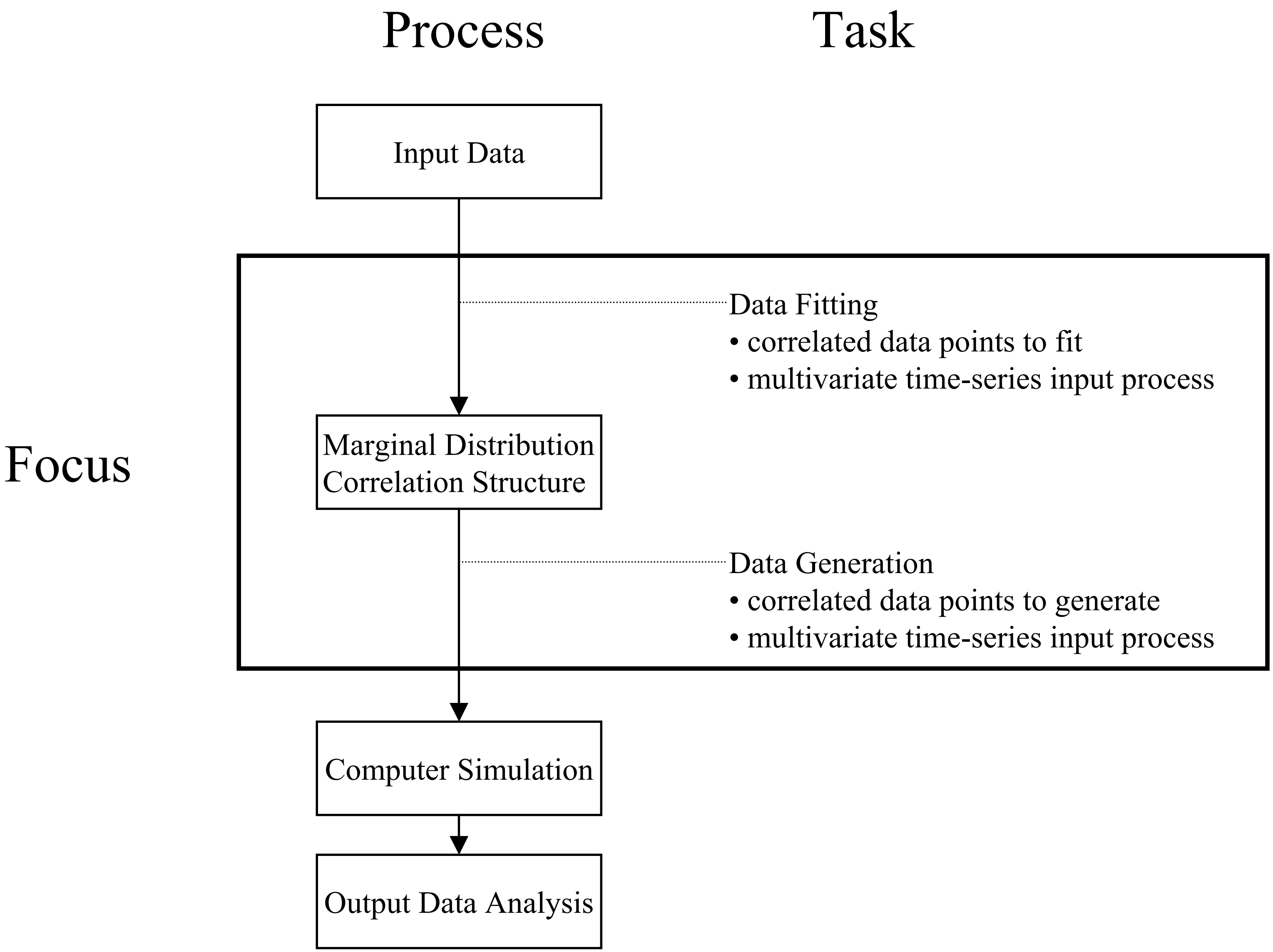
Towards a Comprehensive Stochastic Input-Modeling Framework for Continuous Distributions*

Bahar Deler & Barry L. Nelson
Northwestern University

July 26, 2001

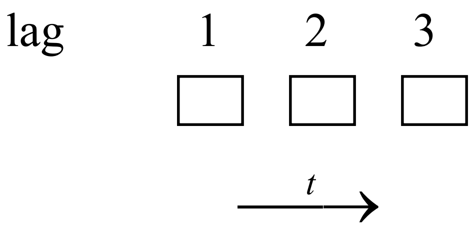
*Research supported by NSF grant DMI 9821011

