

# ESTIMATION CONSERVATIVE D'UN RISQUE EN FIABILITÉ DES STRUCTURES AVEC DÉPENDANCE ENTRE VARIABLES ALÉATOIRES INCOMPLÈTEMENT SPÉCIFIÉE

Nazih Benoumechiara<sup>1,2,\*</sup> & Gérard Biau<sup>1</sup> & Bertrand Michel<sup>3</sup> & Philippe Saint-Pierre<sup>4</sup>  
& Roman Sueur<sup>2</sup> & Nicolas Bousquet<sup>2,4</sup> & Bertrand Iooss<sup>2,4</sup>

<sup>1</sup> *Laboratoire de Statistique Théorique et Appliquée, 4 place Jussieu, 75005 Paris, France*

<sup>2</sup> *EDF R&D, 6 quai Watier, 78400 Chatou, France*

<sup>3</sup> *Ecole Centrale de Nantes, 1 Rue de la Noë, 44300 Nantes, France*

<sup>4</sup> *Institut de Mathématiques de Toulouse, 31062, Toulouse, France*

\* *nazih.benoumechiara@upmc.fr*

**Résumé.** Dans un problème de fiabilité des structures, les paramètres d'entrée incertains sont décrits par un modèle probabiliste généralement obtenu à partir d'avis d'experts ou de mesures issues d'essais mécaniques. Cependant dans de nombreux cas, ce modèle est incomplet. En effet, certaines informations peuvent être indisponibles ou trop coûteuses à obtenir, en particulier sur la structure de dépendance entre les variables d'entrée. La pratique industrielle la plus courante consiste alors à faire l'hypothèse de l'indépendance des entrées pour les paires de variables dont la structure de dépendance est inconnue. Cette solution, facile à mettre en œuvre mais sans fondement théorique, peut alors mener à une évaluation trop optimiste du risque. C'est pourquoi, dans le but de garantir le conservatisme de la méthode, nous proposons d'explorer un ensemble de scénarios de dépendance afin de déterminer la structure la plus pénalisante. Cette démarche mène ainsi à une estimation plus pessimiste des marges mais également plus robuste vis-à-vis des critères réglementaires.

**Mots-clés.** Fiabilité des structures, Expériences Numériques, Copules, R-Vines

**Abstract.** In a structural reliability problem, the random input parameters are described by a probabilistic model generally obtained from expert feedbacks or data from experimental tests. However, in numerous cases, this model is incomplete. Indeed, some information may be unavailable or too costly to obtain, especially for the dependence structure between the input variables. The most common industrial practice consists in assuming the independence of the inputs for the pairs of variables for which the dependence structure is unknown. This solution, easy to implement but without any theoretical foundation, can lead to an overly optimistic evaluation of the risk. Therefore, in order to guarantee the conservatism of the method, we suggest to explore a set of dependency scenarios in order to determine the most penalizing structure. This approach leads to a more pessimistic estimate of margins, but is also more robust with respect to regulatory criteria.

**Keywords.** Structural reliability, Computer Experiments, Copulas, R-Vines

# 1 Introduction

L'étude de certains composants présentant un haut niveau de fiabilité requiert l'utilisation de méthodes de fiabilité des structures. La démarche consiste à évaluer un indicateur de risque  $y \in \mathbb{R}$  à partir d'un modèle numérique  $\eta : \mathbb{R}^d \rightarrow \mathbb{R}$  représentant le comportement du composant en fonction de paramètres d'entrée  $\mathbf{x} = (x_1, \dots, x_d)$ . Ces paramètres, considérés comme incertains, sont modélisés par des variables aléatoires, de sorte que l'indicateur de risque  $Y = \eta(\mathbf{X})$  est lui-même une variable aléatoire. Il est nécessaire de caractériser la loi du vecteur  $\mathbf{X}$  pour toute étude portant sur un critère d'intérêt  $\mathcal{C}(Y)$  de la distribution de sortie du modèle, par exemple le quantile de la distribution de sortie auquel on s'intéresse ici plus particulièrement.

