Pirenópolis EMR 2019
XVI Escola de Modelos de Regressão

24 a 27 de março de 2019
Pousada dos Pireneus
PIRENÓPOLIS/GO

Maynooth University
National University of Ireland Maynooth

Construction and implementation of multivariate dispersion models

Bruna Wundervald [1*], Dennis Leão[2], Wagner Bonat[3]

[1]**Hamilton Institute, Maynooth University** ***e-mail: brunadaviesw@gmail.com**

[2,3] **Federal University of Paraná**

Introduction

- Generalized linear models: usual in statistical modelling;
  $\rightarrow$ *mostly univariate cases.*
- Multivariate frameworks for GLM are still rare.
- Most multivariate techniques are based on the Multivariate Normal distribution;
  - Suitable only for continuous and symmetrical data.
- Statistical models are realistic when can describe the dependency structure, when it exists (Temporal, Spatial, Genetic, Longitudinal, etc.)
- We can be interested in more than one response variable, possibly correlated.

The main goals of this work are to:

- Build probability distributions for multivariate, non-normal random variables and its regressions models;
  - discrete, strong asymmetrical and heavy tailed data.
- Implement the models in R [3].

Methods and Materials

Methods

The generalization of a **dispersion model** is defined as

$$p(y;\mu,\sigma^2) = a(y;\sigma^2)\exp\left\{-\frac{1}{2\sigma^2}d(y;\mu)\right\}, \quad y \in C, \tag{1}$$

where $a \geq 0$ is an adequate function, C is the smallest interval containing the realizable values of y, d is a unit deviance in $C \times \Omega$, $\mu \in \Omega$ and $\sigma^2 \in \Re_+$.

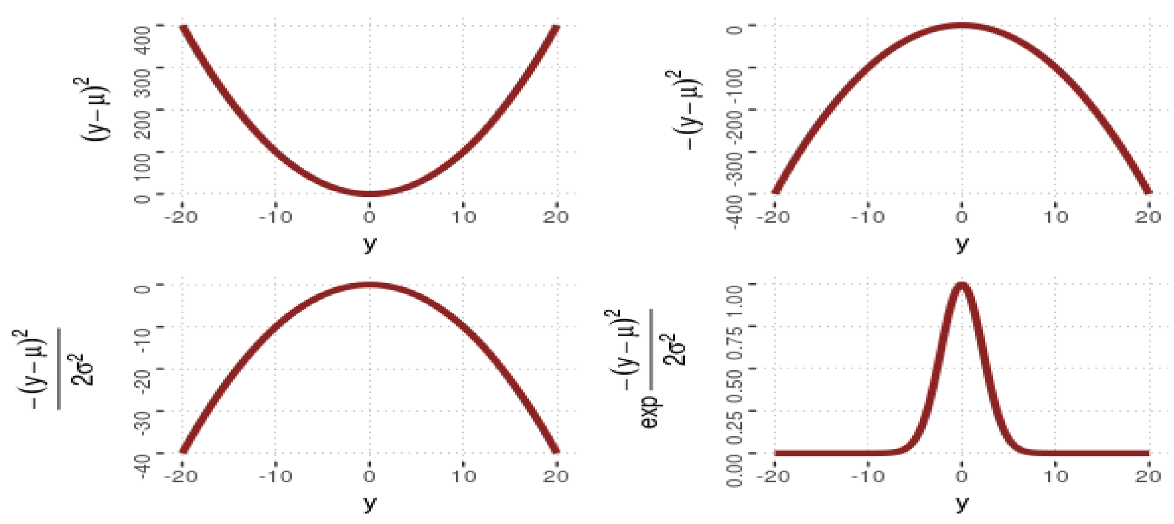

Figure 1: Core of a normal distribution, that can be generalized by a the equation of a dispersion model.

In [1], Jørgensen introduced the theory of the dispersion models, that are based on deviance residuals, a function that satisfies

$$d(y;y) = 0 \quad \forall\ y \in \Omega, \text{ and } d(y;\mu) > 0 \quad \forall\ y \neq \mu. \tag{2}$$

being Ω the parametric space for μ, $\Omega \subseteq \Re$. The multivariate extension of the dispersion model was proposed in [2], and has the form

$$p(\boldsymbol{y};\boldsymbol{\mu},\Sigma) = a(\boldsymbol{y};\Sigma)\exp\left\{-\frac{1}{2}t(\boldsymbol{y};\boldsymbol{\mu})^\top\Sigma^{-1}t(\boldsymbol{y};\boldsymbol{\mu})\right\}, \tag{3}$$

where $\boldsymbol{\mu} \in \Omega$ is a open interval in $\Re^p$, Σ is a positive-definite symmetric matrix $p \times p$, and $t(\boldsymbol{y};\boldsymbol{\mu})$ is a vector of deviance residuals, given by

$$t(\boldsymbol{y};\boldsymbol{\mu}) = sign(\boldsymbol{y}-\boldsymbol{\mu})\sqrt{d(\boldsymbol{y};\boldsymbol{\mu})},$$

and $t(\boldsymbol{\mu};\boldsymbol{\mu}) = \boldsymbol{0}$, for $\boldsymbol{\mu} \in \Omega$.

