



➤ Sur l'estimation REML des composantes de variance du modèle de génétique animal bivarié en présence de fortes corrélations dans les queues de distributions

Tom Rohmer¹, Anne Ricard² & Ingrid David¹

¹ GenPhySE, ² GABI-IFCE

Plan

Introduction

Simulation

Results



Plan

Introduction

Simulation

Results



INRAE

Copula miss-specification on REML estimation

13-15 sept 2022 / Journées Scientifiques Département de Génétique Animal 2022/Tom Rohmer

➤ Example, Large-White dataset

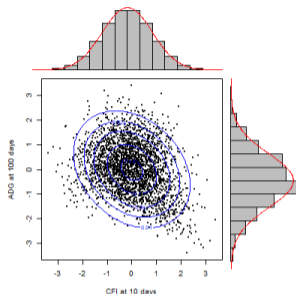
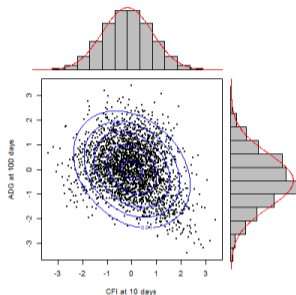


Figure: Plot of gaussian quantiles of the ranks for 2 phenotypes and contour plot of a bivariate Gaussian distribution

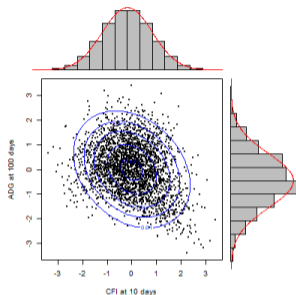
- ▶ the hypothesis of bivariate normality for the bivariate phenotypes seems to be unrealistic.
- ▶ an asymmetric bivariate distribution for the phenotypes.

➤ Example, Large-White dataset



- ▶ Mardia's normality test based on the asymmetry (skewness) of the distribution: p-value was 10^{-6} leading to a rejection because of the asymmetry of the distribution

➤ Example, Large-White dataset



- ▶ The bivariate normality is questionable
- ▶ What about the REML estimations of the bivariate animal model, which assume the bivariate normality?

➤ Example, Large-White dataset

- ▶ Even if the marginals are Gaussian, the bivariate distribution may be non-Gaussian.
- ▶ In fact, the *copula* of the random vectors is not the Gaussian Copula

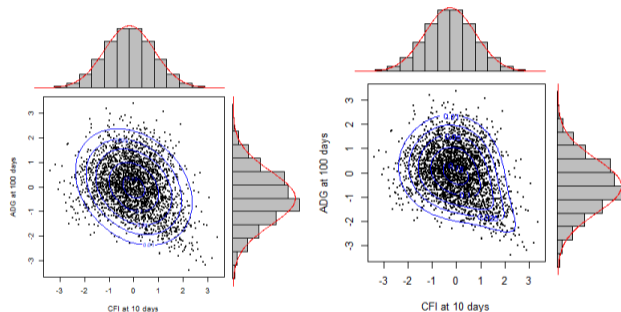


Figure: (right) contour plot of a Joe-Frank copula

➤ Some mathematics, copulas

Theorem of [Sklar(1959)]

Let $\mathbf{X} = (X_1, X_2)$ be a 2-dimensional random vector with c.d.f. \mathbf{F} and let F_1, F_2 be the marginal c.d.f. of \mathbf{X} assuming continuous. Then it exists a unique function $C : [0, 1]^2 \rightarrow [0, 1]$ such that:

$$\mathbf{F}(\mathbf{x}) = C\{F_1(x_1), F_2(x_2)\}, \quad \mathbf{x} = (x_1, x_2) \in \mathbb{R}^2.$$

- ▶ The copula C characterizes the dependence structure of vector \mathbf{X} .

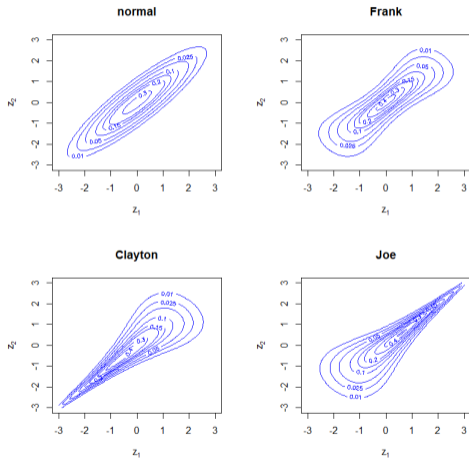


A. Sklar.

Fonctions de répartition à n dimensions et leurs marges.

Publications de l'Institut de Statistique de l'Université de Paris, 8:229–231, 1959.

