Modeling Dependency with Copula: Implications to Engineers and Planners

Haizhong Wang, Ph.D, Assistant Professor
School of Civil and Construction Engineering
Oregon State University, Corvallis, OR
Haizhong.Wang@oregonstate.edu

Seminar at Portland State University
Portland, OR

September 28, 2012
Essentially, all models are wrong, but some are useful.


Deterministic vs. Stochastic

- Greenshield Model
- Greenberg Model
- Underwood Model
- Northwestern Model
- Drew Model

Speed (km/hr) vs. Density (veh/km)

- Greenshields Model
- A Simple Stochastic Model

Speed (km/hr) vs. Density (veh/mile)
Monte Carlo Simulation

How do we measure dependence between random variables?

- Correlation coefficient: a measure of linear dependence between random variables

- Concordance: if “large” values of one random variable tend to be associated with “large” values of the other and “small” values of one with “small” values of the other.

- Discordance: vice versa
Measure of Dependence

- Concordance
- Kendall’s tau
- Spearman’s rho
- Linear correlation: nonelliptical distributions
Dependence of random variables

- Spatial dependencies: the dependence between a number of variables at the same time
- Temporal dependencies: the inter-temporal dependence structure of a process

**The fact:** Covariance only captures the linear dependence relationships for special classes of distributions such as normal distribution

**The question:** Is there a possibility to capture the whole dependence structure without any disturbing effects coming from the marginal distributions?
What is a Copula?

- A Latin noun that means “a link, tie, bond”
- Copulas are used to describe dependence between random variables

Statistics about Copula

- Google 2003: Copula → 10,000 results
- Google 2005: Copula → 650,000 results
- Google 2012: Copula → 1950,000 results

Copulas: Tales and Facts, Thomas Mikosch, 2005. Citation: 140
The first appearance

- 1940/1941: Hoeffding studied nonparametric measures of association such as Spearman’s rho in multivariate distributions

- 1959: The word copula appears for the first time (Sklar, 1959)

- 1999: Introduced to financial applications (Embrechts et al., 1999)

- 2008: Widely adopted in insurance, finance, energy, hydrology, survival analysis, etc.

Source: Daniel Berg, Using Copulas: an Introduction to Practitioners
Definition and Sklar Theorem (1959)

- Sklar theorem describes “join together one-dimensional distribution functions to form multivariate distribution functions”

Let $H$ be a joint distribution function with margins $F_1, \ldots, F_d$. Then there exists a copula $C : [0, 1]^d \rightarrow [0, 1]$ such that

$$H(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d))$$

Theoretically, $C$ captures all aspects of dependence and $F_i$ captures all aspects of marginal distributions
Applications

- Civil engineering - reliability of analysis of highway bridges
- Climate and weather related research
- Analysis of extrema in financial assets and returns
- Failure of paired organs in health science
- Human mortality in insurance (actuarial science)
- Mortalities of spouses
- Mortalities of parents and children twins (identical or nonidentical)
Applications - Choice Modeling


Implications to Planners

Applications - Behaviour Modeling


Applications - Hydrology


Create a copula model for the distribution of \((X_1, \cdots, X_d)\) generally takes two steps

**Model**
- Set a model for marginal distribution \(F_i\)
- Set a model for copula \(C\)

\(C\) is the cdf of a random vector \((U_1, \cdots, U_d)\) with uniform margins

**Simulation**
- Draw a sample \((U_1, \cdots, U_d) \approx C\)
- Set \((X_1, \cdots, X_d) = (F_1^{-1}(U_1), \cdots, F_d^{-1}(U_d))\)
Gaussian Copula

Figure: Bivariate Gaussian copula with varying parameters

- \( \rho = 0.9 \)
- \( \rho = 0.7 \)
- \( \rho = -0.7 \)
- \( \rho = -0.9 \)
Student t Copula

Figure: Bivariate Student copula with varying parameters

\( \rho = 0.8 \) and \( \nu = 1 \)

\( \rho = 0.8 \) and \( \nu = 3 \)

\( \rho = -0.8 \) and \( \nu = 1 \)

\( \rho = -0.8 \) and \( \nu = 3 \)
Implications to Planners

Example - Entrance Ramp Flow Dependency
Example - Entrance Ramp Flow Dependency

Flow profile of a day

No. of observations

Probability

Flow (mi/hr)

