Génération de distributions gamma bivariées par la méthode des copules.

B. Mesnil et S. Mahévas

Ecologie et modèles pour l'halieutique

Ifremer, Nantes

Forum AFH – Juin 2011

Contexte

- La C. E. demande au CIEM des avis incluant des options MSY / Fmsy
- ***OR*** MSY implique de combiner analyses Par-Recrue et relations Stock-Recrutement (RS-R)
- ***ET*** les RS-R sont les objets les plus incertains (après M) de la science halieutique (Hilborn et Walters 1992)

- ***DONC*** les estimations des points MSY sont entachées de grosses incertitudes ⇨ les évaluer & les refléter dans les avis

Option spontanée

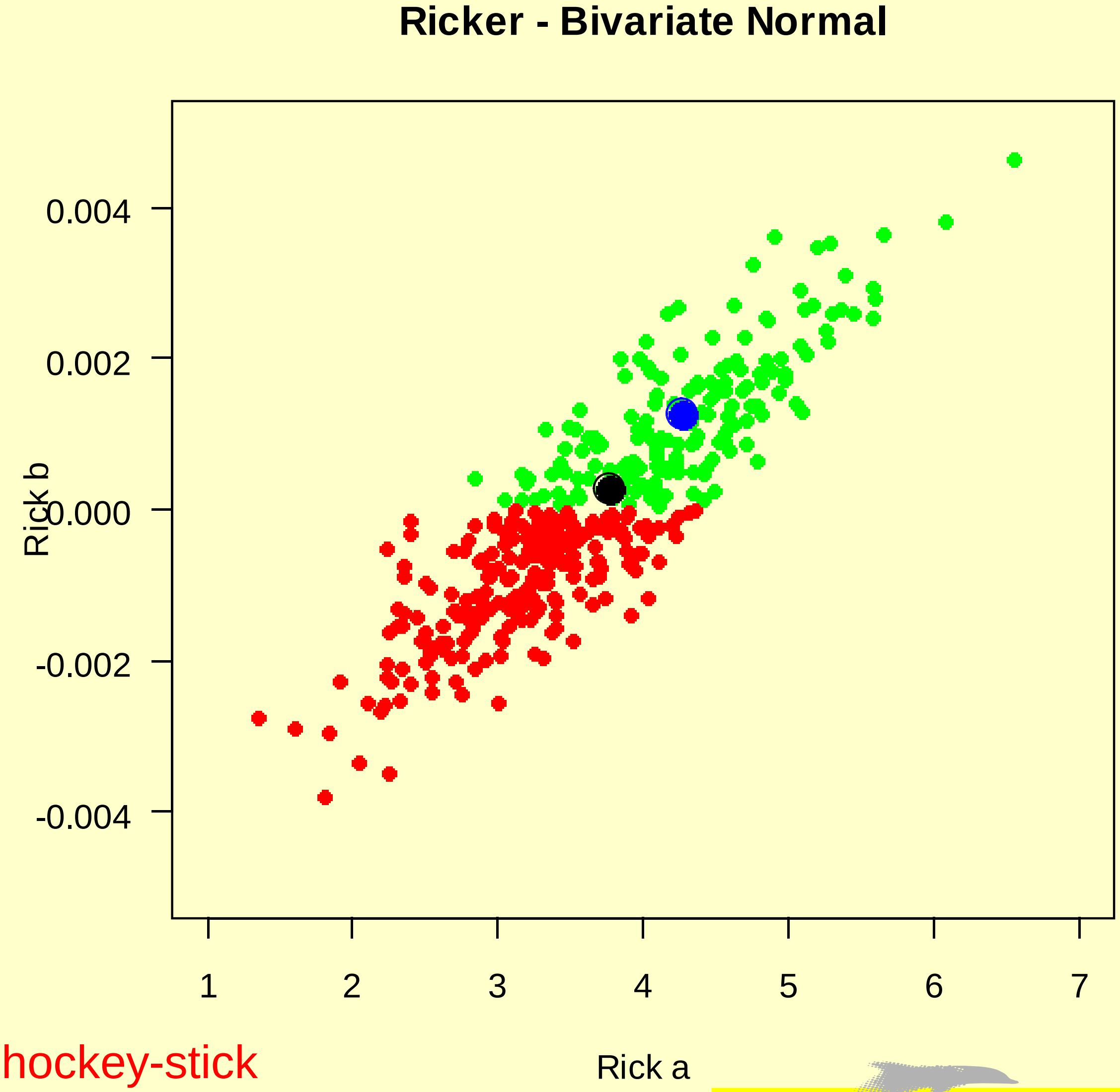
- Ajustement classique R-S : estimation des valeurs des param + variance et covariance par nls
- Faire de multiples tirages d'une loi bivariée, prenant en compte la structure de dépendance entre les paramètres des RS-R ;
  - calculer les points MSY pour chaque réplicat ;
  - estimer un Intervalle de Confiance (IC) d'après les quantiles de la distribution obtenue.
- ***MAIS*** on sait peu de choses sur la distribution des param RS-R (valeur nominale, variance-covariance), mais quid symétrie / skewness ?
- Par défaut (et facilité), la loi Normale bivariée

```
library(MASS)
tmp <- mvrnorm(n=1000,mu[1:2],vcov,empirical=TRUE)
```