Dans de nombreux cas, le modèle probabiliste utilisé est incomplet car obtenu à partir de  $d$  jeux de données expérimentales distincts. En l'absence de données jointes pour le vecteur  $\mathbf{X}$ , il n'est en effet pas possible de caractériser la structure de dépendance de  $\mathbf{X}$ , mais uniquement la loi des marginales  $X_1, \dots, X_d$ . La pratique industrielle la plus courante est alors d'effectuer l'étude en supposant l'indépendance de ces variables. Malheureusement, sans fondement théorique ou justification physique, cette approche peut mener à des résultats trop optimistes. C'est ce qu'ont montré [1, 2], illustrant l'influence de la corrélation de variables aléatoires sur la fiabilité d'un système. Ils concluent à la nécessité de prendre en compte la corrélation dans les calculs fiabilistes. De même, l'article [3], qui se borne au cas d'une dépendance bi-dimensionnelle, indique que le choix de la structure de dépendance peut aussi fortement influencer le critère d'intérêt  $\mathcal{C}(Y)$ .

La démarche proposée ici vise à garantir le conservatisme de l'estimation de  $\mathcal{C}(Y)$ , par l'exploration d'un ensemble de structures de dépendance. La valeur de  $\mathcal{C}(Y)$  finalement retenue sera celle jugée pénalisante, c'est à dire celle menant à l'estimation la plus pessimiste. La structure de dépendance correspondante, définie par la minimisation d'une quantité stochastique et dépendante des données, est alors un extremum-estimateur [4]. Pour cette exploration, différentes approches sont possibles, basées sur différentes caractérisations de la copule de la loi de  $\mathbf{X}$  [5]. Nous choisissons un cadre paramétrique de copules multi-dimensionnelles construites à l'aide de *Regular Vines* [6], ou R-Vines.

L'approche retenue consiste à minimiser le quantile d'ordre  $\alpha$  de  $Y$  sur une grille fine dans le domaine  $\Theta$  des paramètres  $\theta$  de la copule  $C_\theta$  de  $\mathbf{X}$ . Nous proposons aussi un algorithme consistant à pénaliser itérativement ce quantile par l'ajout des paires de variables diminuant le plus la valeur du critère. Notons que cette méthodologie peut être appliquée pour tout problème où l'information sur la dépendance est manquante ou incomplète, et pour toute quantité d'intérêt  $\mathcal{C}(Y)$ .

## 2 Estimation du quantile le plus pénalisant

Soit  $P_1, \dots, P_i, \dots, P_d$  les distributions marginales des variables  $X_1, \dots, X_d$ . Nous introduisons l'ensemble  $\mathcal{P}(P_1, \dots, P_d)$  regroupant toutes les distributions multivariées  $P$

sur  $\mathbb{R}^d$  telle que chaque marginale  $i$  de  $P$  est égale à  $P_i$ . Par la suite,  $\mathcal{P}$  désignera  $\mathcal{P}(P_1, \dots, P_d)$ . Ainsi,  $\mathcal{P}$  représente l'ensemble des distributions possibles pour  $\mathbf{X}$ , dont seules les marginales sont connues.

Nous nous intéressons à  $(P_\theta)_{\theta \in \Theta}$ , sous ensemble de  $\mathcal{P}$  associé à un ensemble de familles paramétriques de copules  $(C_\theta)_{\theta \in \Theta}$ . D'après le théorème de Sklar [5], la fonction de répartition  $F_\theta$  associée à la mesure  $P_\theta$  peut s'écrire, pour tout  $\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$ , comme

$$F_\theta(\mathbf{x}) = C_\theta(F_1(x_1), \dots, F_d(x_d)),$$

où les  $(F_i)$  sont les fonctions de répartition continues des distributions marginales  $P_i$ .