Results

- Construction of non-normalized distributions and characterizing them.
- Parameter interpretation.

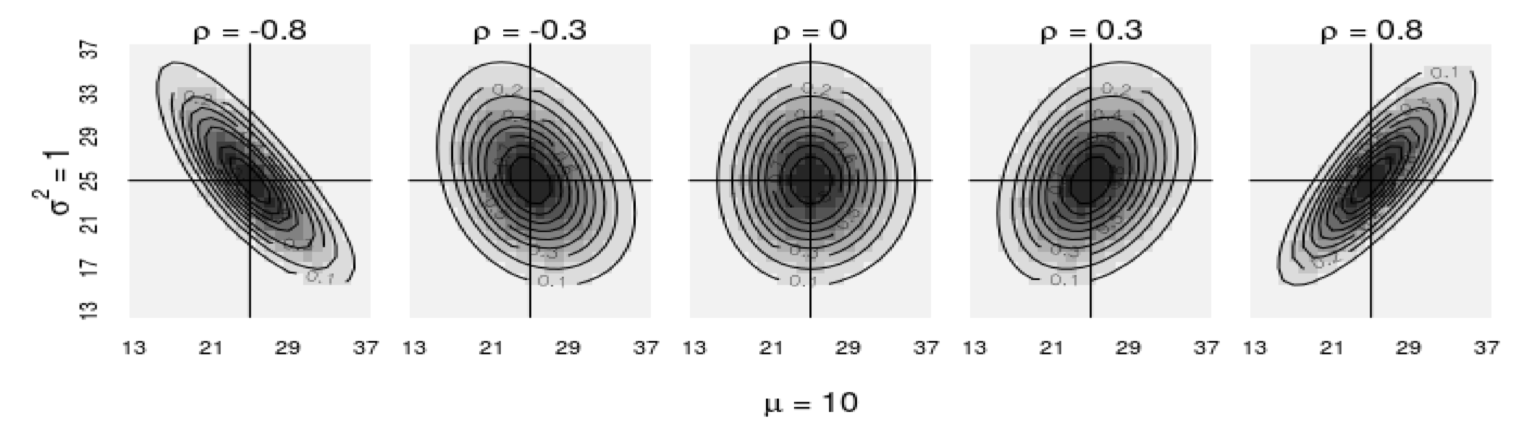

Figure 2: Core of the non-normalized bivariate Binomial distribution.

