

## Nouvelles fonctionnalités du package `fitdistrplus`

**Marie Laure Delignette-Muller**

Université de Lyon

UMR CNRS 5558, Laboratoire de Biométrie et Biologie Evolutive

VetAgro Sup, Campus vétérinaire de Lyon

et

**Christophe Dutang**

Université de Strasbourg

UMR CNRS 7501, Institut de Recherche Mathématique Avancée

Deuxièmes rencontres R, Lyon, 27-28 juin 2013



## Présentation du package `fitdistrplus`

`http://cran.r-project.org/package=fitdistrplus`  
`http://riskassessment.r-forge.r-project.org/`

Package pour **l'ajustement de distributions paramétriques à des données univariées**,

proposant diverses fonctions pour :

- le choix de distributions candidates pour décrire les données,
- l'ajustement de chacune des distributions candidates aux données,
- la comparaison des ajustements en vue de choisir la distribution la plus adaptée,
- le calcul, par bootstrap, de l'incertitude sur les paramètres estimés de la distribution choisie.

## Quelques spécificités du package

- **Différentes méthodes d'estimation des paramètres** sont proposées :
  - méthode du **maximum de vraisemblance (MLE)**,
  - méthode des **moments (MME)**,
  - méthode des **quantiles (QME)**,
  - méthode de **minimisation d'une statistique d'ajustement (MGE)** (8 distances proposées).
- **Prise en compte**
  - des données discrètes
  - et des **données censurées** quel que soit le type de censures (à droite, à gauche ou par intervalle).
- Possibilité de modifier l'algorithme d'optimisation (par défaut `optim`).

## Utilisation actuelle du package

**Package développé initialement pour l'appréciation quantitative du risque** notamment pour permettre la caractérisation hiérarchique de la variabilité et de l'incertitude dans le cadre de simulations de Monte carlo à deux dimensions (package `mc2d`).

**Package généraliste simple à utiliser,**  
utilisé dans des domaines variés :

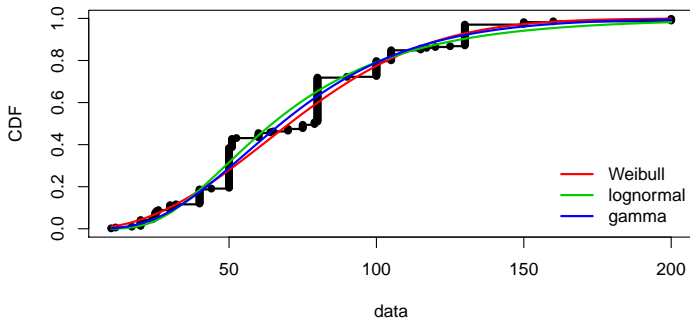
- risque alimentaire,
- épidémiologie,
- biologie moléculaire,
- bioinformatique,
- mathématiques financières et actuarielles, . . .

→ les retours des utilisateurs nous ont incité à développer de **nouvelles fonctionnalités**.

## Comparaison visuelle de plusieurs ajustements (cdfcomp)

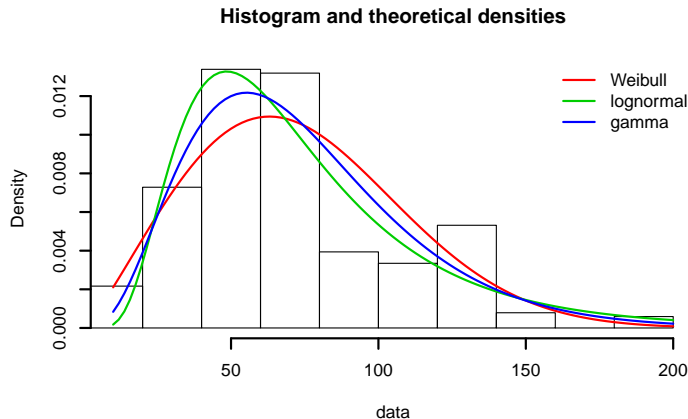
```
> data(groundbeef)
> fw <- fitdist(groundbeef$serving, "weibull")
> fg <- fitdist(groundbeef$serving, "gamma")
> fln <- fitdist(groundbeef$serving, "lnorm")
> cdfcomp(list(fw, fln, fg),
+         legendtext=c("Weibull", "lognormal", "gamma"))
```

Empirical and theoretical CDFs



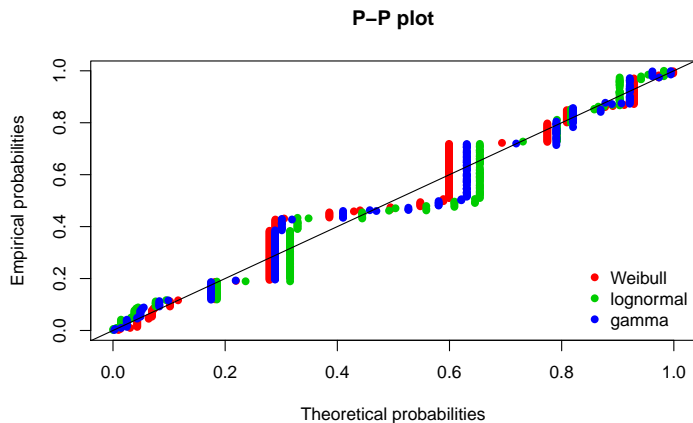
## Comparaison visuelle de plusieurs ajustements : représentation en densité de probabilité (denscomp)

```
> denscomp(list(fw,fln,fg),  
+          legendtext=c("Weibull", "lognormal", "gamma"))
```



