`fitdistrplus`, un package pour l'ajustement de distributions, et son utilisation dans le cadre de la simulation de Monte Carlo à deux dimensions avec le package `mc2d`

M. L. Delignette-Muller - VetAgro Sup - LBBE

30 novembre 2012

Développement de fitdistrplus au sein du projet "Risk Assessment with R"

Projet R-Forge

`http://riskassessment.r-forge.r-project.org/`

Partenaires :

- Régis Pouillot (FDA, Washington)
- Jean-Baptiste Denis (INRA-MIA, Jouy-en-Josas)
- Christophe Dutang (IRMA UMR CNRS 7501, Strasbourg))

Packages :

- fitdistrplus (développement MLDM et CD) : outils pour l'ajustement de distributions univariées (sur le CRAN depuis 2009)
- mc2d (développement RP) : outils pour la simulation de Monte Carlo à 2 dimensions (sur le CRAN depuis 2009)
- rebastaba (développement JBD) : outils pour la manipulation de réseaux bayésiens (sur Rforge)

Contexte nous ayant poussé à développer `fitdistrplus`

Décrire une distribution empirique par une distribution paramétrique choisie dans une famille prédéfinie de distributions : besoin fréquent (notamment en appréciation quantitative des risques)

- la fonction `fitdistr(MASS)` permet de réaliser l'étape d'estimation des paramètres et les autres étapes peuvent être implémentées en utilisant **R**
- `fitdistr(MASS)` ne propose pas d'autres méthodes d'estimation que le maximum de vraisemblance
- `fitdistr(MASS)` ne couvre pas l'ensemble du processus de choix et d'ajustement d'une distribution paramétrique
- aucun package disponible pour ajuster des distributions à des données censurées

Outils commerciaux proposés

Existence de logiciels commerciaux couramment utilisés en appréciation quantitative des risques, proposant :

- l'**ajustement systématique d'un très grand nombre de distributions** paramétriques
- Le **choix quasi automatique** de la meilleure distribution sur des critères d'ajustement criticables

Inconvénients :

- non transparence quant à la méthode d'estimation des paramètres
- non transparence quant à la définition des valeurs initiales des paramètres (échecs d'ajustement parfois surprenants)
- risques liés à l'automatisation de la démarche de choix d'une distribution, reposant souvent sur la seule comparaison de statistiques d'ajustement criticables

Notre vision de la démarche de description d'une distribution empirique par une distribution paramétrique

- Choix de distributions candidates parmi un petit nombre de distributions classiques à partir des caractéristiques de la variable étudiée, de l'examen visuel de la distribution empirique et du calcul des coefficients d'asymétrie et de pointicité.
- Ajustement des distributions candidates
- Choix de la distribution la plus adaptée en fonction des exigences de l'utilisateur, notamment à partir de l'examen visuel des ajustements

Notre vision de la démarche de description d'une distribution empirique par une distribution paramétrique

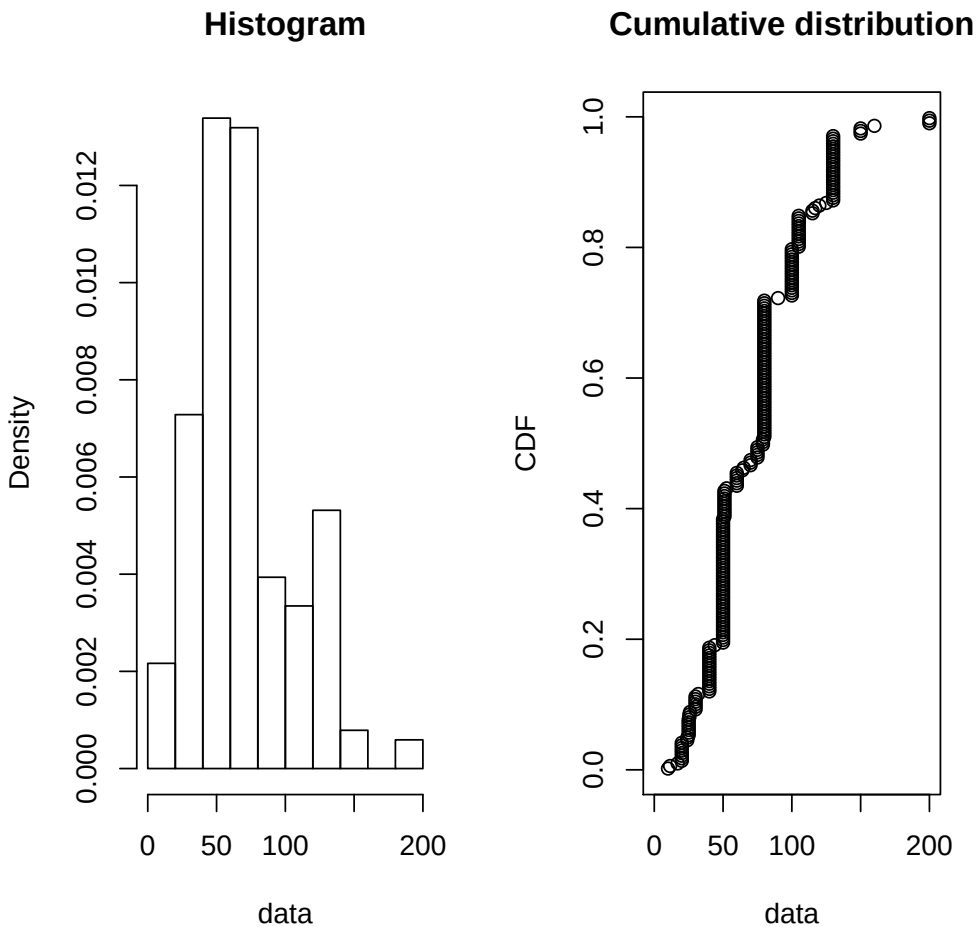
- Choix de distributions candidates parmi un petit nombre de distributions classiques à partir des caractéristiques de la variable étudiée, de l'examen visuel de la distribution empirique et du calcul des coefficients d'asymétrie et de pointicité.
- Ajustement des distributions candidates
- Choix de la distribution la plus adaptée en fonction des exigences de l'utilisateur, notamment à partir de l'examen visuel des ajustements

Notre vision de la démarche de description d'une distribution empirique par une distribution paramétrique

- Choix de distributions candidates parmi un petit nombre de distributions classiques à partir des caractéristiques de la variable étudiée, de l'examen visuel de la distribution empirique et du calcul des coefficients d'asymétrie et de pointicité.
- Ajustement des distributions candidates
- Choix de la distribution la plus adaptée en fonction des exigences de l'utilisateur, notamment à partir de l'examen visuel des ajustements

Cas d'une variable continue : taille de portions de steak haché ingérées par des enfants de moins de 5 ans

```
> data(groundbeef)
> plotdist(groundbeef$serving)
```

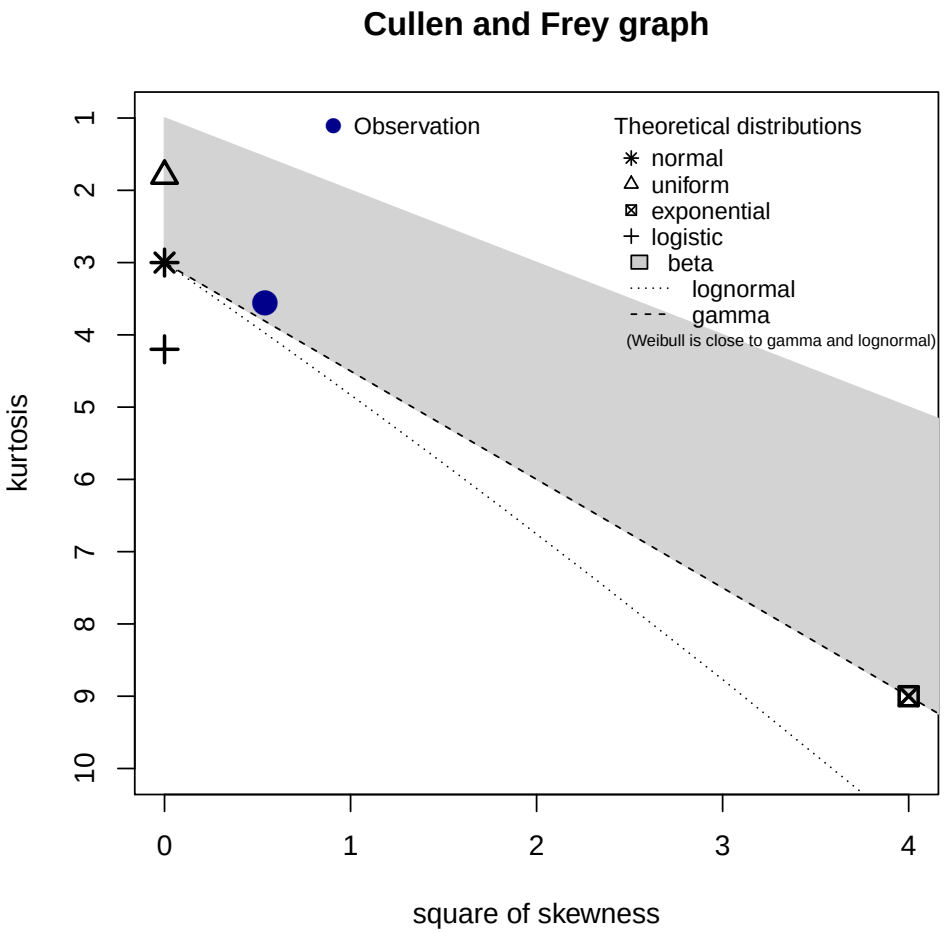


Calcul des coefficients d'asymétrie et de pointicité

- **coefficient d'asymétrie** : $skewness = \frac{\frac{1}{n} \times \sum (x_i - \mu)^3}{\sigma^3}$
  - $skewness = 0$: distribution symétrique
  - $skewness > 0$: asymétrie positive
  - $skewness < 0$: asymétrie négative
- **coefficient de pointicité** : $kurtosis = \frac{\frac{1}{n} \times \sum (x_i - \mu)^4}{\sigma^4}$
  - $kurtosis = 3$: dist. mésokurtique (distribution normale)
  - $kurtosis > 3$: dist. leptokurtique (pointue et à queues épaisses) :
  - $kurtosis < 3$: dist. platikurtique (plate)

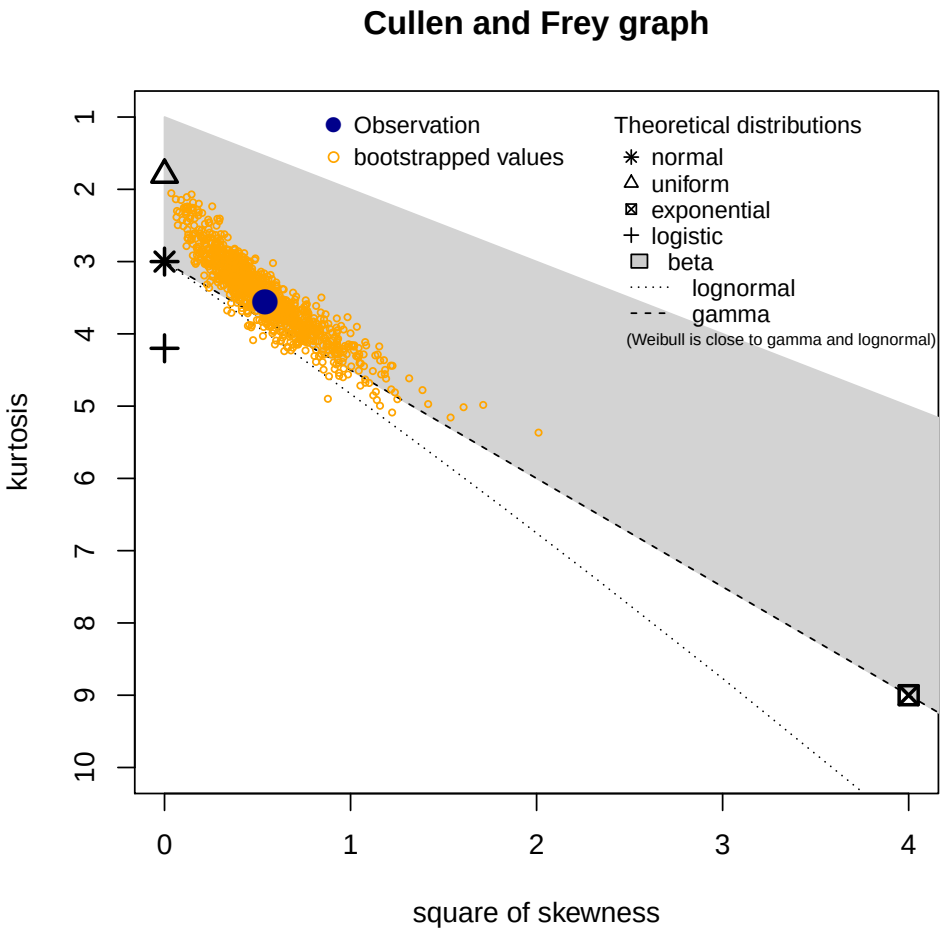
Graphe de Cullen et Frey

```
> descdist(groundbeef$serving)
```