Nous notons respectivement  $G_\theta$  et  $G_\theta^{-1}$ , la fonction de répartition et la fonction quantile de  $Y$ . Pour une famille de copule paramétrique  $C = (C_\theta)_{\theta \in \Theta}$  et pour un alpha fixé  $\alpha \in (0, 1)$ , le quantile minimum  $G_C^{-1*}(\alpha)$  est défini par

$$G_C^{-1*}(\alpha) := \inf_{\theta \in \Theta} G_\theta^{-1}(\alpha).$$

Nous avons alors

$$\theta_C^* \in \operatorname{argmin}_{\theta \in \Theta} G_\theta^{-1}(\alpha).$$

où  $\theta_C^*$  est la structure de dépendance la plus pénalisante.

Nous considérons  $\eta$  comme une boîte noire. Cependant, le quantile  $G_\theta^{-1}(\alpha)$  peut être estimé à partir de  $n$  réalisations  $(y^{(i)})_{i \in \{1, \dots, n\}}$  de la variable aléatoire  $Y$ . Nous choisissons le quantile empirique  $\hat{G}_\theta^{-1}(\alpha)$  comme estimateur de  $G_\theta^{-1}(\alpha)$ . L'extremum-estimateur s'écrit alors

$$\hat{\theta} = \operatorname{argmin}_{\theta \in \Theta} \hat{G}_\theta^{-1}(\alpha). \quad (1)$$

La consistance de l'estimateur  $\hat{\theta}$  est montrée par le théorème 1.

**Théorème 1** *Sous des hypothèses de régularités du modèle  $\eta$  et de la loi de probabilité  $P_\theta$ , nous avons pour  $\alpha \in (0, 1)$  et pour tout  $\varepsilon > 0$ ,*

$$P \left( \left| \hat{G}_\theta^{-1}(\alpha) - G_C^{-1*}(\alpha) \right| > \varepsilon \right) \xrightarrow{n \rightarrow \infty} 0. \quad (2)$$

De plus si,  $\theta_C^* \in \Theta$  minimise  $\theta \mapsto G_\theta^{-1}(\alpha)$ , donc pour tout  $h > 0$  nous avons,

$$\mathbb{P}[|\hat{\theta} - \theta_C^*| > h] \xrightarrow{n \rightarrow \infty} 0.$$

Puisque l'ensemble  $\Theta$  n'est pas explorable aisément de façon continue, nous choisissons de l'approximer par une grille fine  $\Theta_K$  de cardinalité  $K$ . L'estimateur  $\hat{\theta}$  de l'équation (1) est alors approximé par

$$\hat{\theta}_K = \operatorname{argmin}_{\theta \in \Theta_K} \hat{G}_\theta^{-1}(\alpha). \quad (3)$$

De plus, la consistance de  $\hat{\theta}_K$  est vérifiée pour une grille  $\Theta_K$  suffisamment fine et sous les mêmes hypothèses que pour le théorème 1.

En pratique, décrire une structure de dépendance multidimensionnelle n'est pas chose aisée. Nous choisissons alors de construire la copule par une structure de R-Vines, basée sur une *Pair Copula Construction* (PCC). Ceci permet de définir la densité d'une copule multidimensionnelle comme produit de densités de copules bivariées. Cependant, de très nombreuses PCC existent, ce nombre augmentant de surcroît significativement avec la dimension du problème. Voir [6] pour plus de détails sur la PCC et les R-Vines.

Par ailleurs, pour certaines familles de copule, le support de  $\theta$  est un sous ensemble non borné de  $\mathbb{R}$ . Dans ce cas,  $\Theta$  est un sous ensemble de  $\mathbb{R}^p$  (où  $p$  est le nombre de paires de variables potentiellement dépendantes), rendant problématique la construction de la grille  $\Theta_K$ . C'est pourquoi, en pratique, nous choisissons de construire la grille non pas sur  $\Theta$ , mais sur l'espace  $[-1, 1]^p$  correspondant au domaine de variation d'une mesure de concordance telle que le Tau de Kendall. Ce dernier présente l'avantage d'être explicitement relié à  $\theta$  pour de nombreuses familles de copules. Voir [7] pour une description détaillée des mesures de concordances et de leurs liens avec les copules.