Contour plots of bivariate distributions with Gaussian margins and several copula



Plan

Introduction

Simulation

Results

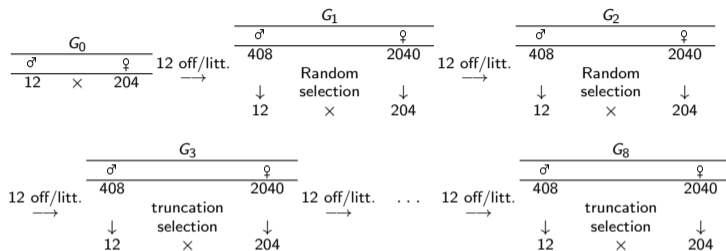


INRAE

Copula miss-specification on REML estimation

13-15 sept 2022 / Journées Scientifiques Département de Génétique Animal 2022/Tom Rohmer

Population simulation



Selection:

- ▶ $G_1 - G_3$ the reproducers were chosen at random
- ▶ $G_4 - G_8$ selection from a combination of their EBV
- ▶ Full/half siblings were not mated
- ▶ selection rate: 2.9% for the males and 10% for the females

➤ Phenotypes simulation

Simulated distribution:

- ▶ $(\mathbf{a}_1, \mathbf{a}_2) \sim \mathcal{N}(0, \mathbf{G} \otimes \mathbf{A})$.

- ▶ \mathbf{G} genetic covariance matrix

- ▶ \mathbf{A} additive genetic relationship matrix associated to the pedigree;

$$\rightarrow a_{i,j} = 0.5(a_{i_S,j} + a_{i_D,j}) + M_{i,j} \quad j = 1, 2, \text{ with } (M_{i,1}, M_{i,2}) \sim \mathcal{N}(0, \mathbf{G}/2)$$



➤ Phenotypes simulation

Simulated distribution:

- ▶ $(\mathbf{a}_1, \mathbf{a}_2) \sim \mathcal{N}(0, G \otimes A)$.

- ▶ G genetic covariance matrix

- ▶ A additive genetic relationship matrix associated to the pedigree;

$$\rightarrow a_{i,j} = 0.5(a_{i_S,j} + a_{i_D,j}) + M_{i,j} \quad j = 1, 2, \text{ with } (M_{i,1}, M_{i,2}) \sim \mathcal{N}(0, G/2)$$

- ▶ $(\varepsilon_{i,1}, \varepsilon_{i,2})$, $i = 1, \dots, n$, have standard Gaussian margins and copula C .

Copula C were Gaussian, Frank, Clayton and Joe with Kendall's correlation to 0.7 (Pearson's correlation between 0.85 and 0.89).



➤ Phenotypes simulation

Simulated distribution:

▶ $(\mathbf{a}_1, \mathbf{a}_2) \sim \mathcal{N}(0, \mathbf{G} \otimes \mathbf{A})$.

▶ \mathbf{G} genetic covariance matrix

▶ \mathbf{A} additive genetic relationship matrix associated to the pedigree;

$$\rightarrow a_{i,j} = 0.5(a_{i_S,j} + a_{i_D,j}) + M_{i,j} \quad j = 1, 2, \text{ with } (M_{i,1}, M_{i,2}) \sim \mathcal{N}(0, \mathbf{G}/2)$$

▶ $(\varepsilon_{i,1}, \varepsilon_{i,2})$, $i = 1, \dots, n$, have standard Gaussian margins and copula C .

Copula C were Gaussian, Frank, Clayton and Joe with Kendall's correlation to 0.7 (Pearson's correlation between 0.85 and 0.89). The phenotypes vector $\mathbf{y}_j = (y_{1,j}, \dots, y_{n,j})$, $j = 1, 2$ were obtained following the bivariate animal model:

$$\begin{cases} \mathbf{y}_1 = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{a}_1 + \boldsymbol{\varepsilon}_1 \\ \mathbf{y}_2 = \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{a}_2 + \boldsymbol{\varepsilon}_2. \end{cases}$$

\mathbf{X}_j the design matrices for the fixed effects and $\boldsymbol{\beta}_j$ associated parameter.