Problème ! `any(tmp<0) = TRUE`

La bi-Normale conserve mu et covariance/corrélation mais n'est pas bornée $\Rightarrow$ produit des params < 0 : **IMPOSSIBLES**

On ne peut pas non plus ne garder que les positifs : moyenne (pt bleu vs. noir) et corrélation n'ont plus rien à voir; pire si on rescale pour rétablir la corrélation



Idem avec Beverton-Holt & hockey-stick

Que faire ?

- Choisir une distribution restreinte à l'axe positif, telle que la loi Gamma ou la Loi Lognormale

- ***MAIS*** peu de générateurs disponibles et un souci avec la corrélation (linéaire) ρ : il n'est pas possible de la conserver dès lors qu'il y a transfo non linéaire (toujours $\rho_{trans} < \rho_{orig}$) ; certains algorithmes imposent une ρ maxi et, parfois, une ρ positive seulement

- MIRACLE : **les copules !**

Refs: googler 'Embrechts' ou 'Genest'

Cas bi-varié Z=(X,Y)

- Fonction de répartition de Z=(X,Y)

$$F_Z(x,y) = P(X<=x,\ Y<=y)$$

- Fonction de répartition marginale de X

$$F_X(x) = P(X<=x)$$

- Fonction de répartition marginale de Y

$$F_Y(y) = P(Y<=y)$$

Propriétés :

- Si X et Y indépendantes alors $F_Z=F_X*F_Y$

Copule bi-variée Z=(X,Y)

- **Définition :** Une copule est une fonction de répartition bivariée *C* définie sur l'hypercube $[0, 1]^2$ et dont les deux marginales sont uniformes sur [0, 1]

- **Théorème de Sklar** [1959]

  Soient *C* une copule bi-variée, et *FX et FY* deux fonctions de répartition

  alors *C(FX(x), FY(y))* est une fonction de répartition de marginales *FX et FY*

Copule bi-variée Z=(X,Y)

- **Théorème de Sklar [1959] - suite**

Soit F_Z une fonction de répartition à 2 dimensions avec des marginales F_X *et* F_Y *continues*, alors il existe une copule *C bi-variée* telle que $F_Z(x,y) = C(F_X(x), F_Y(y))$

- La loi de *Z* c'est deux fonctions de répartition qui capturent les marginales et une copule qui mesure la dépendance
- En général, on connaît F_X et F_Y, on maîtrise moins bien la dépendance…

Comment choisir la copule ?

