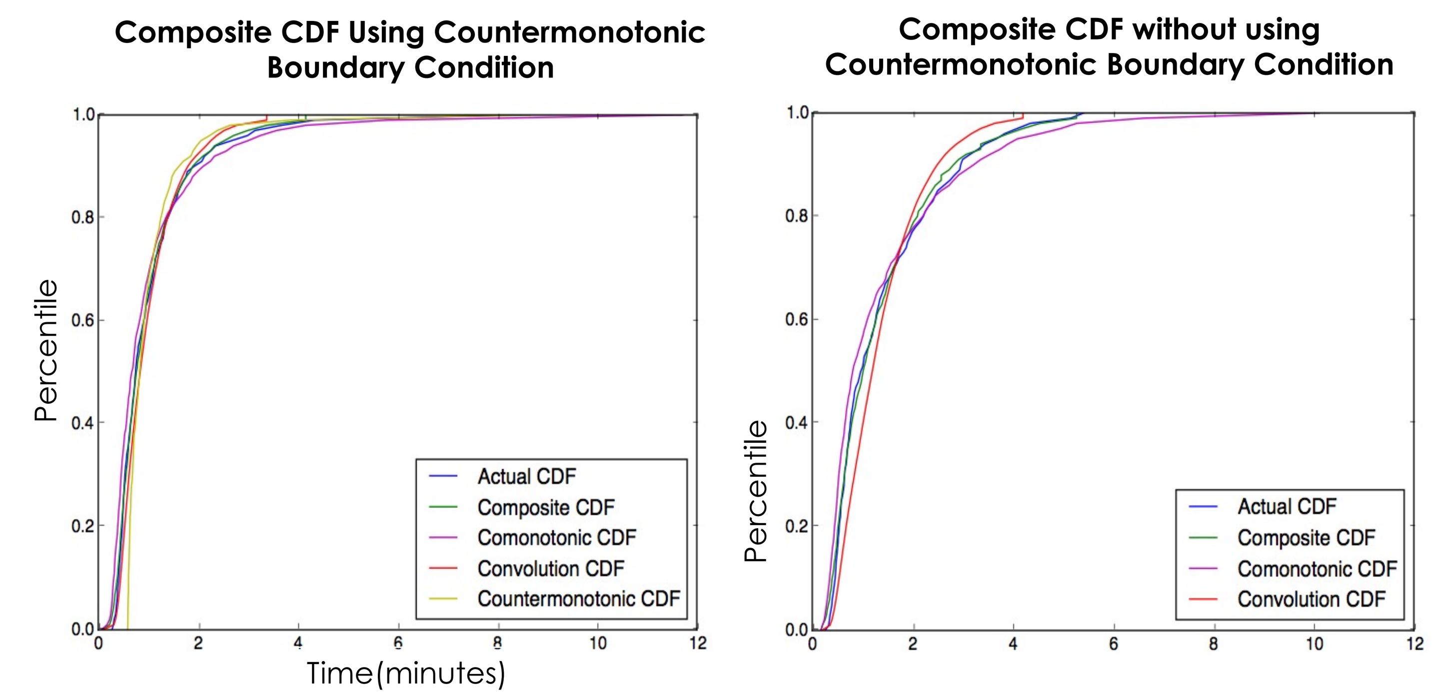


Figure 3) The left figure shows the CDFs formed using the Comontonic Boundary Condition, Convolution Boundary Condition, and Countermonotonic Boundary Condition, as well as the Composite CDFs of all three, and the actual CDF. The right figure shows the same without the Countermonotonic Condition.

Cluster	Composite S-Score using -1 Boundary Condition	Composite S-Score not using -1 Boundary Condition	Comontonic S-Score	Convolution S-Score	Countermonotonic S-Score
1	78	61	28	7	7
2	71	65	6	15	11
3	72	65	10	40	19
4	81	79	17	13	18
5	79	51	10	16	17

Table 1) This table shows the S-Scores, or the number of percentiles within 5 percent of the corresponding Actual percentile, of the composite CDF using Countermonotonic Boundary Condition, the Composite CDF not using the Countermonotonic Condition, the Comonotonic CDF, the Convolution CDF, and the Countermontonic CDF.

Conclusions

- A combination of statistical and computational methods allows accurate prediction of events that are not feasible to predict using either method on its own.
- Assuming Countermonotonicity, boundary conditions can be created for processes that contain negative correlations.

Further Research

- Our next steps will include finding a way to use the α and β conditions to automate the prediction process using real time events.
- A further step to expand our sampling techniques to deal with CDFs that do not follow the traditional mono-modal shape.

Acknowledgements

- Dr. Isaac Iskupati, Carnegie Mellon University Robotics Institute.