Graphe de Cullen et Frey avec bootstrap

```
> descdist(groundbeef$serving,boot=1000)
```



Ajustement des distributions candidates choisies, par défaut par maximum de vraisemblance

Maximum de vraisemblance

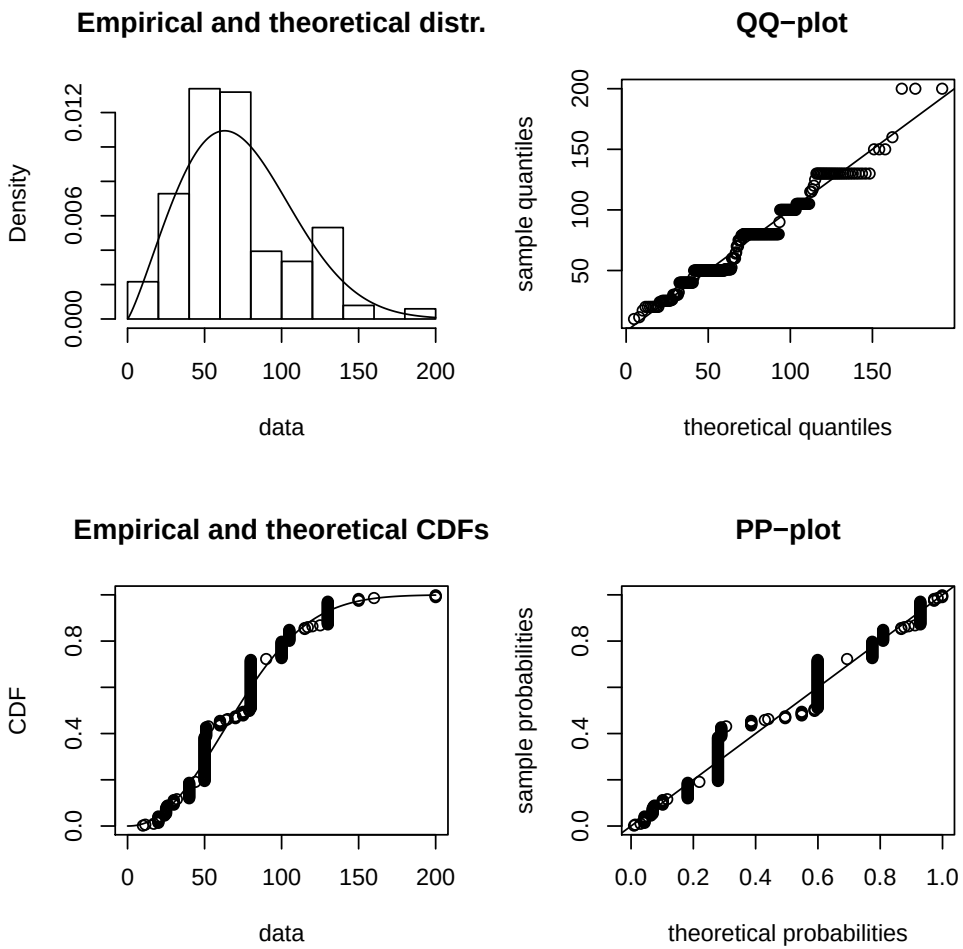
$$\theta_{MLE} = \underset{\theta \in \Theta}{\operatorname{argmax}}\left(\prod_{i=1}^{n} f(x_i|\theta)\right)$$

Exemple : ajustement d'une loi de Weibull

```
> fw <- fitdist(groundbeef$serving,"weibull")
> summary(fw)

Fitting of the distribution ' weibull ' by maximum likelihood
Parameters :
      estimate Std. Error
shape     2.19      0.105
scale    83.35      2.527
Loglikelihood:  -1255   AIC:  2514   BIC:  2522
Correlation matrix:
      shape scale
shape 1.000 0.322
scale 0.322 1.000
```

Graphes d'ajustement

```
> plot(fw)
```



Comparaison de plusieurs ajustements sur la base de la vraisemblance éventuellement pénalisée

Exemple :
logvraisemblance associée aux ajustements des distributions de Weibull, gamma et lognormale.

```
> fg <- fitdist(groundbeef$serving,"gamma")
> fln <- fitdist(groundbeef$serving,"lnorm")
> cbind(Lweibull = fw$loglik, Lgamma = fg$loglik, Llnorm = fln$loglik)
```

```
     Lweibull Lgamma Llnorm
[1,]    -1255  -1254  -1261
```

Dans le cas de la comparaison des distributions non toutes définies par un même nombre de paramètres, utilisation de l'AIC ou du BIC.

Calcul d'autres statistiques d'ajustement

- **distance de Kolmogorov-Smirnov** :
  $\Delta_{KS} = \sup_x(|F_{obs}(x) - F_{theo}(x)|)$
- **distance de Cramér-von Mises** :
  $\Delta^2_{CvM} = n \times \int_{-\infty}^{\infty}(F_{obs}(x) - F_{theo}(x))^2 dx$
  Ces deux distances n'accordent que peu de poids aux queues de distribution : inconvénient en appréciation quantitative des risques
- **distance d'Anderson-Darling** :
  $\Delta^2_{AD} = n \times \int_{-\infty}^{\infty}(F_{obs}(x) - F_{theo}(x))^2 \psi(x) dx$
  avec les poids $\psi(x)$ donnant plus de poids aux queues de distributions : $\psi(x) = (F_{theo}(x) \times (1 - F_{theo}(x)))^{-1}$
  Inconvénient dans le cadre de la comparaison de plusieurs distributions : les poids dépendent de la distribution théorique

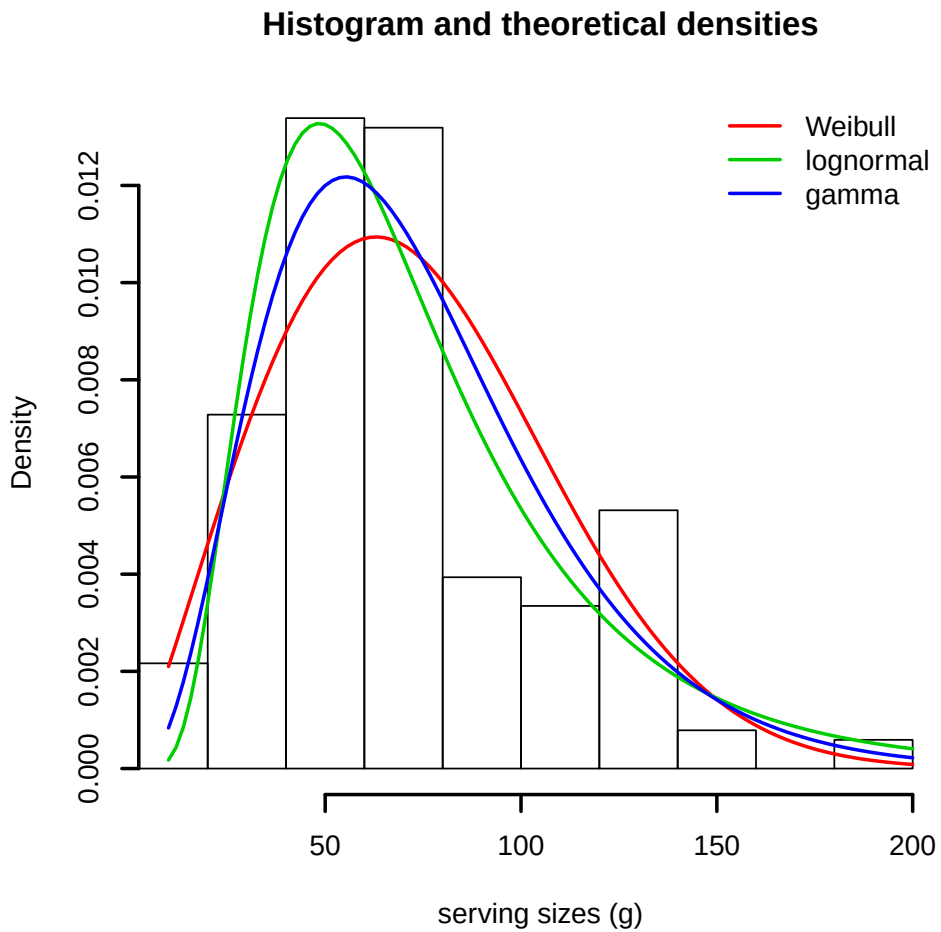
Comparaison des 3 statistiques d'ajustement sur l'exemple

```
> gofstat(fw)

Kolmogorov-Smirnov statistic:  0.14
Cramer-von Mises statistic:  0.684
Anderson-Darling statistic:  3.57

> gofstat(fg)

Kolmogorov-Smirnov statistic:  0.128
Cramer-von Mises statistic:  0.693
Anderson-Darling statistic:  3.57

> gofstat(fln)

Kolmogorov-Smirnov statistic:  0.149
Cramer-von Mises statistic:  0.828
Anderson-Darling statistic:  4.54
```

Choix d'une distribution sur la comparaison systématique d'une seule statistique d'ajustement souvent pratiquée mais non recommandée

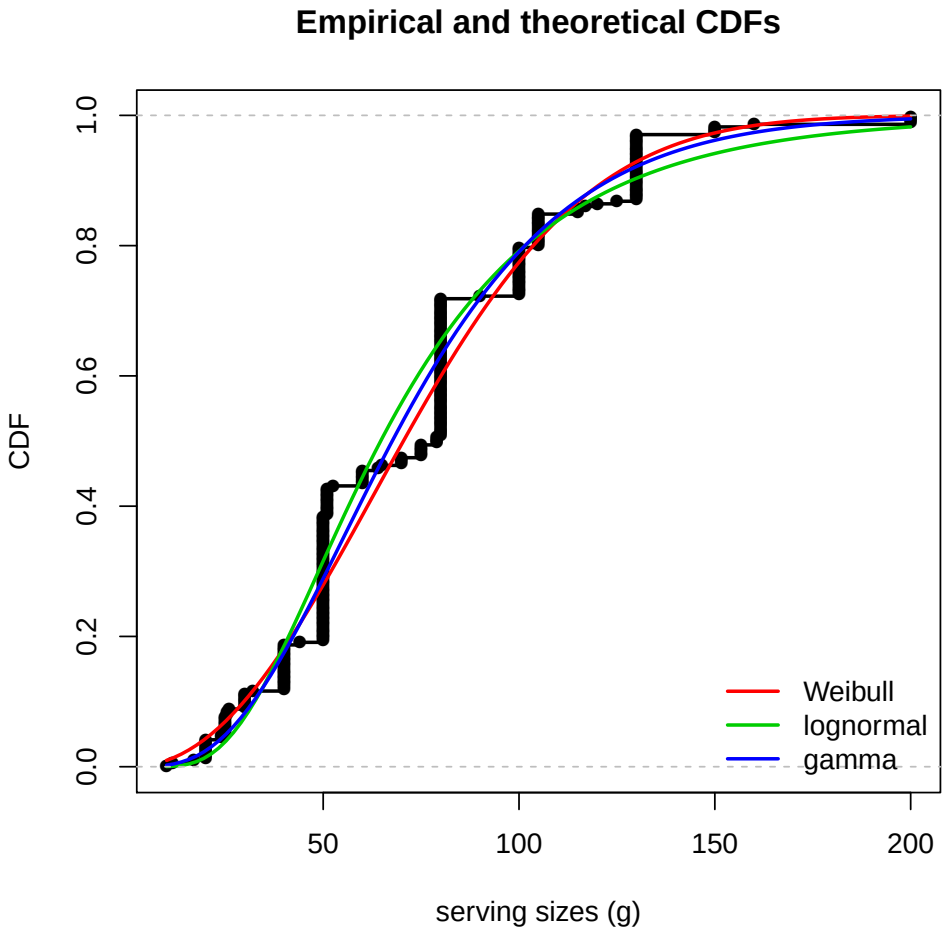
Comparaison graphique de plusieurs ajustements en densité de probabilité avec denscomp

```
> denscomp(list(fw,fln,fg),legendtext=c("Weibull","lognormal","gamma"),
+ xlab="serving sizes (g)",lwd=2,fitlty=1)
```