Stochastic Input Modeling



Literature Review - Data Generation

Univariate Time Series

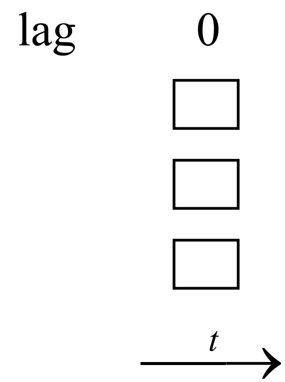


<u>Approaches:</u>

– Exploit properties specific to the marginal distribution (Lewis et al. 1989)

– Construct a series of autocorrelated uniform random variables as the base process and apply the inverse transform method (Melamed 1991, Cario and Nelson 1996)

Random Vectors

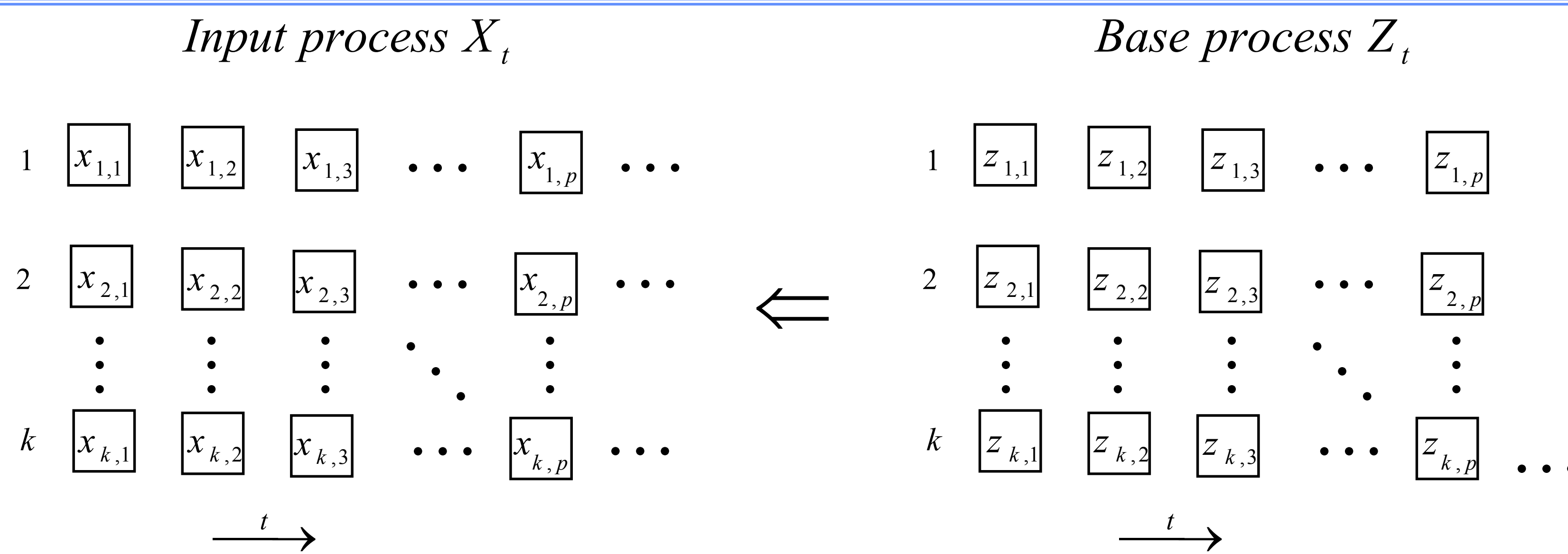


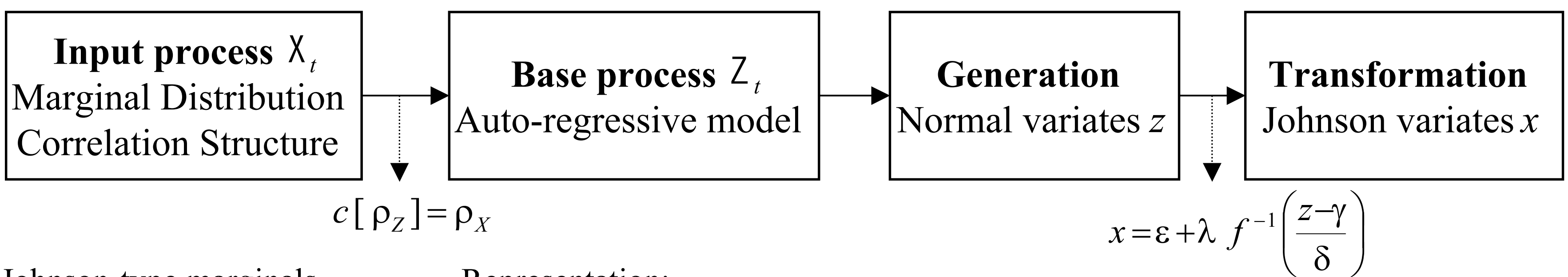
<u>Approach:</u>