➤ A web Rshiny application

- ▶ https://testmyshinyapply.shinyapps.io/Shiny_bivariate_phen/



Plan

Introduction

Simulation

Results

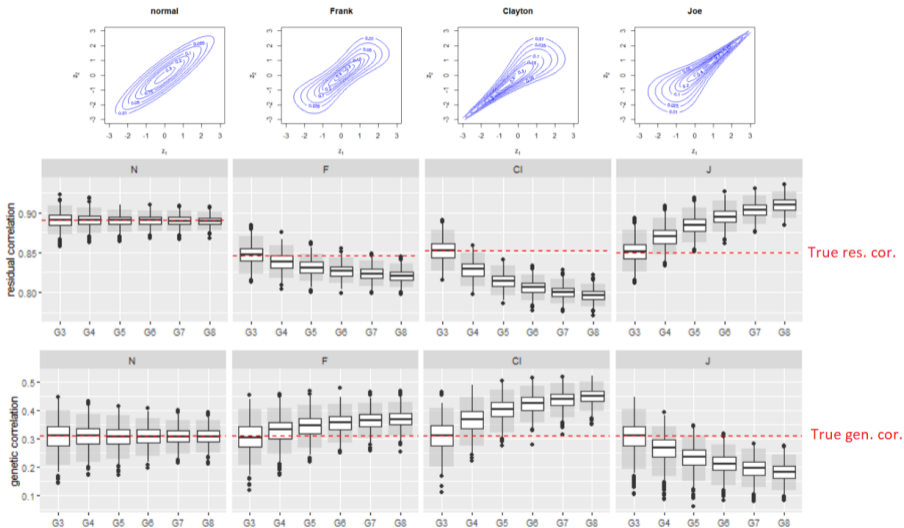


INRAE

Copula miss-specification on REML estimation

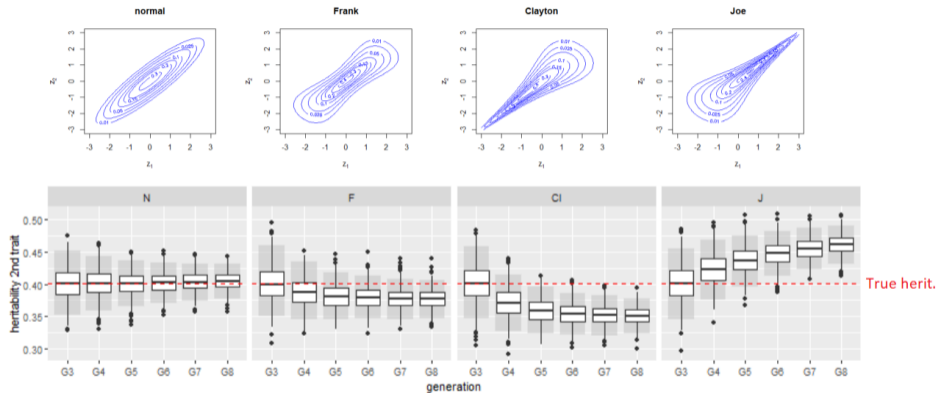
13-15 sept 2022 / Journées Scientifiques Département de Génétique Animal 2022/Tom Rohmer

➤ 1000 runs, Estim. correlations, $h_1^2 = h_2^2 = 0.40$, $\rho_a = 0.31$





1000 runs, Estim. heritability $h_1^2 = 0.15$, $h_2^2 = 0.40$



INRAE

Copula miss-specification on REML estimation

13-15 sept 2022 / Journées Scientifiques Département de Génétique Animal 2022/Tom Rohmer

Conclusion

1. With Random selection: no impact of the copula
2. With truncation selection;
 - ▶ For balanced heritabilities:
 - ▶ Significant impact on correlations;
 - ▶ very low biases for heritability
 - ▶ For unbalanced heritabilities:
 - ▶ Significant impact on the estim. heritabilities for the trait with moderate heritability
 - ▶ Significant impact on residual correlations;
 - ▶ moderate biases (but non-significant) on genetic correlations;
 - ▶ no impact on the estim. heritabilities for the trait with low heritability



Perspectives

GOS 4 Prise en compte de la structure de dépendance des résidus dans l'estimation des paramètres dans les modèles animal classiques, multivariées voir de random régression:

- ▶ Mise en place d'un algorithme SAEM pour l'estimation jointe des paramètres de variance-covariance via le modèle copule.
 - ▶ Meilleure connaissance des héritabilités des caractères et des corrélations génétiques dans le cas de forte corrélation dans les queues de distribution.
 - ▶ Amélioration des processus de sélection



Perspectives

GOS 4 Prise en compte de la structure de dépendance des résidus dans l'estimation des paramètres dans les modèles animal classiques, multivariées voir de random régression:

- ▶ Mise en place d'un algorithme SAEM pour l'estimation jointe des paramètres de variance-covariance via le modèle copule.
 - ▶ Meilleure connaissance des héritabilités des caractères et des corrélations génétiques dans le cas de forte corrélation dans les queues de distribution.
 - ▶ Amélioration des processus de sélection

GOS 1 Modélisation de la variabilité des caractères et amélioration des processus de sélection: intégration de données de capteurs haut-débits, nombreuses, à grandes dimensions et mesurées à des échelles de plus en plus fines dans ces modèles



Perspectives

GOS 4 Prise en compte de la structure de dépendance des résidus dans l'estimation des paramètres dans les modèles animal classiques, multivariées voir de random régression:

- ▶ Mise en place d'un algorithme SAEM pour l'estimation jointe des paramètres de variance-covariance via le modèle copule.
 - ▶ Meilleure connaissance des héritabilités des caractères et des corrélations génétiques dans le cas de forte corrélation dans les queues de distribution.
 - ▶ Amélioration des processus de sélection

GOS 1 Modélisation de la variabilité des caractères et amélioration des processus de sélection: intégration de données de capteurs haut-débits, nombreuses, à grandes dimensions et mesurées à des échelles de plus en plus fines dans ces modèles

- ▶ De l'information non résumé
- ▶ analyse conjointe de caractères de natures très différentes
- ▶ des modèles potentiellement non-linéaires

➤ Merci de votre attention