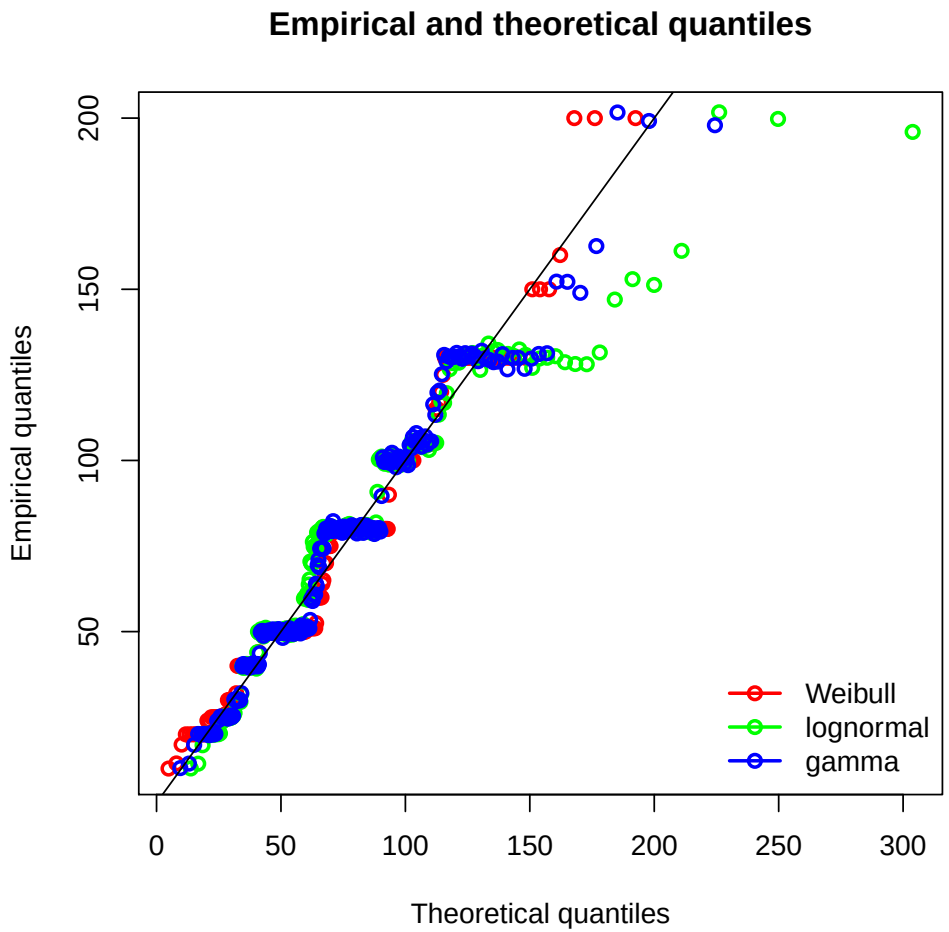
Comparaison graphique de plusieurs ajustements en fréquences cumulées avec cdfcomp

```
> cdfcomp(list(fw,fln,fg),legendtext=c("Weibull","lognormal","gamma"),
+ xlab="serving sizes (g)",lwd=2,lines01=TRUE,fitlty=1)
```



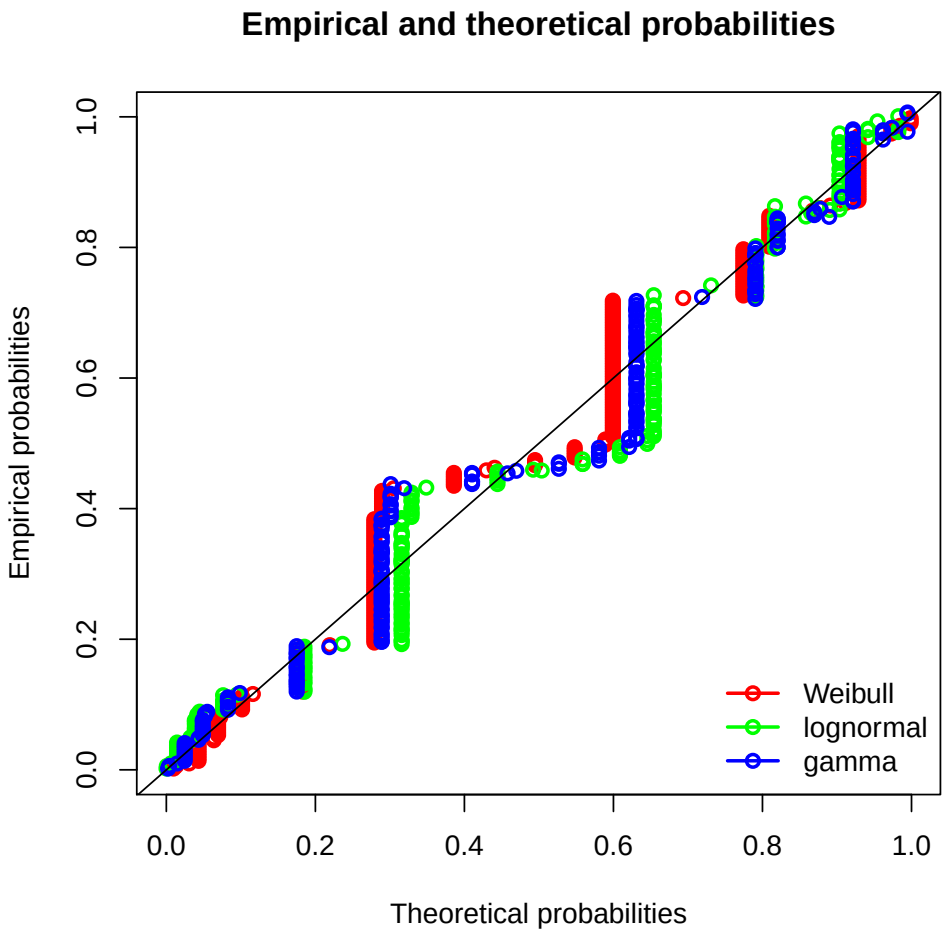
Comparaison graphique de plusieurs ajustements en QQ-plot avec qqcomp

```
> qqcomp(list(fw,fln,fg),legendtext=c("Weibull","lognormal","gamma"),
+ lwd=2,fitcol=c("red","green","blue"))
```



Comparaison graphique de plusieurs ajustements en PP-plot avec ppcomp

```
> ppcomp(list(fw,fln,fg),legendtext=c("Weibull","lognormal","gamma"),
+ lwd=2,fitcol=c("red","green","blue"))
```



Bilan sur les outils offerts par `fitdistrplus` dans le cadre classique

Dans le cas le plus classique d'ajustement par maximum de vraisemblance d'une distribution paramétrique sur des données observées d'une variable quantitative continue, `fitdistrplus` offre

- une fonction d'aide à la caractérisation de la distribution observée et au choix de distributions paramétriques candidates (`plotdist, descdist`)
- une fonction d'ajustement (`fitdist`) accompagnée de fonctions génériques (`plot`, `print`, `summary`, `quantile`)
- des outils complémentaires d'évaluation et de comparaison de qualités d'ajustement
  - statistiques d'ajustement (`gofstat`)
  - graphes d'ajustement (`denscomp`, `cdfcomp`, `qqcomp`,`ppcomp`)

Autres cas envisagés

- Autres méthodes d'ajustement pour les variables quantitative continues
- Autres types de données (variable discrète, variable censurée)

Autres méthodes d'estimation des paramètres dans le cas d'une variable quantitative continue

- Minimisation d'une statistique d'ajustement
- Méthode des moments
- Méthode des quantiles

Choix de distances à minimiser

Distance de Kolmogorov-Smirnov

$$\Delta_{KS} = \sup_x(|F_{obs}(x) - F_{theo}(x)|)$$

Distance de Cramér-von Mises

$$\Delta^2_{CvM} = n\int_{-\infty}^{\infty}(F_{obs}(x) - F_{theo}(x))^2 dx$$

Distance de Anderson-Darling et variantes

$$\Delta^2_{AD} = n\int_{-\infty}^{\infty}(F_{obs}(x) - F_{theo}(x))^2 \psi(x) dx$$

avec $\psi(x) = (F_{theo}(x) \times (1 - F_{theo}(x)))^{-1}$

ou (ADR) :$\psi(x) = (1 - F_{theo}(x))^{-1}$

ou (ADL) : $\psi(x) = F_{theo}(x)^{-1}$

ou (AD2R) : $\psi(x) = (1 - F_{theo}(x))^{-2}$

ou (AD2L) : $\psi(x) = F_{theo}(x)^{-2}$

ou (AD2) : $\psi(x) = F_{theo}(x)^{-2} + (1 - F_{theo}(x))^{-2}$

Méthode des moments

Principe : sur la base d'un nombre k de moments égal au nombre de paramètres de la loi ajustée, calculer les paramètres de façon à ce qu'il y ait égalité des moments empiriques et théoriques

Méthode des moments

$\theta \in \Theta$ tel que $E(x^k) = \frac{1}{n}\sum_{i=1}^{n} x_i^k$ pour $k = 1, \ldots, p$

Calcul analytique lorsqu'il est disponible, ou numérique par minimisation de la distance quadratique entre les moments empiriques et théoriques

Méthode des quantiles

Principe : sur la base d'un nombre k de quantiles égal au nombre de paramètres de la loi ajustée, calculer les paramètres de façon à ce qu'il y ait égalité des quantiles empiriques et théoriques

Méthode des quantiles

$\theta \in \Theta$ tel que $F^{-1}_{theo}(p_k) = Q_{n,p_k}$ pour $k = 1, \ldots, p$

Calcul numérique par minimisation de la distance quadratique entre les quantiles empiriques Q_{n,p_k} et théoriques $F^{-1}_{theo}(p_k)$ pour les k probabilités p_k données

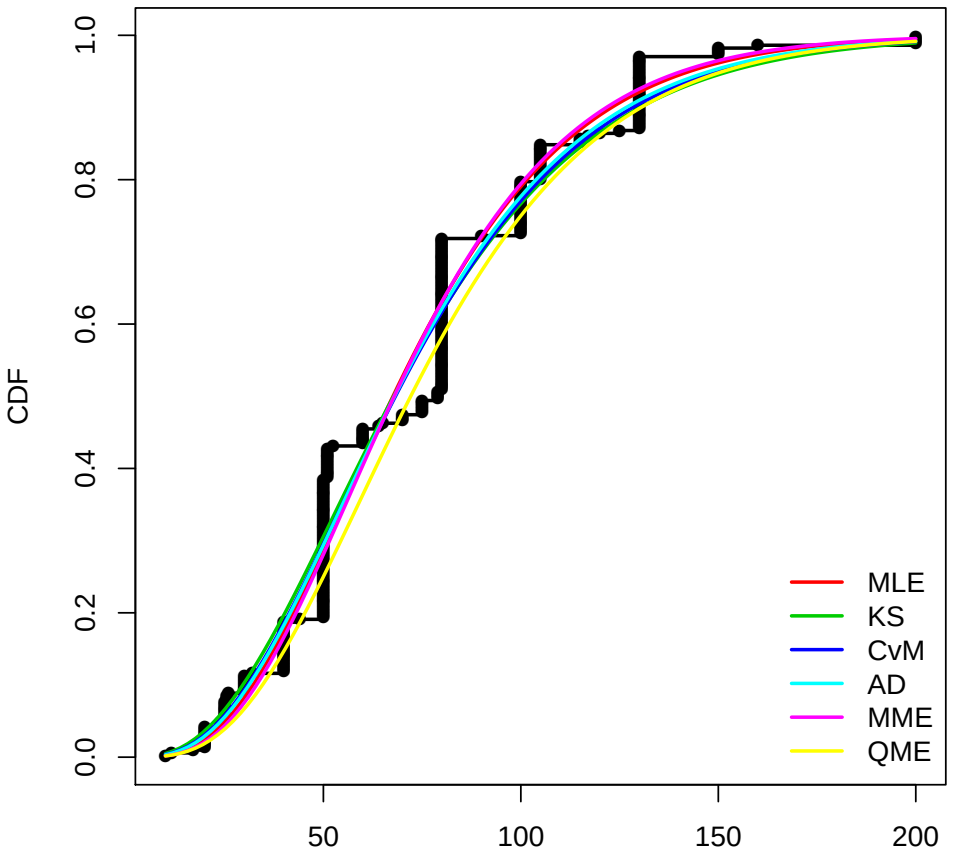
Comparaison des diverses méthodes d'ajustement sur un exemple