« *the question "which copula to use?" has no obvious answer* » Embrechts 2009

Exemples de copules et caractéristiques

		Fonctions	Spearman	Kendall
Elliptiques	Gauss	$\int_{-\infty}^{\Phi^{-1}(u)}\int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi(1-\rho^2)^{1/2}} \times \exp\left[\frac{-(x_1^2-2\rho x_1x_2+x_2^2)}{2(1-\rho^2)}\right] dx_1dx_2$	$\frac{6}{\pi}\arcsin(\frac{\rho}{2})$	$\frac{2}{\pi}\arcsin(\rho)$
	Student	$\int_{-\infty}^{t_\nu^{-1}(u)}\int_{-\infty}^{t_\nu^{-1}(v)} \left(\frac{\Gamma(\frac{\nu+2}{2})}{\Gamma(\frac{\nu}{2})\pi\nu\sqrt{1-\rho}}\right) \times \left(1+\frac{(x_1,x_2)^t\begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}(x_1,x_2)}{\nu}\right)^{\frac{-(\nu+2)}{2}} dx_1dx_2$	$\frac{6}{\pi}\arcsin(\frac{\rho}{2})$	$\frac{2}{\pi}\arcsin(\rho)$
Archimédiennes	Franck	$\frac{1}{\theta}\log\left(1+\frac{(\exp(\theta u)-1)(\exp\theta v)-1)}{\exp(\theta)-1}\right)$		$1-\frac{4(1-D(\theta))}{\theta}$

Ifremer

Copule bi-variée Z=(X,Y)

Liens (pas toujours simples) entre mesure de dépendance et C

- Pour cet exemple, la métrique de dépendance est la corrélation linéaire $\rho = \mathrm{Cov}(X; Y)^*(\mathrm{Var}(X)V(Y))^{-1/2}$ (ou de Pearson – mais les experts* préfèrent Kendall ou Spearman)
- et parmi les différents types de copules, une copule Gaussienne (elliptique) est choisie car adaptée aux dépendances symétriques et aisément paramétrée avec ρ

* Genest, C. & A. C. Favre (2007). "Everything you always wanted to know about copula modeling but were afraid to ask." *Journal of Hydrologic Engineering* **12**(4): 347-368