– Transform multivariate copula into vectors with arbitrary marginal distributions (Cario et al. 2000, Chen 2000, Ghosh and Henderson 2001)

Develop a comprehensive framework that integrates the theory of univariate time series and random vectors with dependent components and extend it to multivariate time series

Data Generation Model: Overview





- Johnson-type marginals

$$F_X(x) = \Phi\left\{\gamma + \delta f\left(\frac{x-\varepsilon}{\lambda}\right)\right\}$$

- Product-moment correlations

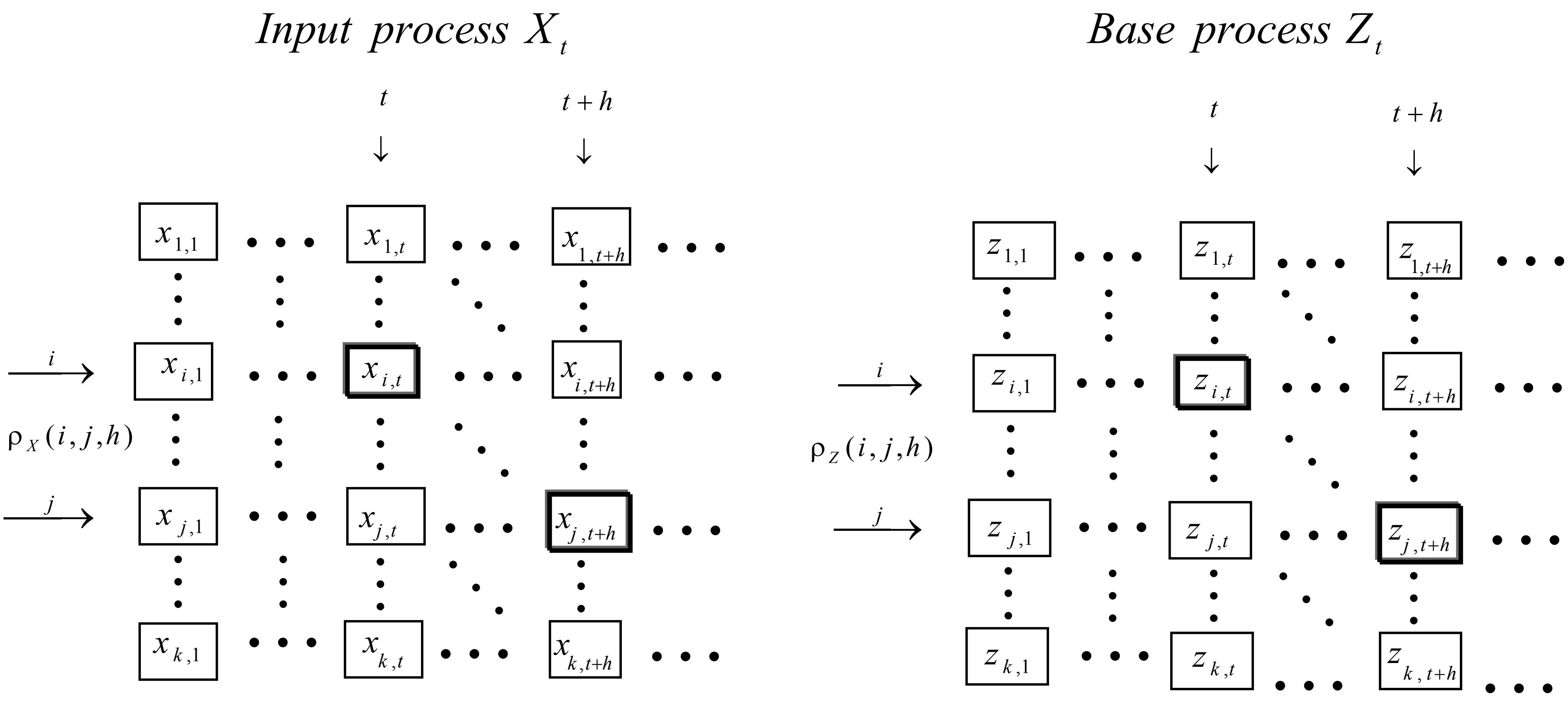
- Representation:

$$Z_t = \sum_{h=1}^{p} \boldsymbol{\alpha}_h Z_{t-h} + u_t$$

- Assumptions:
  - Stationarity of Z_t
  - Gaussian white noise u_t

Correlation Matching Problem

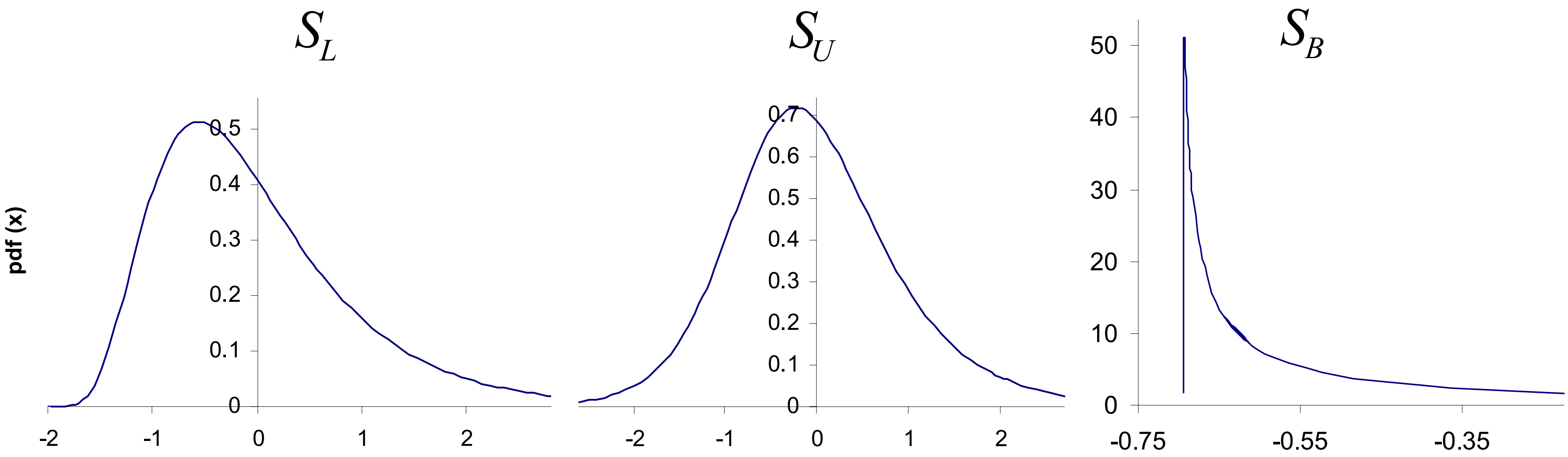
<u>Objective</u>: Select a correlation structure for the base process that gives the desired correlation structure for the input process



$$c_{ijh}\left[\rho_Z(i,j,h)\right] = \rho_X(i,j,h)$$ ($pk^2 + k(k-1)/2$ **independent matching problems)**