Flow (Vehs/hr)
### Example - Morning Peak Hour

<table>
<thead>
<tr>
<th></th>
<th>101</th>
<th>102</th>
<th>103</th>
<th>104</th>
<th>105</th>
<th>005</th>
<th>006</th>
<th>008</th>
<th>009</th>
<th>010</th>
</tr>
</thead>
<tbody>
<tr>
<td>101</td>
<td>1.0000</td>
<td>0.0503</td>
<td>-0.1529</td>
<td>0.0680</td>
<td>0.0209</td>
<td>-0.0284</td>
<td>0.0767</td>
<td>0.2414</td>
<td>-0.0061</td>
<td>-0.4539</td>
</tr>
<tr>
<td>102</td>
<td>0.0503</td>
<td>1.0000</td>
<td>-0.0108</td>
<td>0.1021</td>
<td>0.1678</td>
<td>0.0901</td>
<td>0.1178</td>
<td>-0.0376</td>
<td>0.1488</td>
<td>-0.0591</td>
</tr>
<tr>
<td>103</td>
<td>-0.1529</td>
<td>-0.0108</td>
<td>1.0000</td>
<td>-0.0548</td>
<td>0.2614</td>
<td>0.0699</td>
<td>0.0104</td>
<td>-0.2420</td>
<td>0.0231</td>
<td>0.1339</td>
</tr>
<tr>
<td>104</td>
<td>0.0680</td>
<td>0.1021</td>
<td>-0.0548</td>
<td>1.0000</td>
<td>0.2458</td>
<td>0.1749</td>
<td>0.1910</td>
<td>0.1267</td>
<td>0.0948</td>
<td>-0.0287</td>
</tr>
<tr>
<td>105</td>
<td>0.0209</td>
<td>0.1678</td>
<td>0.2614</td>
<td>0.2458</td>
<td>1.0000</td>
<td>0.3052</td>
<td>-0.0164</td>
<td>0.0759</td>
<td>0.5157</td>
<td>0.0111</td>
</tr>
<tr>
<td>005</td>
<td>-0.0284</td>
<td>0.0901</td>
<td>0.0699</td>
<td>0.1749</td>
<td>0.3052</td>
<td>1.0000</td>
<td>0.0297</td>
<td>0.1201</td>
<td>0.1346</td>
<td>0.1041</td>
</tr>
<tr>
<td>006</td>
<td>0.0767</td>
<td>0.1178</td>
<td>0.0104</td>
<td>0.1910</td>
<td>-0.0164</td>
<td>0.0297</td>
<td>1.0000</td>
<td>-0.0344</td>
<td>-0.1562</td>
<td>0.0034</td>
</tr>
<tr>
<td>008</td>
<td>0.2414</td>
<td>-0.0376</td>
<td>-0.2420</td>
<td>0.1267</td>
<td>0.0759</td>
<td>0.1201</td>
<td>-0.0344</td>
<td>1.0000</td>
<td>-0.0087</td>
<td>0.0309</td>
</tr>
<tr>
<td>009</td>
<td>-0.0061</td>
<td>0.1488</td>
<td>0.0231</td>
<td>0.0948</td>
<td>0.5157</td>
<td>0.1346</td>
<td>-0.1562</td>
<td>-0.0087</td>
<td>1.0000</td>
<td>-0.0006</td>
</tr>
<tr>
<td>010</td>
<td>-0.4539</td>
<td>-0.0591</td>
<td>0.1339</td>
<td>-0.0287</td>
<td>0.0111</td>
<td>0.1041</td>
<td>0.0034</td>
<td>0.0309</td>
<td>-0.0006</td>
<td>1.0000</td>
</tr>
</tbody>
</table>
## Example - Afternoon Peak Hour