### 3 Méthodologie

Nous considérons que les familles de copules sont préalablement définies pour chaque paire de variables. Certaines copules présentant des dépendances de queue non symétriques, un certain *a priori* sur le comportement du quantile en fonction de la dépendance des paires de variables est requis pour correctement sélectionner l'orientation des copules. Sans connaissance préalable, il est préférable d'utiliser des copules symétriques (ex : gaussiennes).

Une première approche considère l'ensemble des structures de R-Vines possibles et détermine le paramètre  $\theta$  le plus pénalisant. Pour une structure de R-Vine donnée,  $K$  estimations du quantile sont effectuées selon une grille  $\Theta_K$ . Pour un nombre  $n$  d'évaluations du modèle,  $\hat{\theta}_K$  est alors déterminé à partir de (3). Cependant, le coût de cette approche rend son utilisation très limitée. En effet le nombre de structures de Vine augmente significativement avec la dimension : 480 pour  $d = 5$  et 23,040 pour  $d = 6$  (voir [8] pour plus de détails). C'est pourquoi, nous proposons une seconde approche permettant de déterminer itérativement le paramètre  $\theta$  le plus pénalisant. Cette méthode repose sur l'hypothèse selon laquelle l'influence de la dépendance sur le quantile n'est régie que par quelques paires de variables. Ceci est presque toujours vérifié en pratique. À partir de cela, il nous suffit de déterminer à chaque itération, la paire de variables minimisant le plus le quantile et de l'ajouter dans la structure de R-Vines. L'algorithme 1 décrit ces étapes. Nous considérons  $p_{max}$  comme le nombre maximum de paires et  $\Delta Q$  comme la diminution minimale du quantile entre chaque itération. L'algorithme renvoie l'estimation du quantile minimum  $Q_{min}$  et la structure de dépendance  $\theta_{min}$  associée.

**Data :**  $\alpha \in (0, 1) \mid p_{max} \leq p \mid \Delta Q > 0$   
**Result :**  $Q_{min}$  and  $\theta_k$   
 $\Theta = \emptyset;$   
 $Q_{min} = +\infty ;$   
 $Q_{k-1} = +\infty ;$   
 $k = 1;$   
**while**  $k \leq p_{max}$  **do**  
    Pour chaque paire de variables;  
    **for**  $i \in 1 : d - 1$  **do**  
        **for**  $j \in i + 1 : d$  **do**  
            Si la paire n'appartient pas déjà à la structure de dépendance;  
            **if**  $\Theta_{ij} \notin \Theta$  **then**  
                Estimation du minimum sur  $\Theta \otimes \Theta_{ij};$   
                 $\hat{\theta}_{ij} = \operatorname{argmin}_{\theta \in \Theta \otimes \Theta_{ij}} \hat{G}_{\theta}^{-1}(\alpha);$   
                Si le quantile est minimisé, nous gardons la structure;  
                **if**  $\hat{G}_{\hat{\theta}_{ij}}^{-1}(\alpha) < Q_{min}$  **then**  
                     $Q_{min} = \hat{G}_{\hat{\theta}_{ij}}^{-1}(\alpha);$   
                     $\theta_{min} = \hat{\theta}_{ij};$   
                     $\Theta_k = \Theta_{ij};$   
            La structure de R-Vine est mise à jour ainsi que l'ensemble  $\Theta;$   
             $\Theta = \Theta \otimes \Theta_k;$   
            Si le quantile ne décroît pas suffisamment;  
            **if**  $|Q_{k-1} - Q_{min}| < \Delta Q$  **then**  
                **stop;**  
            **else**  
                 $Q_{k-1} = Q_{min} ;$   
                 $k = k + 1;$