Example

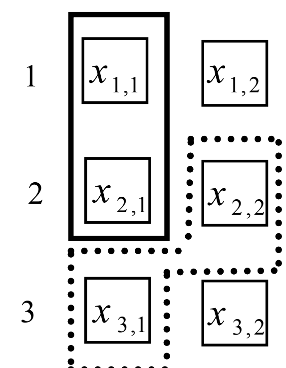
<u>Step 0</u> A trivariate input process with marginals and correlations given as



$$\gamma_1 = -1.462 \quad \delta_1 = 2.236 \qquad \gamma_2 = -0.730 \quad \delta_2 = 1.905 \qquad \gamma_3 = 1.258 \quad \delta_3 = 0.426$$

$$\varepsilon_1 = -2.125 \quad \lambda_1 = 1.000 \qquad \varepsilon_2 = -0.686 \quad \lambda_2 = 1.521 \qquad \varepsilon_3 = -0.694 \quad \lambda_3 = 4.421$$

$$\sum_X(0) = \begin{pmatrix} 1.000 & \boxed{0.364} & 0.408 \\ 0.364 & 1.000 & 0.257 \\ 0.408 & 0.257 & 1.000 \end{pmatrix} \text{ and } \sum_X(1) = \begin{pmatrix} 0.287 & 0.232 & 0.104 \\ 0.129 & 0.281 & 0.289 \\ 0.117 & 0.259 & 0.169 \end{pmatrix}$$



Example (cont'd)

Step 1

– Characterize the underlying process by solving 12 correlation matching problems

$$\sum_Z(0) = \begin{pmatrix} 1.000 & 0.377 & 0.460 \\ 0.377 & 1.000 & 0.299 \\ 0.460 & 0.299 & 1.000 \end{pmatrix} \quad \text{and} \quad \sum_Z(1) = \begin{pmatrix} 0.307 & 0.244 & 0.124 \\ 0.137 & 0.286 & 0.336 \\ 0.141 & 0.302 & 0.219 \end{pmatrix}$$

– Using the correlation structure, determine the system parameters for

$$\mathsf{Z}_t = \boldsymbol{\alpha}_1 \mathsf{Z}_{t-h} + \mathsf{u}_t$$

$$\boldsymbol{\alpha}_1 = \begin{pmatrix} 0.269 & 0.157 & -0.046 \\ -0.091 & 0.228 & 0.309 \\ -0.029 & 0.267 & 0.152 \end{pmatrix} \quad \text{and} \quad \sum_u = \begin{pmatrix} 0.885 & 0.310 & 0.385 \\ 0.310 & 0.843 & 0.175 \\ 0.385 & 0.175 & 0.890 \end{pmatrix}$$

– Simulate the vector autoregressive model to obtain $z_{i,t}$

Step 2

– Using $z_{i,t}$, generate Johnson variates $x_{i,t} = \varepsilon_i + \lambda_i f_i^{-1}\left(\dfrac{z_{i,t} - \gamma_i}{\delta_i}\right)$

Data Fitting Model: Setup

Given a sample from a stationary k-variate time-series input process X_t there are

- k marginals to fit by estimating
  - the families of the Johnson marginals, f_i, $i = 1,2,\ldots,k$
  - the parameters of the Johnson marginals, $\gamma_i, \delta_i, \lambda_i, \varepsilon_i, i = 1,2,\ldots,k$
- $pk^2 + k(k-1)/2$ correlations to fit by estimating the underlying base process
  - the order of dependence, p
  - the autoregressive coefficient matrices, $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_p$
  - the residual covariance matrix, Σ_u

Work iteratively between improving

• the estimates of the base process parameters, $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_p, \Sigma_u$, and

• the estimates of the parameters of the marginal distributions, $f_i, \gamma_i, \delta_i, \lambda_i, \varepsilon_i, i = 1,2,\ldots,k$

Data Fitting Model: Overview

Procedure for Univariate Case

Stage 0 Obtain initial parameters for the Johnson marginal, $\gamma, \delta, \lambda, \varepsilon$, by solving the least-squares problem suggested by Swain et al. (1988)

Stage 1 Reduce the input data to the base process level via $\gamma + \delta f[(x_t - \varepsilon)/\lambda]$ and find the least-square estimators for $\alpha_1, \ldots, \alpha_p, \sigma$, which imply a stationary autoregressive process

Stage 2 Modify the parameters of the Johnson marginal by solving a least- squares fitting problem that ensures the independence and normality of the residuals of the base process

Extensions

- Order of dependence, p
- Type of the Johnson marginal, f

Summary & Future Research

Summary

- Represented and generated stationary multivariate time series with fully specified correlation matrices and marginals from the Johnson system
- Developed software implementing the data-generation procedure

Ongoing Work

- Develop an automated & statistically valid algorithm to fit input models
  - Numerical testing
  - Analysis of the properties of the estimators produced by the fitting procedure
- Make the software available for incorporating into commercial packages