<table>
<thead>
<tr>
<th>4005-</th>
<th>101</th>
<th>102</th>
<th>103</th>
<th>104</th>
<th>105</th>
<th>005</th>
<th>006</th>
<th>008</th>
<th>009</th>
<th>010</th>
</tr>
</thead>
<tbody>
<tr>
<td>101</td>
<td>1.0000</td>
<td>0.4030</td>
<td>0.2727</td>
<td>0.0914</td>
<td>0.0827</td>
<td>0.3679</td>
<td>0.3815</td>
<td>-0.0474</td>
<td>0.0044</td>
<td>0.0204</td>
</tr>
<tr>
<td>102</td>
<td>0.4030</td>
<td>1.0000</td>
<td>0.4563</td>
<td>0.1567</td>
<td>0.0937</td>
<td>0.5894</td>
<td>0.5233</td>
<td>0.0439</td>
<td>-0.1558</td>
<td>0.4893</td>
</tr>
<tr>
<td>103</td>
<td>0.2727</td>
<td>0.4563</td>
<td>1.0000</td>
<td>0.2309</td>
<td>-0.0606</td>
<td>0.4834</td>
<td>0.3606</td>
<td>-0.0407</td>
<td>-0.2626</td>
<td>0.3998</td>
</tr>
<tr>
<td>104</td>
<td>0.0914</td>
<td>0.1567</td>
<td>0.2309</td>
<td>1.0000</td>
<td>0.0008</td>
<td>0.2485</td>
<td>0.1592</td>
<td>-0.0808</td>
<td>-0.0927</td>
<td>0.1064</td>
</tr>
<tr>
<td>105</td>
<td>0.0827</td>
<td>0.0937</td>
<td>-0.0606</td>
<td>0.0008</td>
<td>1.0000</td>
<td>0.0024</td>
<td>0.1417</td>
<td>-0.1339</td>
<td>0.2076</td>
<td>-0.0106</td>
</tr>
<tr>
<td>005</td>
<td>0.3679</td>
<td>0.5894</td>
<td>0.4834</td>
<td>0.2485</td>
<td>0.0024</td>
<td>1.0000</td>
<td>0.5952</td>
<td>0.0249</td>
<td>-0.2089</td>
<td>0.5994</td>
</tr>
<tr>
<td>006</td>
<td>0.3815</td>
<td>0.5233</td>
<td>0.3606</td>
<td>0.1592</td>
<td>0.1417</td>
<td>0.5952</td>
<td>1.0000</td>
<td>-0.0083</td>
<td>0.0249</td>
<td>0.5276</td>
</tr>
<tr>
<td>008</td>
<td>-0.0474</td>
<td>0.0439</td>
<td>-0.0407</td>
<td>-0.0808</td>
<td>-0.1339</td>
<td>0.0249</td>
<td>-0.0083</td>
<td>1.0000</td>
<td>-0.0616</td>
<td>0.1039</td>
</tr>
<tr>
<td>009</td>
<td>0.0044</td>
<td>-0.1558</td>
<td>-0.2626</td>
<td>-0.0927</td>
<td>0.2076</td>
<td>-0.2089</td>
<td>0.0249</td>
<td>-0.0616</td>
<td>1.0000</td>
<td>-0.0787</td>
</tr>
<tr>
<td>010</td>
<td>0.0204</td>
<td>0.4893</td>
<td>0.3998</td>
<td>0.1064</td>
<td>-0.0106</td>
<td>0.5994</td>
<td>0.5276</td>
<td>0.1039</td>
<td>-0.0787</td>
<td>1.0000</td>
</tr>
</tbody>
</table>
Example - Entrance Ramp Flow Dependency

Figure: The joint probability density contour through a Gaussian copulas for the morning peak hour (01/02/2003) dependency among entrance-ramps southbound of GA400.
Example - Entrance Ramp Flow Dependency

Figure: The morning peak (01/02/2003) dependency surface through nonparametric bivariate copulas for entrance-ramps southbound of GA400, including 4005101 to 4005105.
Example - Day-to-Day Analysis

(a) 4005101/102
(b) 4005102/103
(c) 4005103/104
(d) 4005104/105

Figure: The joint probability density contour through a 2d student t copulas for the morning peak hour (01/07/2003) dependency among entrance-ramps southbound of GA400.
Student-t Copula

Figure: The joint probability density contour through a 2d student t copulas for the morning peak hour (01/07/2003) dependency among entrance-ramps southbound of GA400.
Result Analysis - Student t copula

Figure: The joint probability density contour through a 2d student t copulas for the afternoon peak hour (01/09/2003) dependency among entrance-ramps northbound of GA400.
Figure: The dependency structure between ramp flow in 2 and 3 dimensions
Computing/Fitting with Copula

- Matlab - Built-in-functions
- R - Copula package
The copula contains all the information about the dependence between random variables.

Copulas provide an alternative and often more useful representation of multivariate distribution functions compared to traditional approaches such as multivariate normality.

Most traditional representations of dependence are based on the linear correlation coefficient - restricted to multivariate elliptical distributions. Copula representations of dependence are free of such limitations.

Copulas enable us to model marginal distributions and the dependence structure separately.

Copulas provide greater modeling flexibility, given a copula we can obtain many multivariate distributions by selecting different margins.

A copula is invariant under strictly increasing transformations.

Most traditional measures of dependence are measures of pairwise dependence. Copulas measure the dependence between all $d$ random variables.
Questions and Comments?

Thanks!

Jia Li at University of California Davis