Ajustement, par diverses méthodes, d'une loi gamma sur les données de tailles de portions de steak haché consommées par des enfants de moins de 5 ans

```
> fgMLE <- fitdist(groundbeef$serving,"gamma")
> fgKS <- fitdist(groundbeef$serving,"gamma",method="mge",gof="KS")
> fgCvM <- fitdist(groundbeef$serving,"gamma",method="mge",gof="CvM")
> fgAD <- fitdist(groundbeef$serving,"gamma",method="mge",gof="AD")
> fgMME <- fitdist(groundbeef$serving,"gamma",method="mme")
> fgQME <- fitdist(groundbeef$serving,"gamma",method="qme",
+     probs=c(0.25,0.75))
> cdfcomp(list(fgMLE,fgKS,fgCvM,fgAD,fgMME,fgQME),main="",lwd=2,
+     legendtext = c("MLE","KS","CvM","AD","MME","QME"),fitlty=1)
```

Comparaison des diverses méthodes d'ajustement sur un exemple

Ajustement, par diverses méthodes, d'une loi gamma sur les données de tailles portions de steak haché consommées par des enfants de moins de 5 ans



Autres types de donnés envisagés

- Cas des variables discrètes :
  ajustement par maximum de vraisemblance avec adaptation des graphes d'ajustement et des statistiques d'ajustement (utilisation de la statistique du χ^2).
  Utilisation des mêmes fonctions pour certaines avec l'option `discrete = TRUE`.
- **Cas des variables censurées :**
  adaptation du codage des données, de la maximisation de la vraisemblance et des graphes d'ajustement.
  Utilisation de fonctions spécifiques : `plotdistcens`, `fitdistcens`, `cdfcompdens`

Autres types de donnés envisagés

- Cas des variables discrètes :
  ajustement par maximum de vraisemblance avec adaptation des graphes d'ajustement et des statistiques d'ajustement (utilisation de la statistique du χ^2).
  Utilisation des mêmes fonctions pour certaines avec l'option `discrete = TRUE`.
- **Cas des variables censurées :**
  adaptation du codage des données, de la maximisation de la vraisemblance et des graphes d'ajustement.
  Utilisation de fonctions spécifiques : `plotdistcens`, `fitdistcens`, `cdfcompdens`

Codage des données censurées

Deux colonnes : `left` et `right`

- censure à gauche : `left = NA`
- censure à droite : `right = NA`
- censure par intervalle :
  `[left ; right] = [borne inf ; borne sup]`
- donnée non censurée : `left = right = valeur`

Exemple de jeu de données censurées

Exemple : Tolérance à la salinité (LC50) de macroinvertébrés marins.

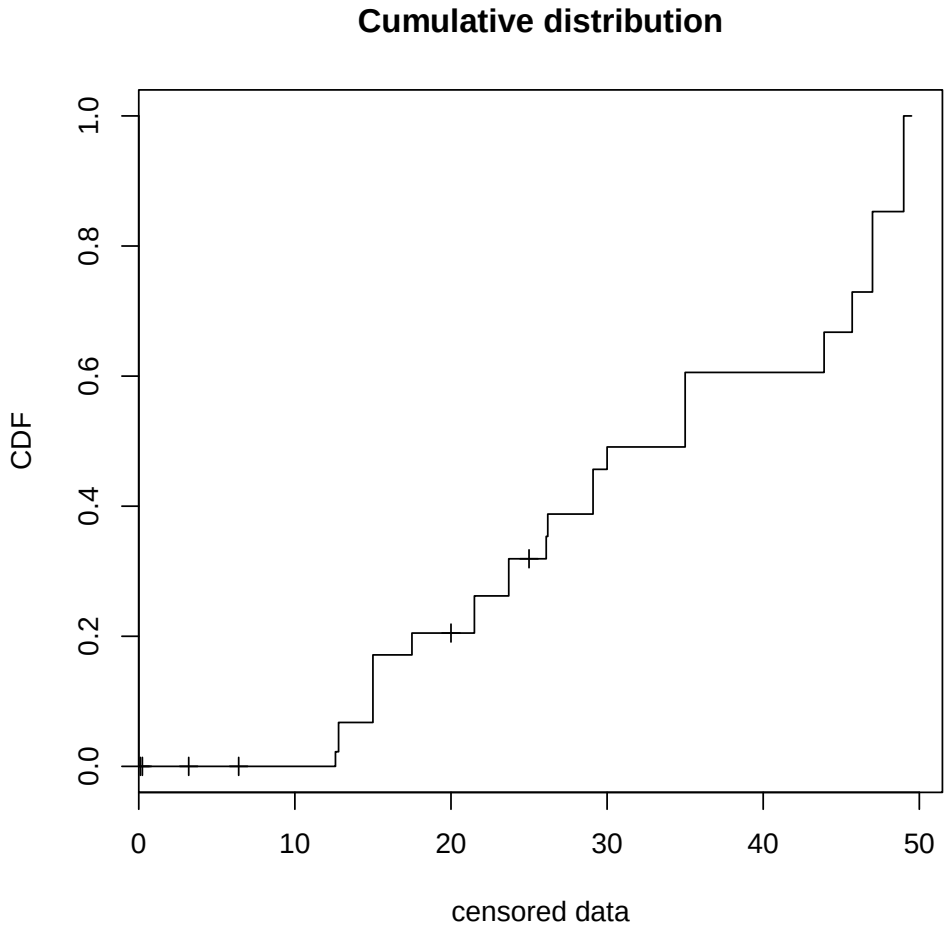
```
> data(salinity)
> salinity[1:10,]
```

```
   left right
1  20.0    NA
2  20.0    NA
3  20.0    NA
4  20.0    NA
5  20.0    NA
6  21.5  21.5
7  15.0  30.0
8  20.0  25.0
9  23.7  23.7
10 25.0    NA
```

Examen de la distribution empirique

Graphe de Turnbull : estimation non paramétrique de la fonction de répartition de la distribution empirique (package `survival`)

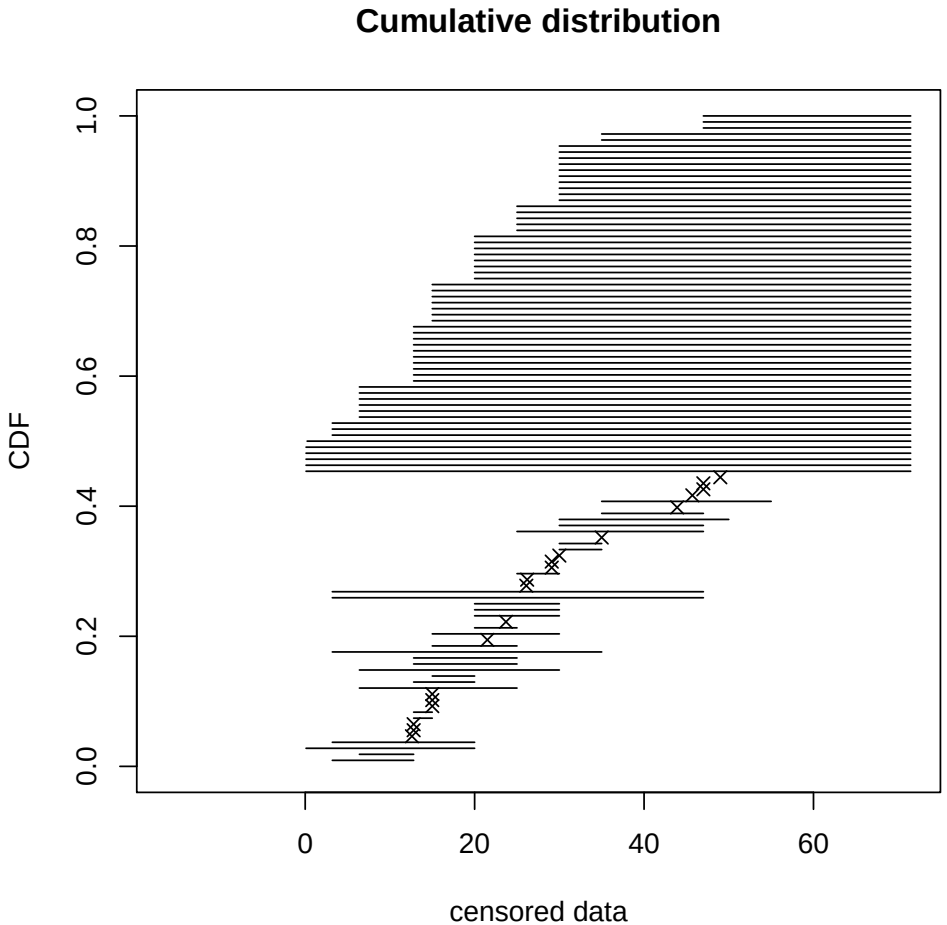
```
> plotdistcens(salinity)
```



Représentation parfois plus parlante mais moins rigoureuse

Graphe représentant les intervalles ou valeurs classés (la difficulté résidant dans ce classement)

```
> plotdistcens(salinity,Turnbull=FALSE)
```



Maximum de vraisemblance pour des données censurées

Définition de la vraisemblance pour des données censurées

$$L(\theta) = \prod_{i=1}^{N_{nonC}} f(x_i|\theta) \times \prod_{j=1}^{N_{leftC}} F(x_j^{upper}|\theta)$$
$$\times \prod_{k=1}^{N_{rightC}} (1 - F(x_k^{lower}|\theta)) \times \prod_{m=1}^{N_{intC}} (F(x_m^{upper}|\theta) - F(x_j^{lower}|\theta))$$

avec

x_i les N_{nonC} observations non censurées,

x_j^{upper} les bornes supérieures définissant les N_{leftC} observations censurées à gauche,

x_k^{lower} les bornes inférieures définissant les N_{rightC} observations censurées à droite,

$[x_m^{lower}; x_m^{upper}]$ les intervalles définissant les N_{intC} observations censurées par intervalle,

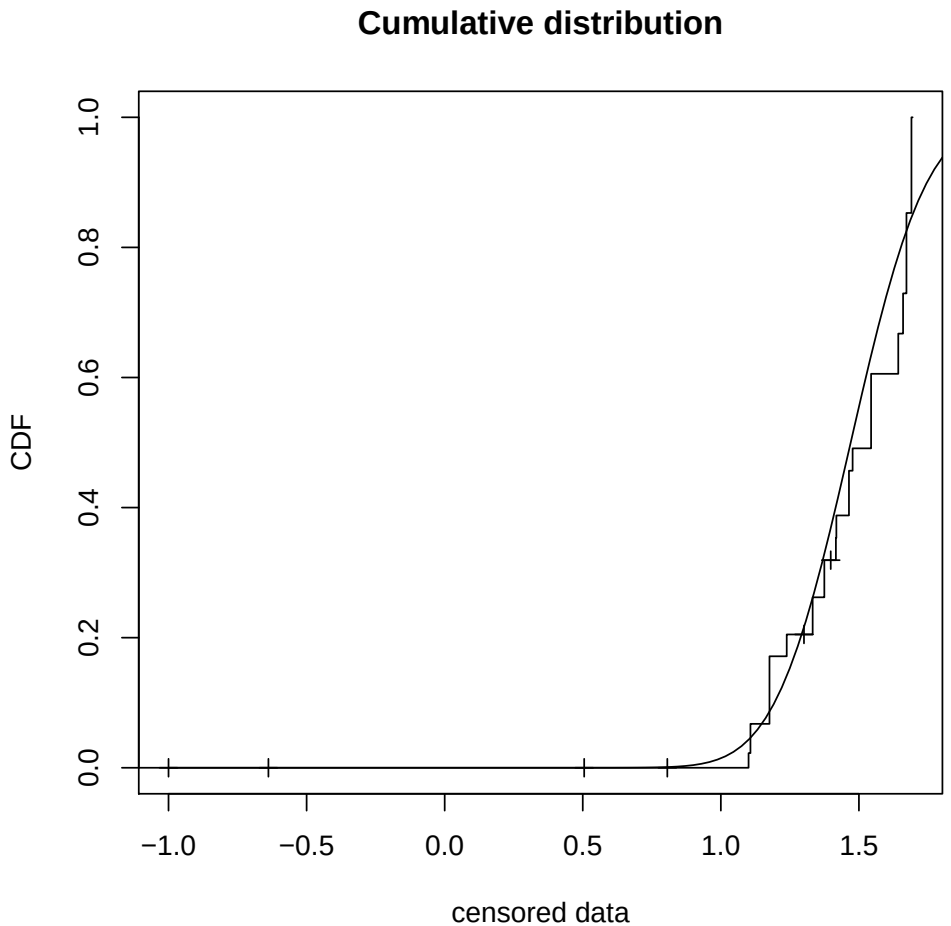
F la fonction de répartition de la distribution paramétrique.

