
INTRODUCTION.....	- 3 -
1. INSTALLATION DU LOGICIEL EXTREMES	- 7 -
1.1. Environnement Windows	- 7 -
1.1.1. Le fichier setup.exe	- 7 -
1.1.2. Organisation des répertoires	- 11 -
1.2. Environnement Linux	- 11 -
1.2.1. Le fichier extr.tar.gz	- 11 -
1.2.2. Organisation des répertoires	- 11 -
2. DEMARRAGE D'EXTREMES AVEC L'INTERFACE GRAPHIQUE MATLAB..	- 12 -
2.1. Première exécution sous Windows	- 12 -
2.2. Exécutions ultérieures sous Windows.....	- 14 -
2.3. Première exécution sous Linux	- 14 -
2.4. Exécutions ultérieures sous Linux.....	- 15 -
3. INTERFACE GRAPHIQUE MATLAB	- 16 -
3.1. Organisation générale.....	- 16 -
3.1.1. Espace de travail.....	- 16 -
3.1.2. Historique des actions	- 17 -
3.1.3. Action en cours.....	- 18 -
3.2. Système de menus	- 19 -
3.2.1. Le menu Fichiers	- 19 -
3.2.2. Le menu Simulations.....	- 22 -
3.2.3. Le menu Graphiques	- 26 -
3.2.4. Le menu Estimations	- 30 -
3.2.5. Le menu Tests	- 57 -
3.2.6. Le menu Régularisation bayésienne.....	- 79 -
3.2.7. Le menu Aide	- 88 -
3.3. Boutons de contrôle du graphique.....	- 88 -
4. EXECUTION MANUELLE	- 95 -
4.1. Ligne de commandes.....	- 95 -
4.2. Fichier d'entrée du programme	- 97 -
4.2.1. Type de chargement et numéros de tâches.....	- 97 -
4.2.2. Modèles des fichiers d'entrée selon le type de chargement	- 97 -
4.3. Fichiers de sortie	- 100 -
ANNEXE A – Paramétrage des lois	- 101 -
ANNEXE B – Nom réel de la fonctionnalité et numéro de tâche.....	- 103 -
ANNEXE C – Exemples de fichiers d'entrée	- 105 -
ANNEXE D – Fichiers de sortie	- 148 -

INTRODUCTION

Le logiciel *EXTREMES* regroupe différents outils dédiés à l'étude des valeurs extrêmes : entre autres estimation des quantiles extrêmes et sélection de modèles pour les queues de distribution. Il est le fruit d'une collaboration entre l'équipe IS2 de l'INRIA Rhône-Alpes et la division Recherche et Développement d'EDF, et l'aboutissement des travaux de thèse de Myriam Garrido¹. Il ne s'adresse pas uniquement aux spécialistes des valeurs extrêmes, même s'il offre de nouveaux outils pour l'étude des queues de distribution.

Dans un 1^{er} point, nous décrivons le contexte mathématique permettant l'étude des événements rares et dans un 2^{ème} temps sont exposées les fonctionnalités du logiciel proprement dites.

- Fondements théoriques

La théorie des valeurs extrêmes (Embrechts² *et al*, 1997) a été développée pour l'estimation de probabilités d'occurrences d'évènements rares. Elle permet d'extrapoler le comportement de la queue de distribution des données à partir des plus grandes données observées. Le résultat suivant sur la loi des valeurs extrêmes est, pour le maximum de n observations, un analogue du théorème central limite pour la moyenne. Il décrit les limites possibles de la loi du maximum de n variables aléatoires indépendantes et identiquement distribuées correctement normalisées à l'aide de deux suites α_n et β_n .

Soit F la fonction de répartition de la loi d'intérêt. Sous certaines conditions de régularité sur F , il existe $\tau \in \mathbb{R}$ et deux suites normalisantes α_n et β_n tels que :

$$\forall x \in \mathbb{R}, \lim_{n \rightarrow \infty} F^n(\alpha_n + \beta_n x) = H_\tau(x),$$

où H_τ est la fonction de répartition de la loi des valeurs extrêmes :

$$\begin{aligned} \text{si } \tau \neq 0, \quad H_\tau(x) &= \exp\left(-\left(1 + \tau x\right)_+^{-1/\tau}\right) \\ \text{si } \tau = 0, \quad H_\tau(x) &= \exp\left(-\exp(-x)\right) \end{aligned}$$

et où la notation y_+ désigne $\max(y, 0)$.