**Algorithme 1 :** Construction itérative de la structure de dépendance la plus pénalisante.

Nous illustrons cette méthodologie sur un problème industriel de fiabilité des structures où la dépendance entre les variables aléatoires est inconnue. Le modèle numérique  $\eta$  prend en entrée le vecteur aléatoire  $\mathbf{X} \in \mathbb{R}^6$ , dont seules les marginales sont connues. Les familles de copules choisies sont sélectionnées parmi un ensemble de familles paramétriques telles que des gaussiennes, gumbel ou clayton. Nous appliquons l'algorithme 1 afin d'obtenir une estimation plus conservative du quantile de la sortie  $Y$  pour  $\alpha = 1\%$ . La figure 1 montre la diminution du quantile par l'ajout successif de la paire la plus pénalisante. L'algorithme commence avec un quantile à l'indépendance de 1.75 et s'arrête au bout de six paires de variables pour un quantile proche des 1.0. Cela montre premièrement que la dépendance à une forte influence sur l'estimation du quantile. Mais aussi que les dépendances stochastiques des paires de variables n'ont pas toutes la même influence sur le quantile.

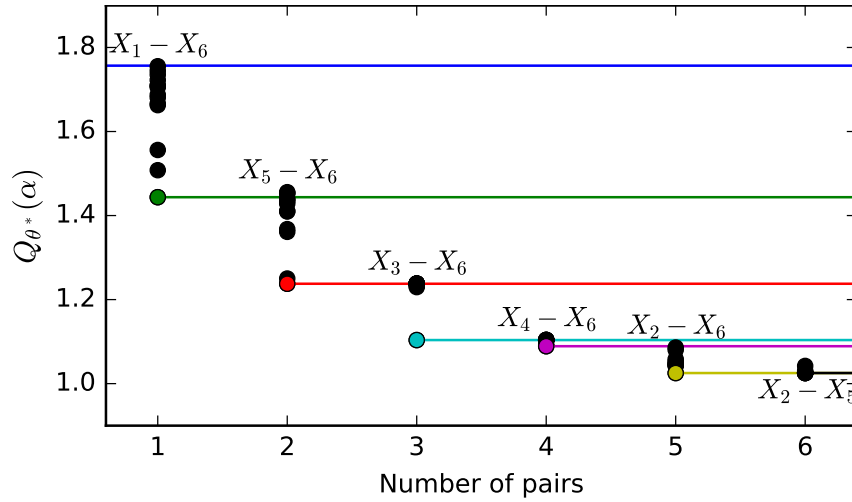


Figure 1: Minimisation du quantile par ajout successif des paires de variables les plus pénalisantes.

## Bibliographie

- [1] Grigoriu, M. and Turkstra, C. (1979). Safety of structural systems with correlated resistances. *Applied Mathematical Modelling*, 3(2):130–136.
- [2] Thoft-Christensen, P. and Sørensen, J. D. (1982). Reliability of structural systems with correlated elements. *Applied Mathematical Modelling*, 6(3):171–178.
- [3] Tang, X.-S., Li, D.-Q., Rong, G., Phoon, K.-K., and Zhou, C.-B. (2013a). Impact of copula selection on geotechnical reliability under incomplete probability information. *Computers and Geotechnics*, 49:264–278.
- [5] Sklar, M. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, 8:229–231.
- [6] Bedford, T. and Cooke, R. M. (2002). Vines: A new graphical model for dependent random variables. *Annals of Statistics*, 30(4):1031–1068.
- [4] Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. *Handbook of econometrics*, 4:2111–2245.
- [7] North Frees, E. W. and Valdez, E. A. (1998). Understanding relationships using copulas. *American actuarial journal*, 2(1):1–25.
- [8] Morales Napoles, O., Cooke, R. M., and Kurowicka, D. (2010). About the number of vines and regular vines on n nodes.