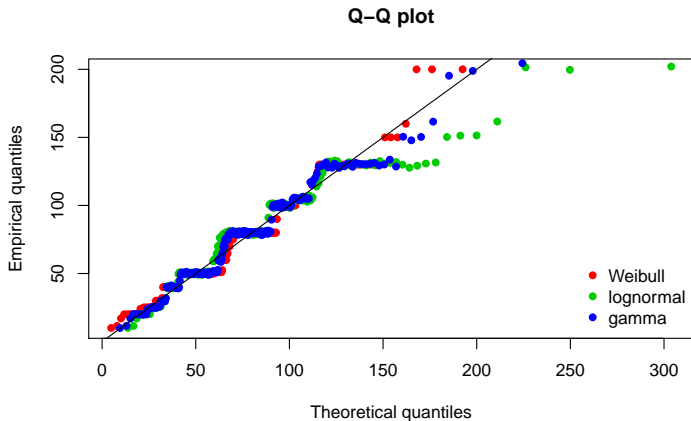
## Comparaison visuelle de plusieurs ajustements : P-P plot (ppcomp)

```
> ppcomp(list(fw,fln,fg),  
+         legendtext=c("Weibull", "lognormal", "gamma"))
```



## Comparaison visuelle de plusieurs ajustements : Q-Q plot (qqcomp)

```
> qqcomp(list(fw,fln,fg),  
+         legendtext=c("Weibull", "lognormal", "gamma"))
```





## Comparaison numérique de plusieurs ajustements

- **Statistiques d'ajustement**,  
Kolmogorov-Smirnov, Cramér-von Mises et Anderson-Darling,
- **et critères d'information** , AIC et BIC.

```
> gofstat(list(fw,fln,fg),  
+         fitnames = c("Weibull", "lognormal", "gamma"))
```

Goodness-of-fit statistics

	Weibull	lognormal	gamma
Kolmogorov-Smirnov statistic	0.140	0.149	0.128
Cramer-von Mises statistic	0.684	0.828	0.693
Anderson-Darling statistic	3.574	4.544	3.566

Goodness-of-fit criteria

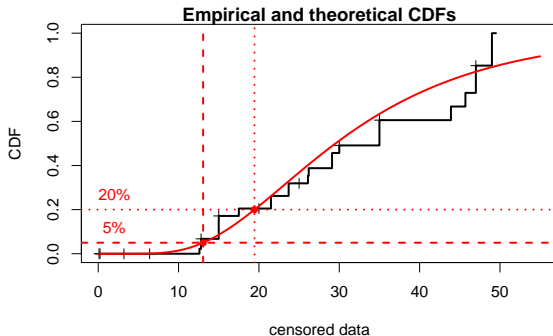
	Weibull	lognormal	gamma
Aikake's Information Criterion	2514	2527	2511
Bayesian Information Criterion	2522	2534	2518

## Calcul de quantiles à partir d'une loi ajustée : exemple sur données censurées en écotoxicologie

```
> data(salinity)
> f <- fitdistcens(salinity, "lnorm")
> quantile(f, probs = c(0.05, 0.20))
```

Estimated quantiles for each specified probability (censored data)

	p=0.05	p=0.2
estimate	13.1	19.5



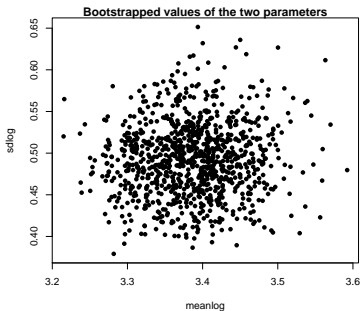
# Bootstrap

```
> b <- bootdistcens(f,niter=1001)
> summary(b)
```

Nonparametric bootstrap medians and 95% percentile CI

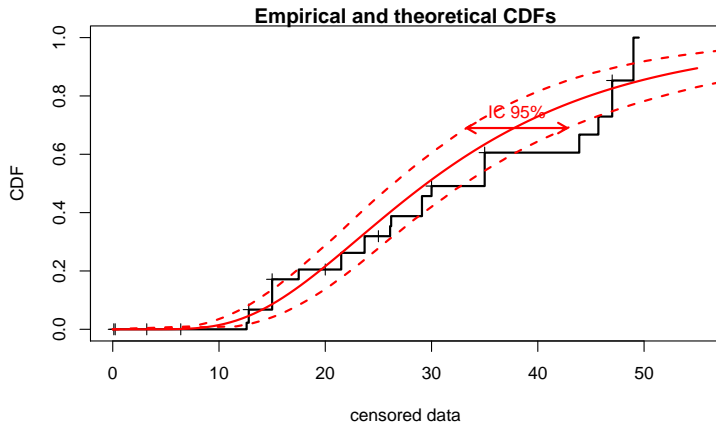
	Median	2.5%	97.5%
meanlog	3.387	3.274	3.500
sdlog	0.492	0.411	0.581

```
> plot(b)
```



## Intervalles de confiance bootstrap sur les quantiles

```
> cdfcompcons(f)
> q <- quantile(b, probs = seq(0,1,length=101))
> points(q$quantCI[1,],q$probs,type="l",lty=2,col="red")
> points(q$quantCI[2,],q$probs,type="l",lty=2,col="red")
```



Nous envisageons

- **d'étendre aux données censurées certaines méthodes** disponibles actuellement sur les données non censurées,
- **d'élargir le choix de distances** utilisables pour l'ajustement par minimisation de ces distances et pour l'évaluation de la qualité de l'ajustement (distances basées sur les quantiles),
- **d'élargir le choix des moments** pour l'ajustement par la méthode des moments ("L-moment" et "TL-moment"),
- **aborder l'ajustement de distributions multivariées.**

N'hésitez pas ? nous envoyer vos retours ou idées en vue de l'amélioration de ce package.