Exemple d'ajustement d'une loi normale

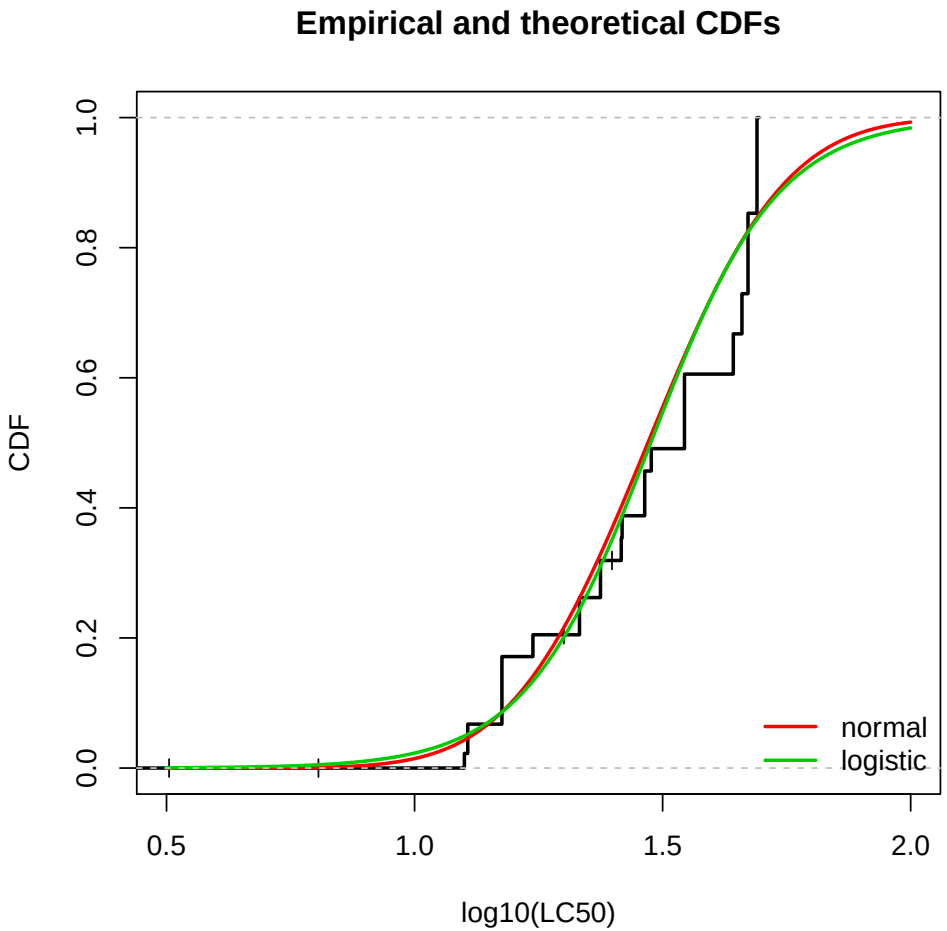
```
> f <- fitdistcens(log10(salinity),"norm")
> summary(f)

FITTING OF THE DISTRIBUTION ' norm ' BY MAXIMUM LIKELIHOOD ON CENSORED D
PARAMETERS
     estimate Std. Error
mean    1.470     0.0282
sd      0.215     0.0237
Loglikelihood:  -61.8   AIC:  128   BIC:  133
Correlation matrix:
      mean    sd
mean 1.000 0.294
sd   0.294 1.000
```

Graphe d'ajustement

```
> plot(f)
```



Comparaison de l'ajustement de plusieurs lois à l'aide de `cdfcompdens`



Incertitude sur les paramètres estimés

- Procédure de bootstrap
- Utilisation des échantillons bootstrap dans le cadre de simulations de Monte Carlo à deux dimensions

Estimation de l'incertitude sur les paramètres par bootstrap

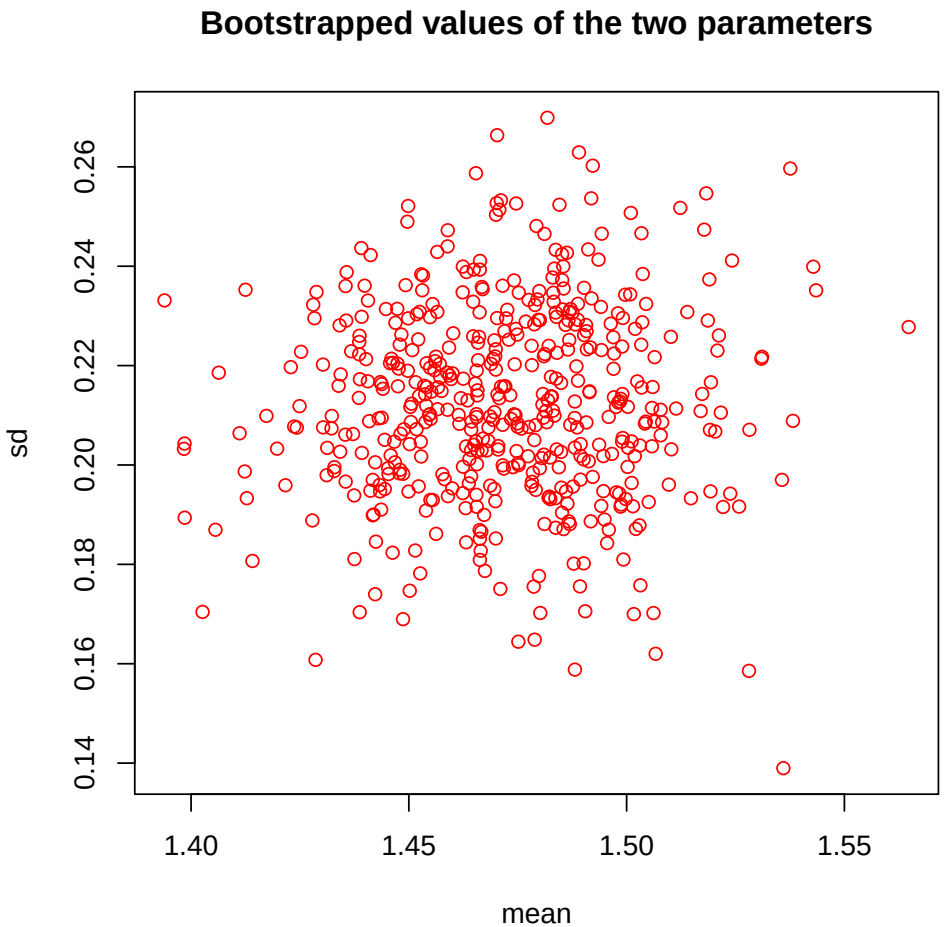
Calcul des intervalles de confiance par bootstrap

- paramétrique (par défaut) ou non paramétrique pour les données non censurées
  `bootdist()`
- non paramétrique pour les données censurées
  `bootdistcens()`

```
> fln <- fitdistcens(log10(salinity),"norm")
> bln <- bootdistcens(fln,niter=501)
> summary(bln)

Nonparametric bootstrap medians and 95% percentile CI
     Median 2.5% 97.5%
mean  1.472 1.42 1.524
sd    0.212 0.17 0.252
```

Représentation de l'échantillon bootstrap

```
> plot(bln,col="red")
```



Calcul de l'incertitude sur une fonction des paramètres

Cas particulier très utile en écotoxicologie : estimation du quantile à 5% de la distribution de sensibilité des espèces à un toxique (SSD) (HC5 : concentration protégeant 5% des espèces) avec son intervalle de confiance à 95%

```
> fln <- fitdistcens(log10(salinity),"norm")
> HC5log10 <- quantile(fln,probs = 0.05,bootstrap=TRUE)

Estimated quantiles for each specified probability
  prob=0.05
1      1.12

two-sided 95% CI of each quantile
      prob=0.05
2.5%       1.05
97.5%      1.19
```

Monte Carlo à deux dimensions

Utilisation de l'échantillon bootstrap dans le cadre de simulations de Monte Carlo à deux dimensions,
visant à séparer la variabilité et l'incertitude
dans les appréciations quantitatives des risques
Lien avec le package `mc2d`
développé par Régis Pouillot (FDA, Washington)

La variabilité

- résulte de l'**hétérogénéité** des composantes du système à modéliser.
- **ne peut être réduite** sans modifier ce système.
- est liée au caractère aléatoire (stochasticité) des phénomènes étudiés.
- est parfois appelée incertitude de type A (**"aleatory uncertainty"**) ou incertitude objective.
- correspond en langage courant à *" Je sais que cela ne se passe pas toujours de la même façon".*

Exemples en écotoxicologie : variabilité inter-individus ou inter-espèces sur la réponse à la contamination, variabilité géographique de la contamination, etc.

L'incertitude

- résulte du **manque de connaissance**
- **peut être réduite** par l'acquisition de nouvelles connaissances.
- est liée au niveau d'ignorance du modélisateur.
- est parfois appelée incertitude de type B (**"epistemic uncertainty"**) ou incertitude subjective.
- correspond en langage courant à *" Je ne sais pas exactement comment cela se passe".*

Exemples : incertitude sur les paramètres des modèles liée au manque de données, à l'incertitude de mesure, à la censure éventuelle de données, à l'incertitude sur le modèle lui-même.