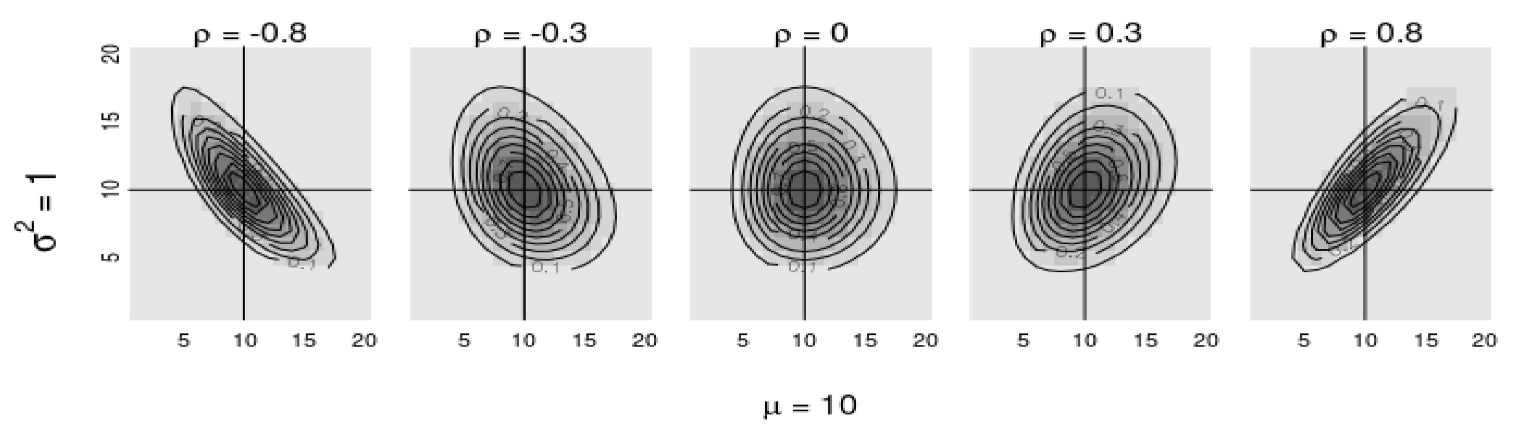

Figure 3: Core of the non-normalized bivariate Poisson distribution

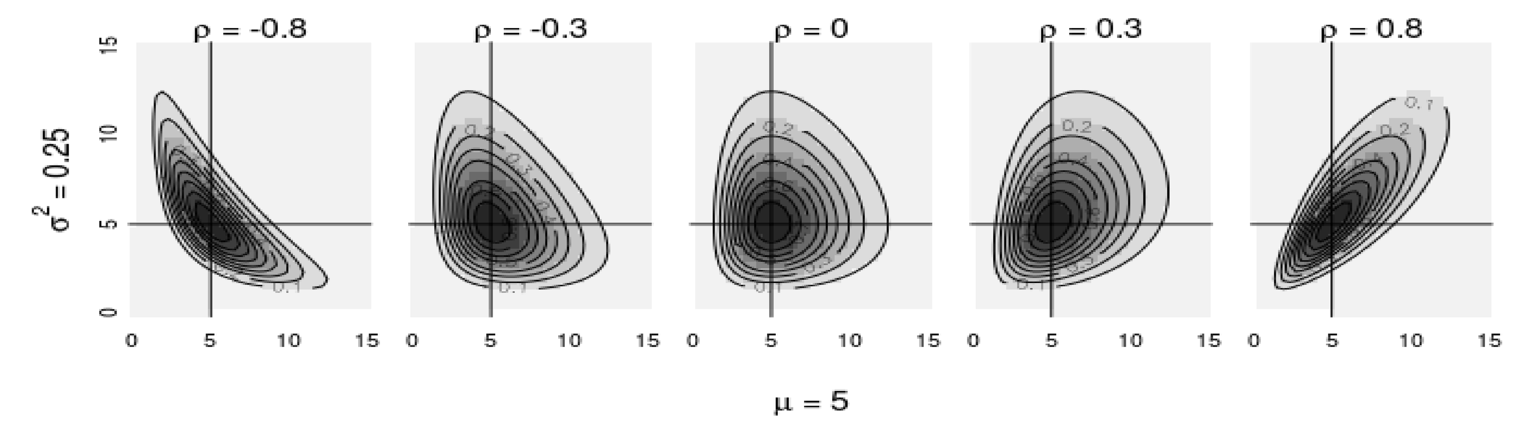

Figure 4: Core of the non-normalized bivariate Gamma distribution.

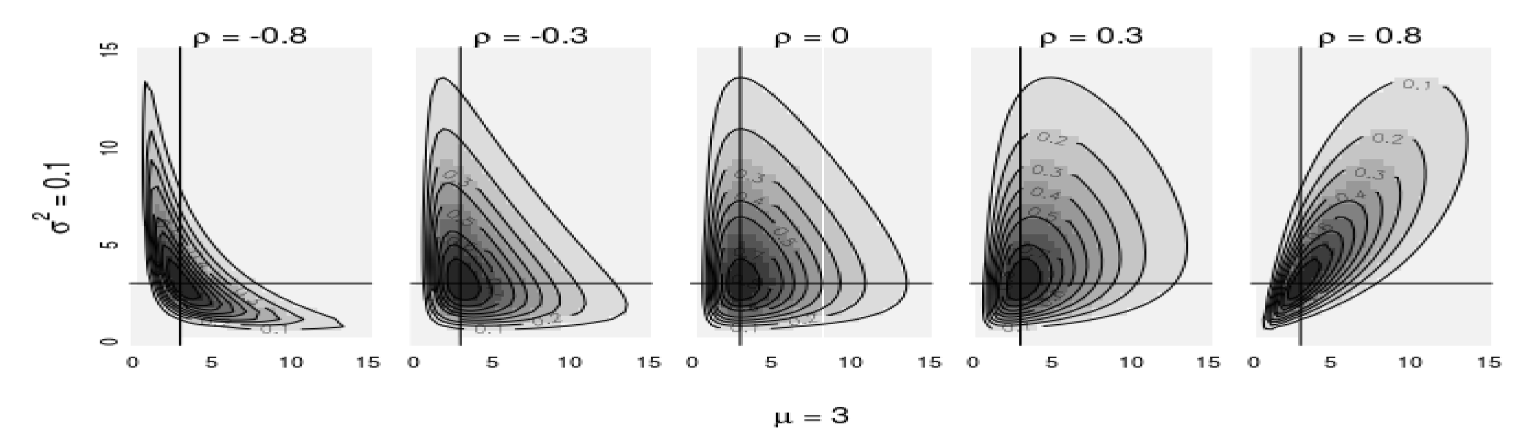

Figure 5: Core of the non-normalized bivariate inverse Normal distribution.
Interpretation

- The parameter ρ controls the correlation.
- The dispersion parameters control the variability and shape.
- Similar to the bivariate normal distribution.
- $\boldsymbol{\mu}$ is not necessarily a vector of expectations: better interpreted as a vector of modes.

Conclusions and discussion

- The method is simple and the interpretation is intuitive.
- Results about the normalizing constants do not influence directly on the construction of regression models for the location parameters.

Future work includes:

- Evaluate the approximations to the normalizing constants.
- Inference and full computational implementation.

Bibliography

[1] Jørgensen, Bent. *The Theory of Dispersion Models*, 1987

[2] Bent Jørgensen and Steffen L. Lauritzen *Multivariate Dispersion Models*.2000.

[3] R Core Team *R: A Language and Environment for Statistical Computing*.2018.