En pratique

- Ajuster la relation S-R; réserver params et var-cov

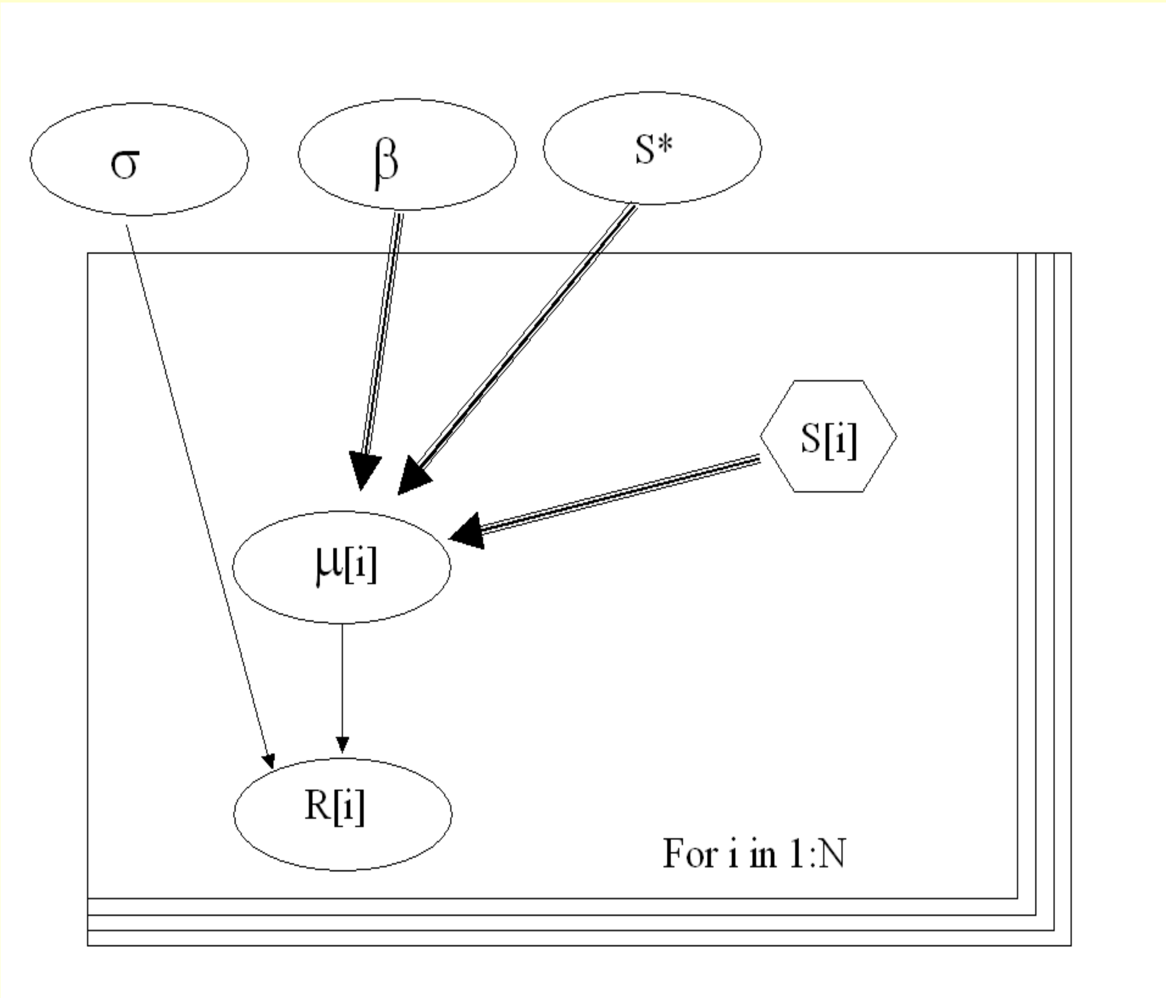
- Construire la copule Gaussienne **U** en appliquant la fonction quantile-Normale à une bi-Normale std

  Puisque l'on souhaite que les marges ~ Gamma, au préalable spécifier les paramètres *shape* et *scale* de chaque marge à partir des résultats de l'ajustement

- Enfin, passer les réplicats au module de calcul MSY, et retourner les distributions de MSY, Fmsy, Bmsy etc.