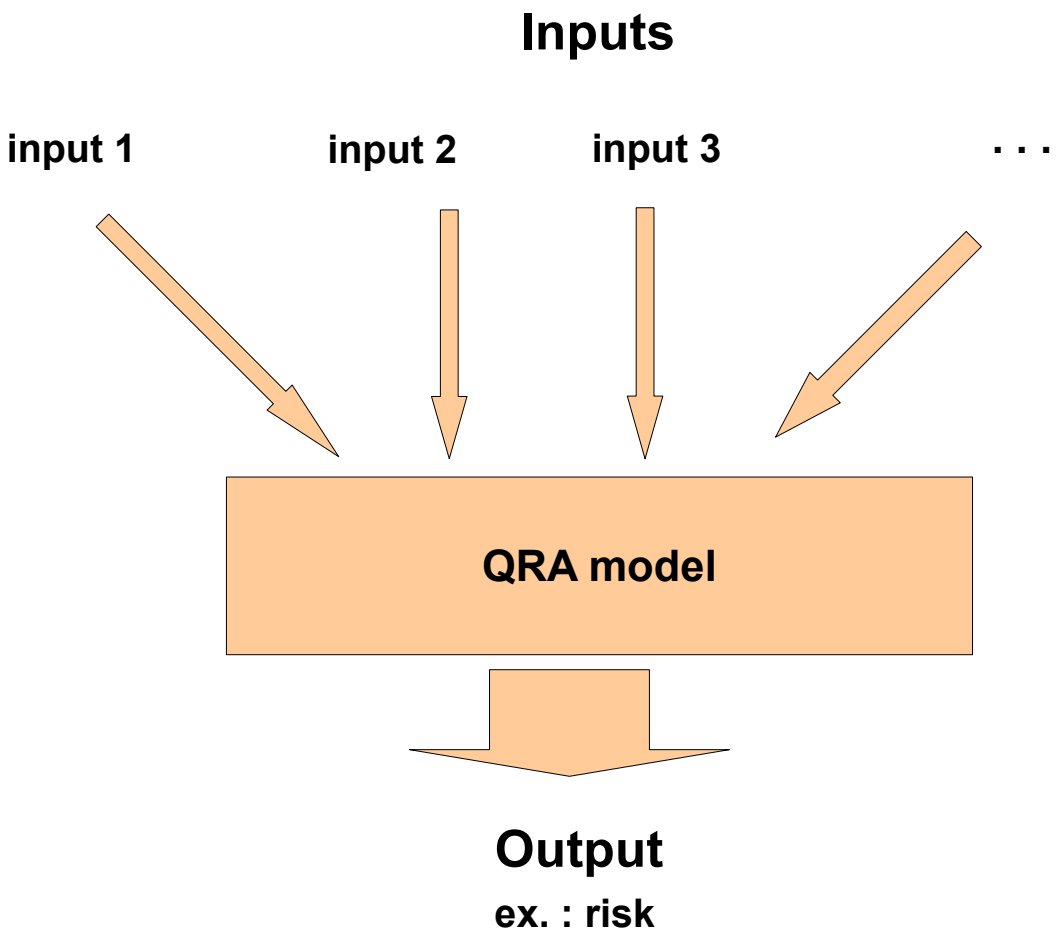
Modélisation hiérarchique de l'incertitude et de la variabilité

- La **variabilité** est décrite par une **distribution de probabilité** caractérisée par des **paramètres**.
  Ex. : NEC (no effect concentration : seuil d'effet biologique)

$$NEC|M_{NEC}, S_{NEC} \sim lognormale(M_{NEC}, S_{NEC})$$

- Ces **paramètres** peuvent eux-mêmes être supposés **incertains** et caractérisés par des **distributions de probabilité** obtenues
  - par inférence bayésienne
  - ou à l'aide de techniques de rééchantillonnage (bootstrap)

Modèle d'appréciation quantitative des risques

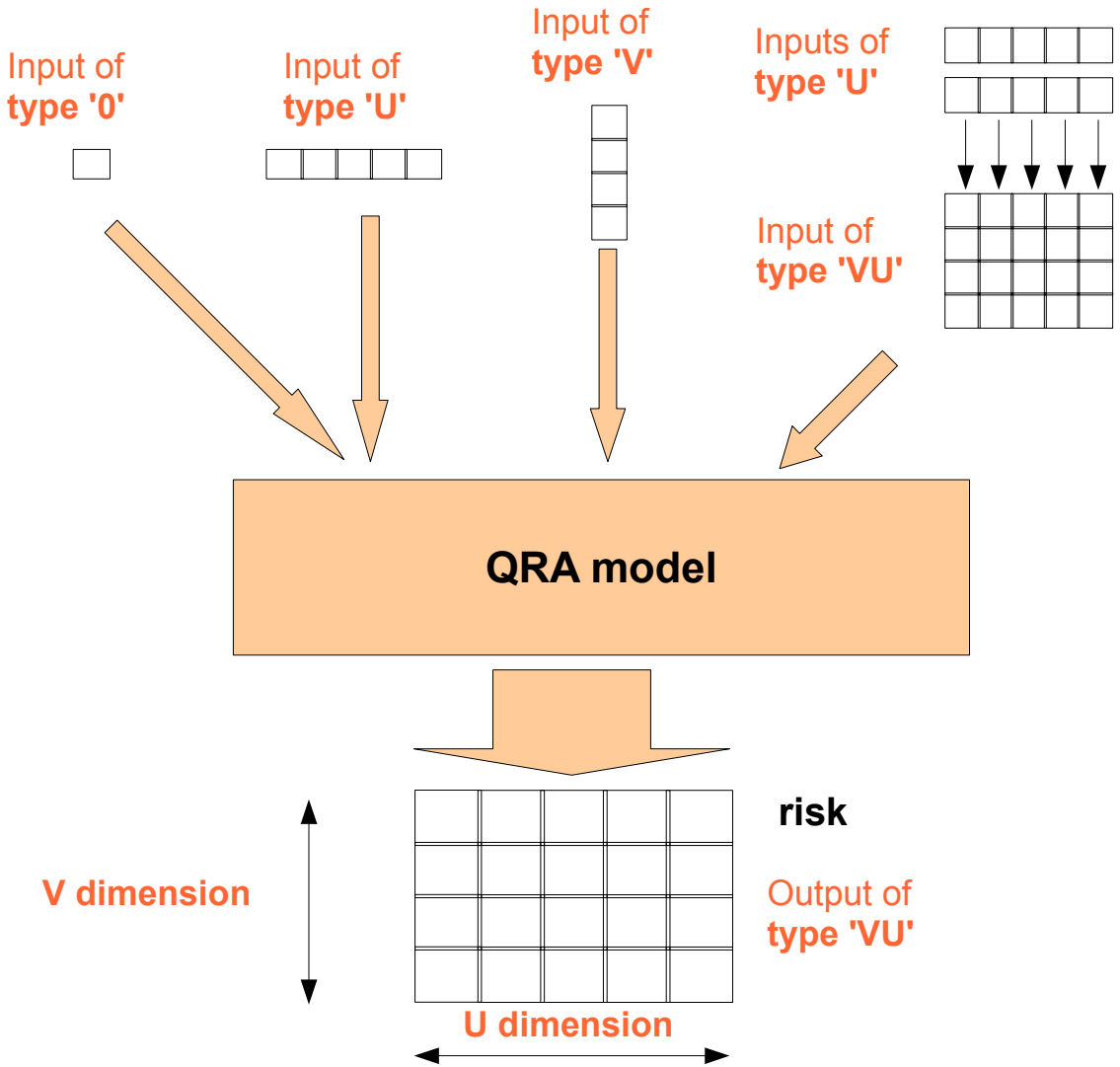


Différents types d'entrées d'un modèle

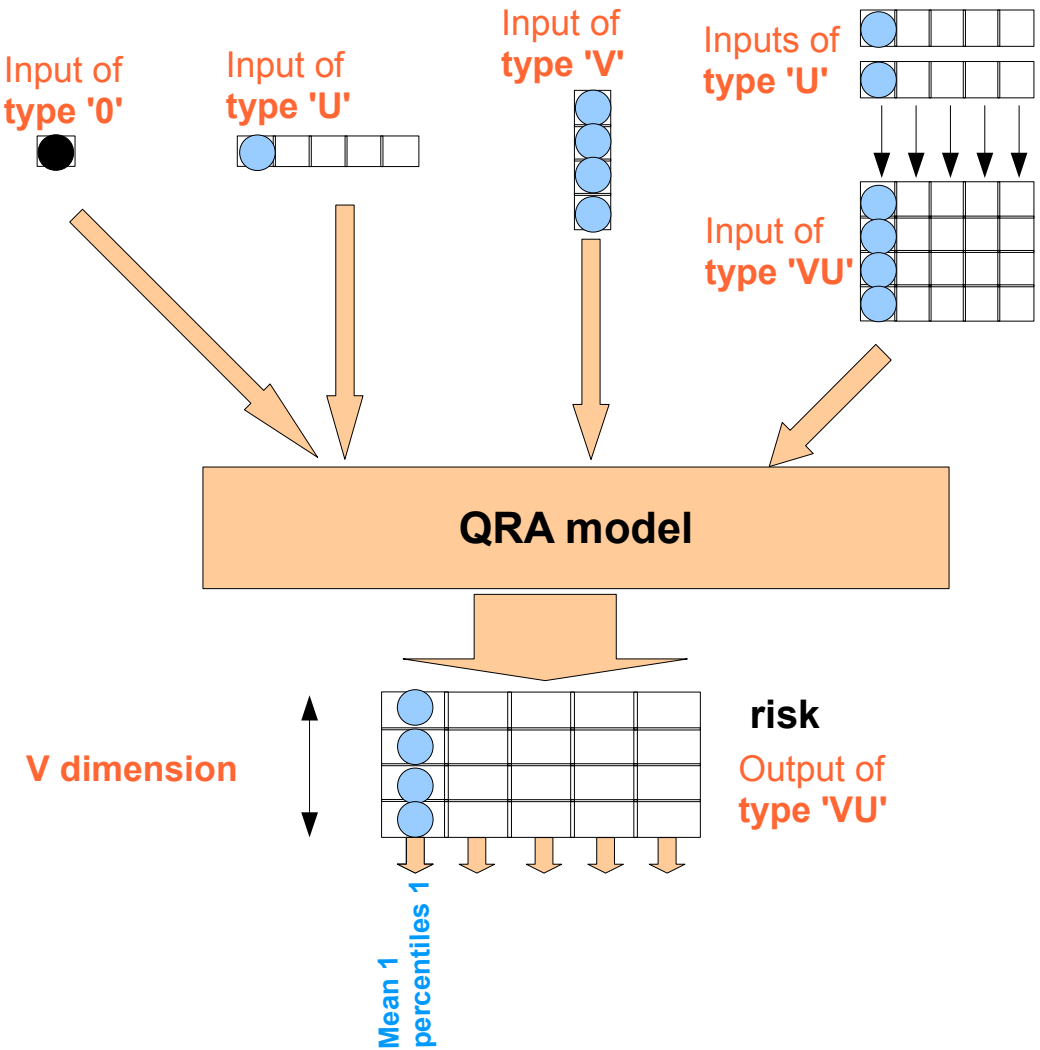
Formalisme adopté dans le package R `mc2d`

- Entrée suposée constante et connue (`type="0"`)
- Entrée supposée uniquement incertaine (variabilité négligée)(`type="U"`)
- Entrée supposée uniquement variable (incertitude négligée)(`type="V"`)
- Entrée supposée variable et incertaine (avec modélisation hiérarchique de l'incertitude et de la variabilité telle que décrite précédemment) (`type="VU"`)

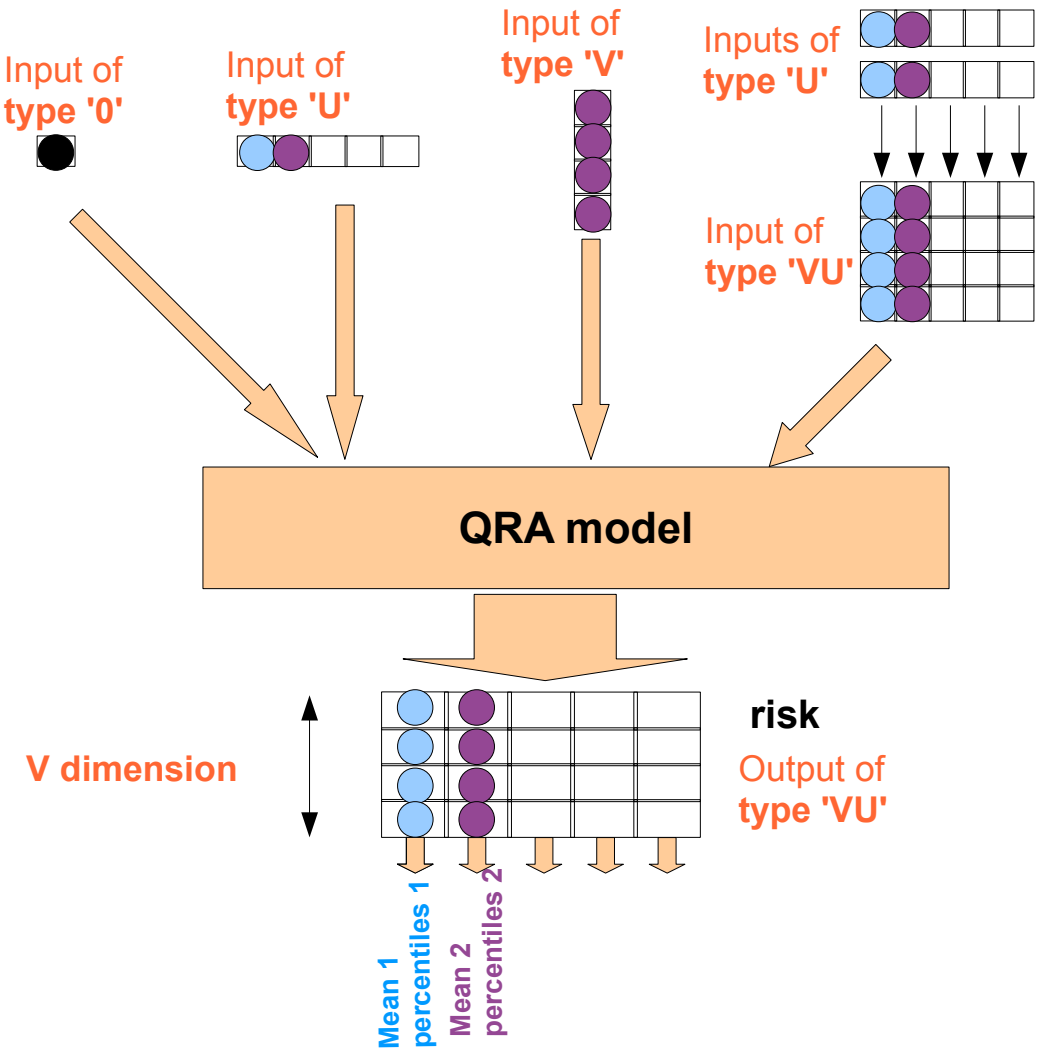
Transferts de l'incertitude et de la variabilité



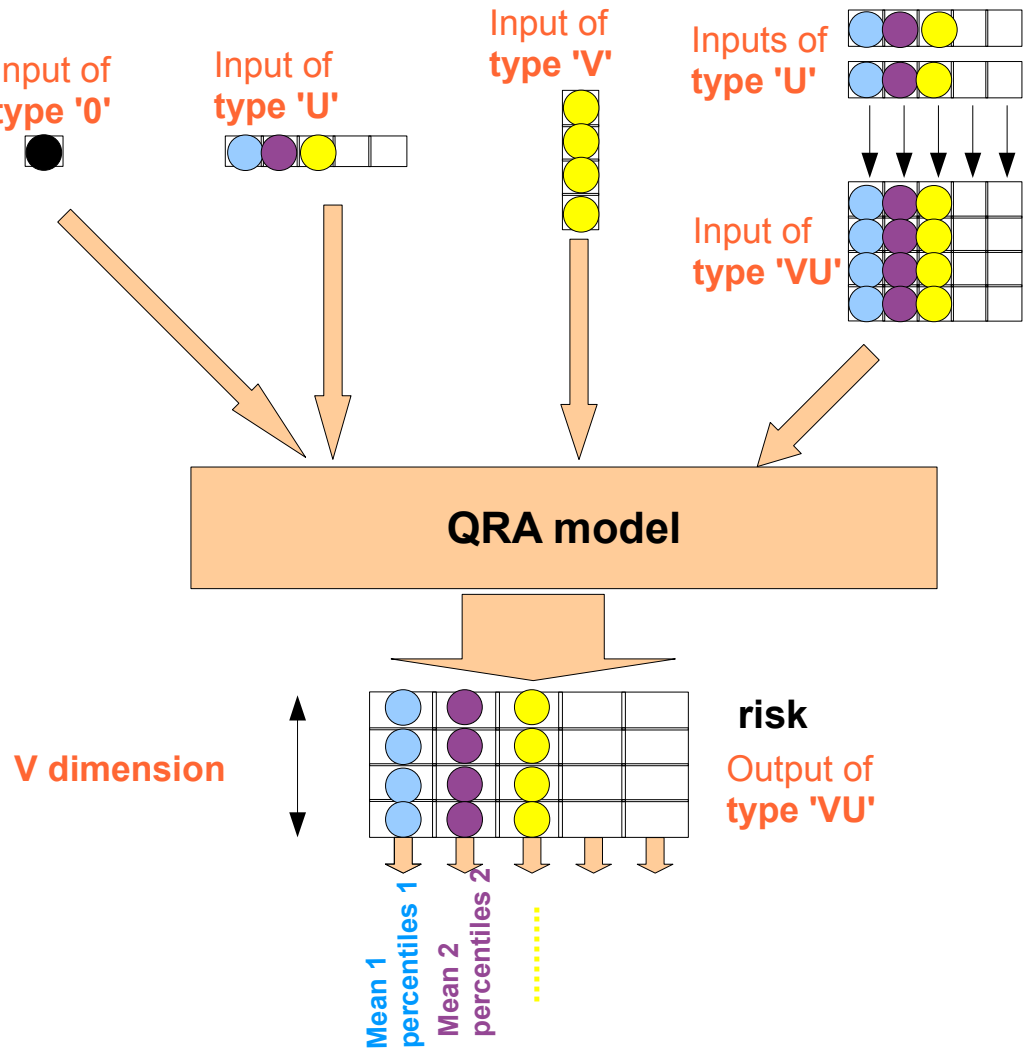
Première simulation dans la dimension d'incertitude



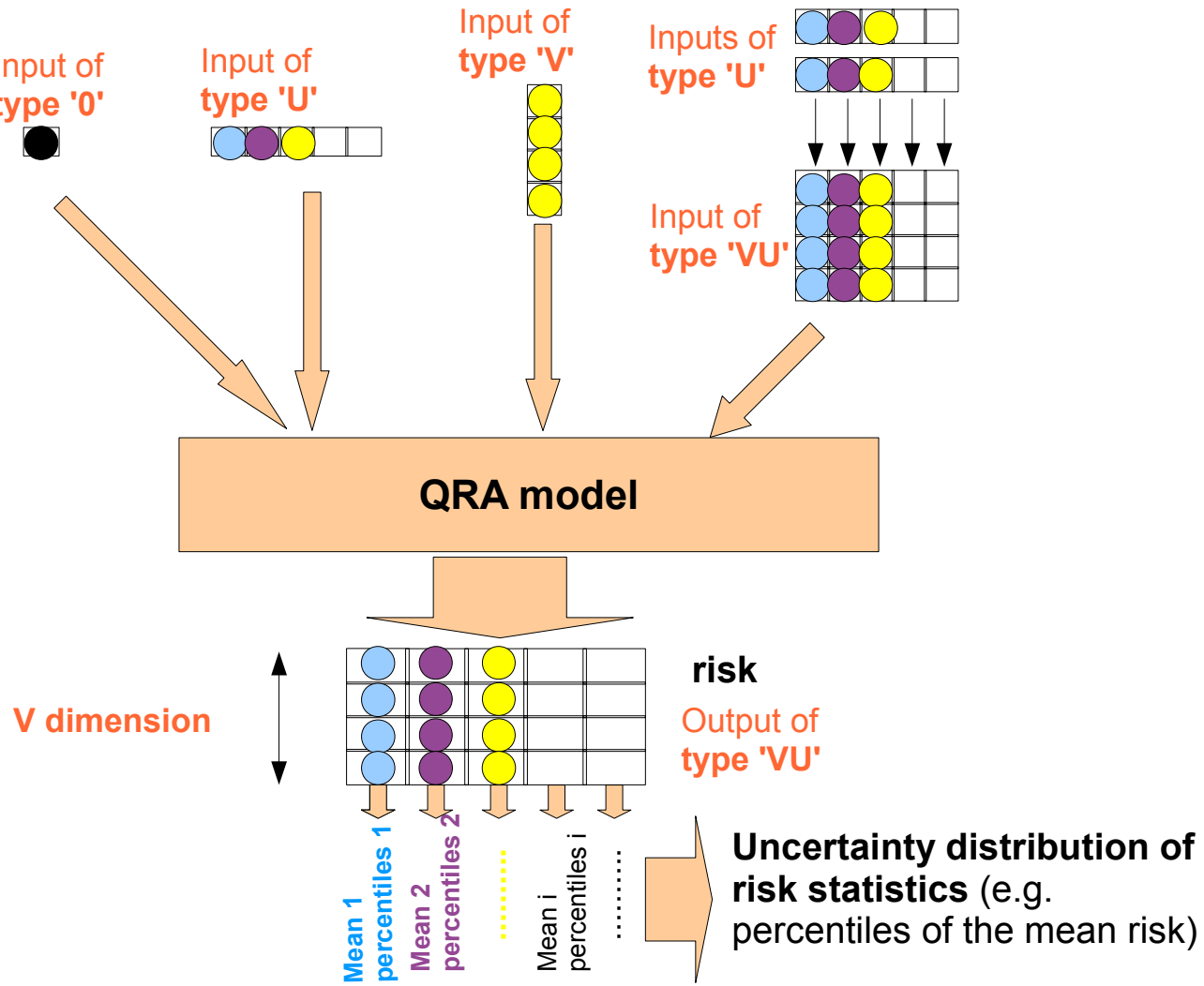
Seconde simulation dans la dimension d'incertitude



Troisième simulation dans la dimension d'incertitude



Sortie des simulations à 2 dimensions



Utilisation des échantillons bootstrap dans le cadre de simulations de Monte Carlo à deux dimensions

Exemple : simulation du risque de SHU associé à l'ingestion par des enfants de moins de 5 ans d'une portion crue de steak haché issu du lot contaminé dont les données ont été analysées précédemment.

- Chargement de la librairie `mc2d`

  ```
  > require(mc2d)
  ```

- Définition des dimensions de variabilité et d'incertitude

  ```
  > ndvar(1001)

  [1] 1001

  > ndunc(1001)

  [1] 1001
  ```

Définition de la taille de la portion ingérée, **variable** et **incertaine**

- Ajustement aux données de tailles avec `fitdistrplus`

```
> fg <- fitdist(groundbeef$serving,"gamma")
> bootg <- bootdist(fg,niter = ndunc())
```

- Simulation des paramètres **incertains** de la loi gamma ajustée

```
> shape.S <- mcdata(bootg$estim$shape, type="U")
> rate.S <- mcdata(bootg$estim$rate, type="U")
```

- Simulation des tailles de portions (**incertaines** et **variables**)

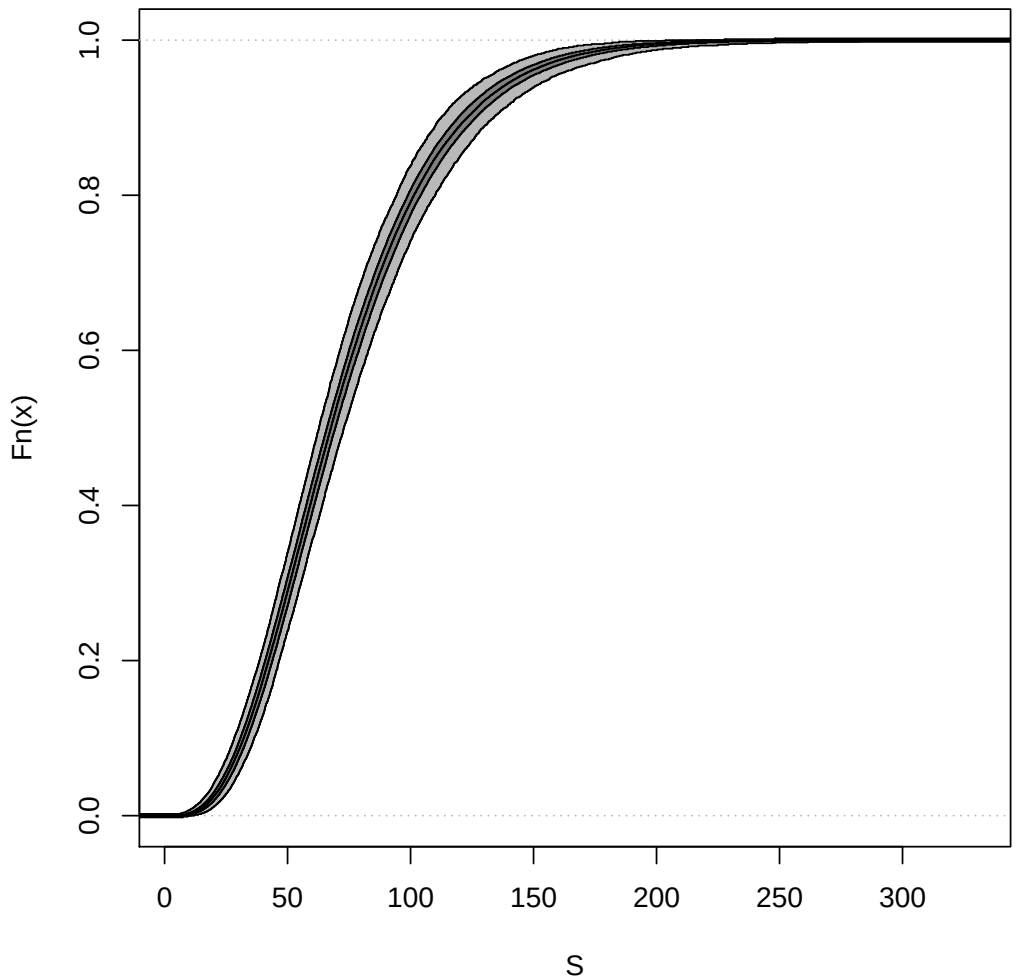
```
> S <- mcstoc(rgamma, type="VU", shape=shape.S, rate=rate.S)
```