Comparaisons avec deux autres approches

- Modèle bayesien



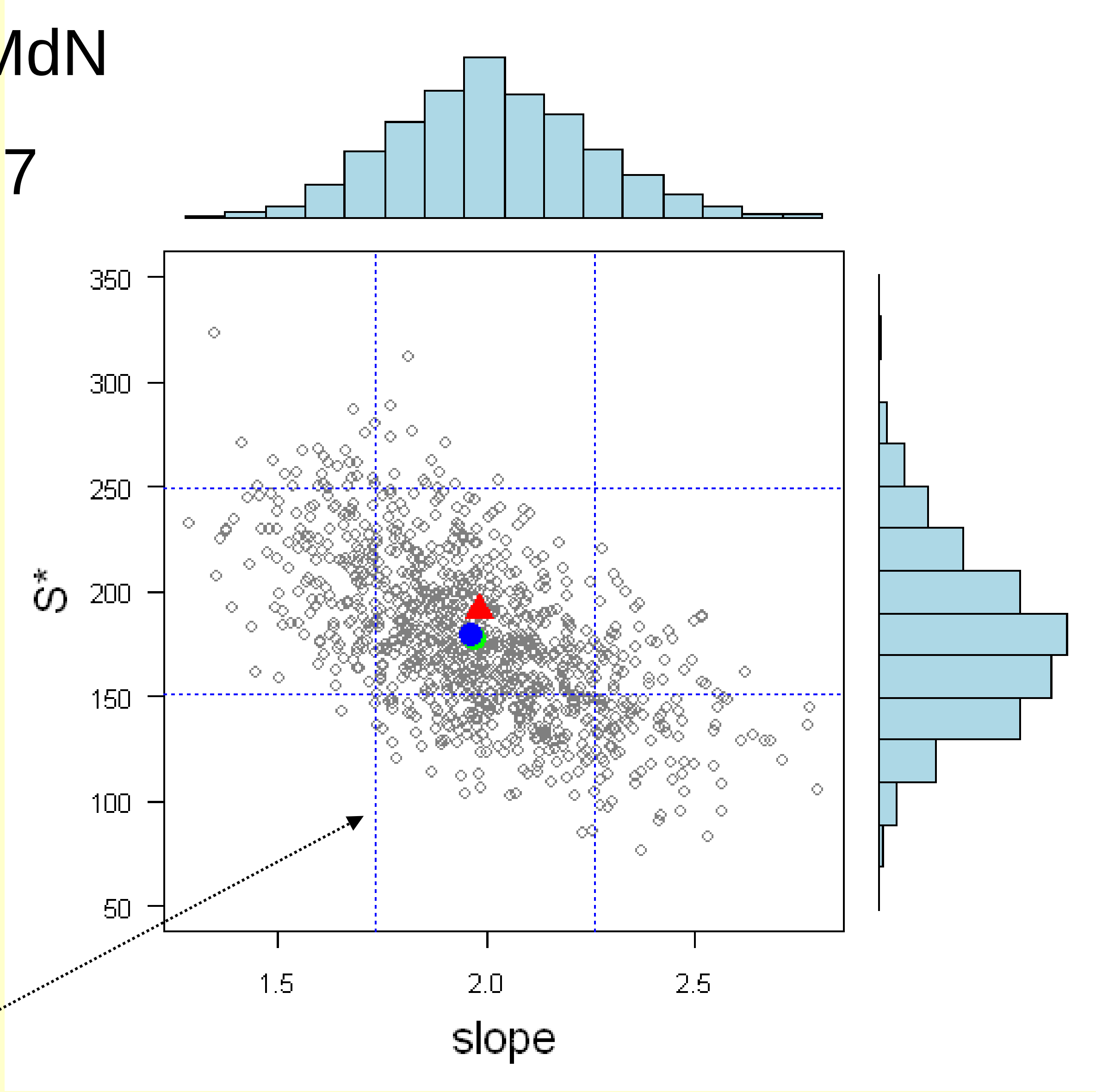
- Bootstrap des résidus

  réajustement sur recrutements ajustés + résidus versus SSB

Distribution jointe des paramètres SR-R (hockey-stick); copule Gauss & marges Γ

Ex. morue MdN

1950-2007



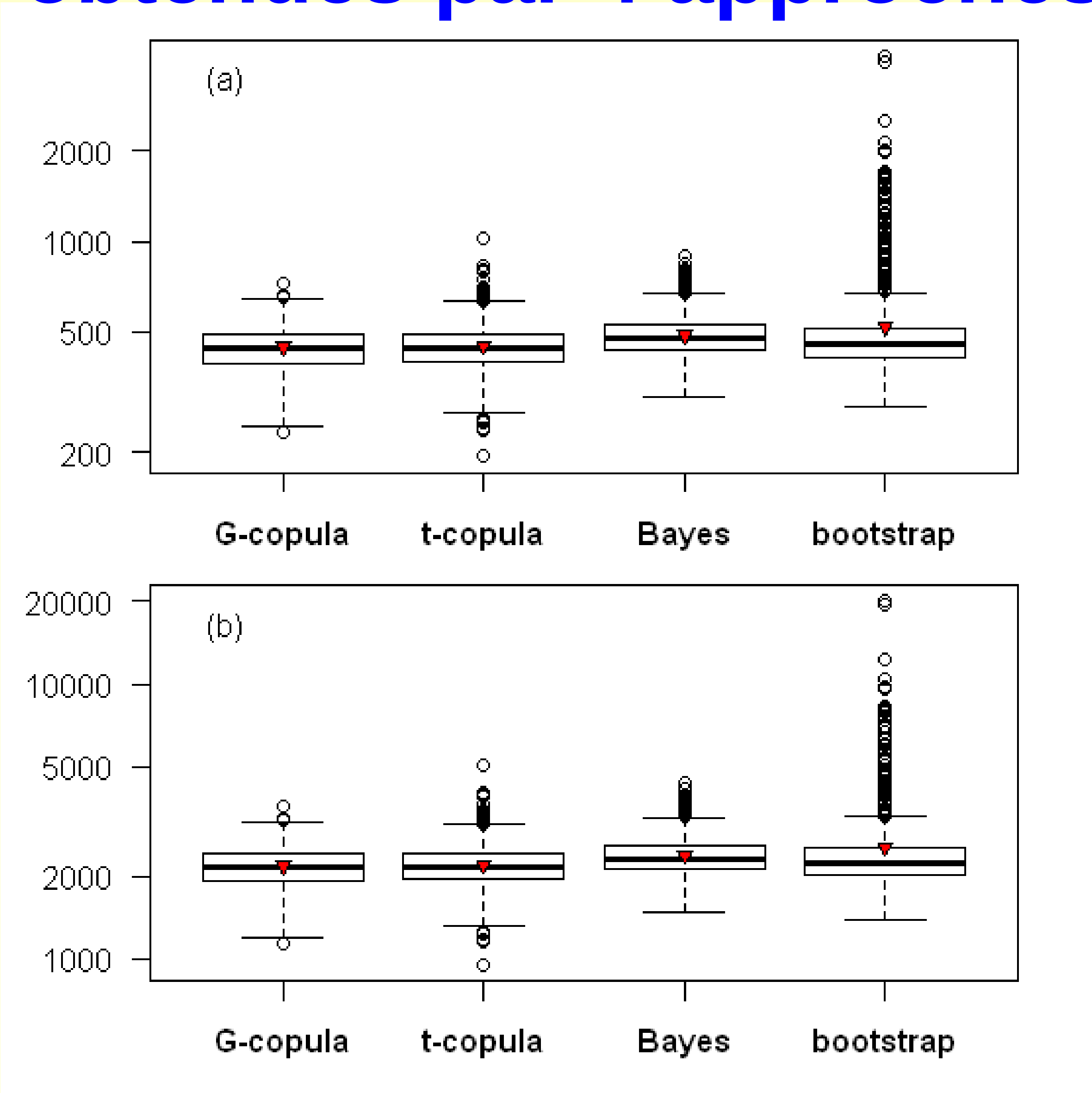
Rond vert = barycentre copule;

Rond bleu = param nominaux

Tiretés bleus = IC 90% d'un fit bayesien, dont barycentre = triangle

Distributions de MSY (a) et Bmsy (b) obtenues par 4 approches

NB: echelle log !



Inter-quartiles & médianes similaires, y compris bootstrap cochon

Que penser de ces copules ?

- Un côté magique, qui donc doit s'accompagner d'une bonne dose de méfiance
- Ca a l'air de faire ce que l'on voulait : paramètres tous positifs, corrélation bien préservée, distributions au choix et générées en un clin d'oeil
- Outil très flexible : on peut générer des bi-gammas, des bi-lognormales de paramètres différents, voire des mélanges gamma-Student ou gamma-lognormale etc., y compris avec ρ négatives
- D'autres types de copules en plus des Gaussiennes, avec d'autres propriétés /limitations.

NB: outils très populaires chez les financiers, mais accusés d'avoir favorisé la crise de Wall Street

FIN

The study of copulas and the role they play in probability, statistics, and stochastic process is a subject still in its infancy. There are many problems and much work to be done

R.B. Nelsen (2006)