Résumé des tailles de portion simulées

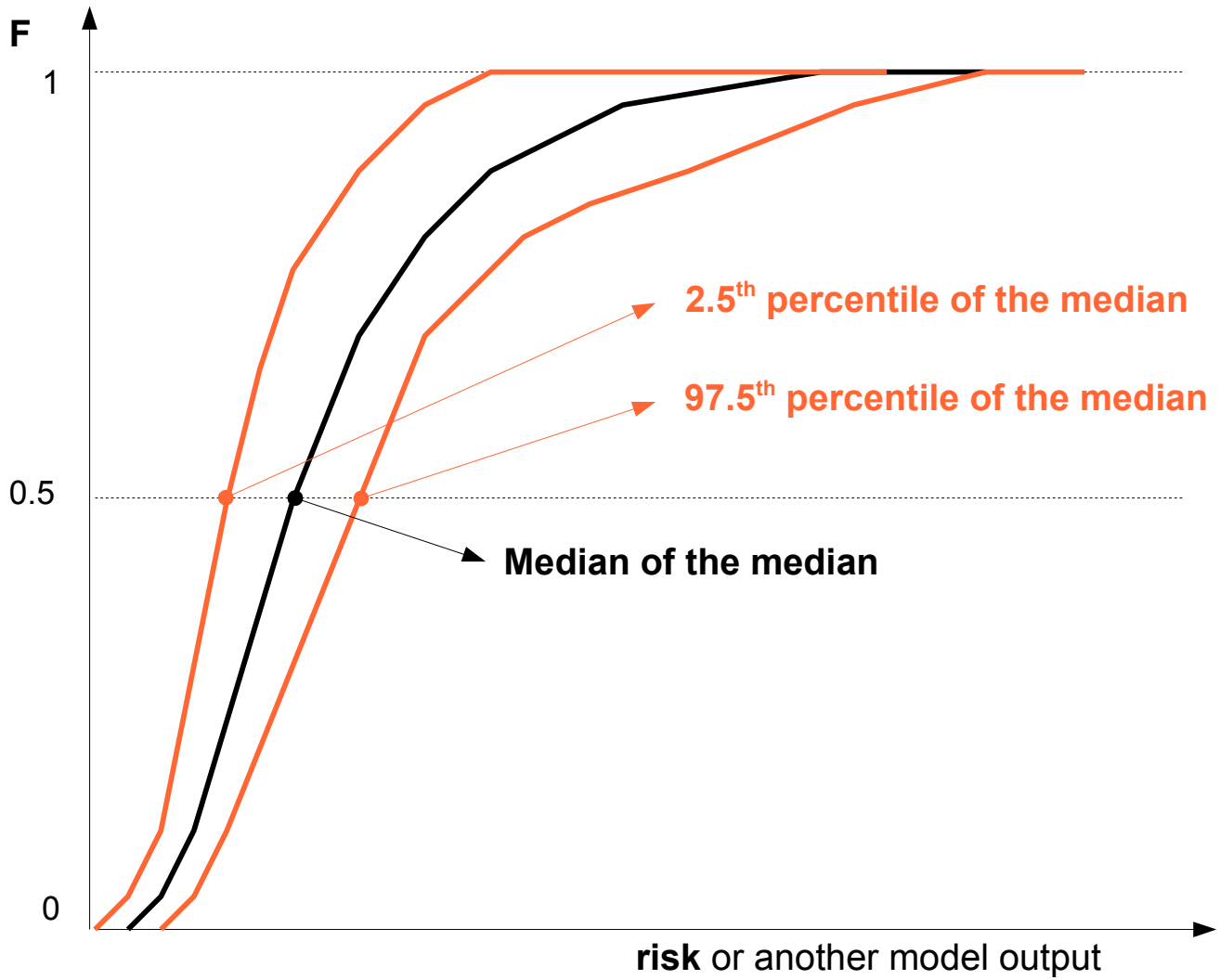
```
> summary(S)

node :
        mean   sd    Min 2.5%  25%  50%   75% 97.5% Max  nsv Na's
median 73.6 36.8   7.19 20.1 46.7 67.6  93.9   161 248 1001    0
mean   73.7 36.8   7.24 20.3 46.8 67.7  94.1   161 254 1001    0
2.5%   68.7 32.5   2.88 16.7 42.9 62.8  87.2   145 204 1001    0
97.5%  79.2 41.2 12.09 24.4 51.0 72.9 101.1   179 331 1001    0
```

Tracé de la fonction de répartition de la taille de portion avec incertitude (bandes à 50% et 95%)

```
> plot(S)
```



Lecture de ce type de graphe



Définition de la concentration en *Escherichia coli* O157 :H7, du nombre de cellules ingérées et du risque de SHU associé

- Ajustement aux données de dénombrement sur boîte de Petri (sur 0.1 g) et bootstrap avec `fitdistrplus`

```
> nbcol <- c(rep(0,24),rep(1,14),rep(2,6),rep(4,1))
> fp <- fitdist(nbcol,"pois")
> bootp <- bootdist(fp,niter  = ndunc())
```

- Définition de la concentration par g (**incertaine**) à partir de l'échantillon bootstrap du paramètre λ de la loi de Poisson estimé

```
> C <- mcdata(bootp$est$lambda,type="U")*10
```

- Définition de la dose en nombre de cellules ingérées

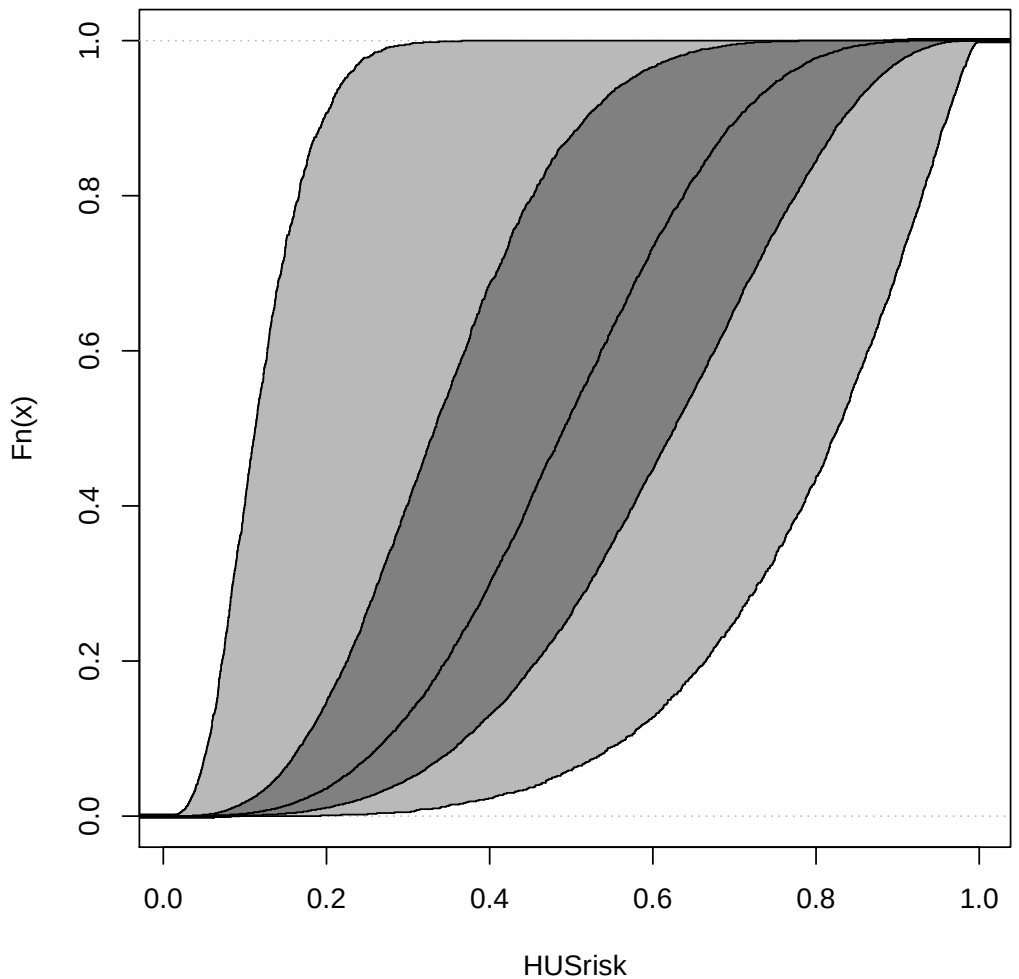
```
> D <- mcstoc(rpois, type="VU", lambda=S*C)
```

- Calcul du risque de SHU à l'aide d'un modèle dose-réponse à un paramètre supposé **incertain**

```
> r <- mcstoc(rpert, type="U", min=1e-4, mode=1.2e-3, max=5e-3)
> HUSrisk <- 1-(1-r)^D
```

Tracé de la fonction de répartition du risque individuel simulé avec incertitude (bandes à 50% et 95%)

```
> plot(HUSrisk)
```



Résultats sur le risque moyen

```
> meanHUSrisk <-  mcapply(HUSrisk,"var",mean)
> summary(meanHUSrisk)

node :
       NoVar
median 0.492
mean   0.477
2.5%   0.120
97.5%  0.790
```

Risque attendu entre 12 et 79% : résultat intéressant directement le gestionnaire du risque, qui lui permet de calculer facilement un nombre de cas intégrant toutes les sources de variabilité au sein de la population, et prenant en compte les sources d'incertitude.

Utilisation des deux packages dans la communauté scientifique

- `fitdistrplus`
  Utilisé dans divers domaines, au vu d'une vingtaine de publications citant le package : risque alimentaire, écologie, entomologie, épidémiologie, biochimie, génomique, bioinformatique, neurobiologie, économie.
- `mc2d`
  Utilisé actuellement dans un cercle plus restreint, essentiellement dans le domaine du risque alimentaire.

Valorisation

- Article présentant les packages `mc2d` et `fitdistrplus` publié en 2010 dans *International Journal of Food Microbiology* primé par la "Society Of Toxicology" en 2011
- Rédaction en cours, avec Christophe Dutang, d'un article en vue d'une soumission à *Journal of Statistical Software*
- Développement d'un outil web basé sur le package `fitdistrplus` pour mettre à disposition des écotoxicologues un outil d'ajustement de SSD (Species Sensitivity Distributions) (Philippe Veber en collaboration avec l'équipe MEPS)

`MOSAIC_SSD`

Perspectives et difficultés rencontrées

De nombreuses idées de développement pour les deux packages mais le développement de packages est très chronophage, difficile à planifier pour un enseignant-chercheur, et souvent compliqué par des difficultés techniques inattendues :

- Evolution constante de **R** impliquant l'apparition de nouvelles erreurs à la compilation avec les versions stabilisées, en dehors de toute modification du package
- Difficulté d'analyse d'erreurs de compilation spécifiques à certaines plateformes alors que le package est entièrement écrit en **R** et fonctionne sur la machine du développeur

En conclusion, on a vraiment besoin d'un appui technique pour continuer !