On dit alors que la fonction de répartition F est dans le domaine d'attraction de Fréchet, de Gumbel ou de Weibull selon que $\tau > 0$, $\tau = 0$ ou $\tau < 0$.

Une 2^{ème} méthode d'estimation de queues de distribution est la méthode des excès ou POT (Peaks over threshold), introduite dans (Pickands³, 1975). Soit u un réel suffisamment grand appelé seuil. La méthode des excès s'appuie sur l'approximation de la loi des excès au-dessus

¹ Garrido M., Modélisation des événements rares et estimations des quantiles extrêmes, Méthode de sélection de modèles pour les queues de distribution, Thèse de doctorat, Université Grenoble 1, 2002.

² Embrechts P., Klüppelberg C., Mikosh T., *Modelling extremal events* – Springer-Verlag, Applications of mathematics, 1997.

³ Pickands J., « Statistical inference using extreme order statistics », The Annals of statistics, vol. 3, 1975, p. 119-131.

du seuil u de la variable aléatoire X , c'est-à-dire de la loi conditionnelle de la variable aléatoire $X - u$ sachant que $X > u$. La fonction de répartition des excès est définie par :

$$F_u(y) = P(X - u < y / X > u).$$

D'après le théorème de Pickands, si F appartient à l'un des 3 domaines d'attraction de la loi des valeurs extrêmes, la fonction de répartition F_u peut être approchée par une loi de Pareto généralisée (GPD) définie pour $\sigma > 0$ par :

$$\begin{aligned} \text{si } \gamma \neq 0, \quad G_{\gamma, \sigma}(x) &= 1 - \left(1 + \frac{x\gamma}{\sigma}\right)_+^{-1/\gamma} \\ \text{si } \gamma = 0, \quad G_{0, \sigma}(x) &= 1 - \exp(-x/\sigma) \end{aligned} \quad [1]$$

Sur la base de ces résultats, il est possible d'estimer des quantiles extrêmes. Un quantile extrême q_n d'ordre $(1 - p_n)$ est défini par l'équation $F(q_n) = 1 - p_n$ avec $p_n \leq 1/n$, n désignant la taille de l'échantillon. Un tel quantile étant généralement situé au-delà de l'observation maximale, des techniques spécifiques d'estimation sont nécessaires. La méthode POT s'appuie sur le théorème de Pickands pour estimer q_n par :

$$\hat{q}_n = u_n + \frac{\hat{\sigma}_n}{\hat{\gamma}_n} \left(\left(\frac{np_n}{k_n} \right)^{-\hat{\gamma}_n} - 1 \right) \quad [2]$$

où k_n désigne le nombre d'excès au-delà du seuil u_n et $\hat{\sigma}_n$ et $\hat{\gamma}_n$ sont des estimateurs des paramètres de la loi GPD. Pour ces derniers, de nombreuses propositions existent, voir par exemple (Embrechts¹).

- Fonctionnalités

Les sources du logiciel **EXTREMES** sont écrites en langage C++ et une interface graphique a été développée sous **Matlab** de façon à allier rapidité d'exécution et convivialité. Les fonctions disponibles sont regroupées en 3 catégories.

- Fonctions statistiques classiques

Les fonctions ci-dessous sont d'intérêt général au sens où elles ne sont pas dédiées à l'étude des valeurs extrêmes.

-- Simulations de variables aléatoires de lois Normale, Lognormale, Exponentielle, Gamma, Weibull, Chi2, Student, Pareto, Beta, Uniforme et Pareto généralisée

¹ Embrechts P., Klüppelberg C., Mikosch T., Modelling extremal events – Springer-Verlag, Applications of mathematics, 1997.

-
-
- Graphique des densités, fonctions de répartition, fonctions de survie, fonctions quantiles des lois précitées
 - Estimation des paramètres des lois précitées
 - Estimation non paramétrique de la densité (méthode de noyau, histogramme)
 - Estimation paramétrique des quantiles
 - Test d'Anderson-Darling et Cramer-Von Mises

- Fonctions extrêmes classiques

Nous regroupons ici les fonctions d'estimation et de test bien connues dans le domaine de la statistique des valeurs extrêmes.

- Vérification de l'exponentialité des excès : il s'agit de s'assurer que la fonction de répartition des données étudiées est dans le domaine d'attraction de Gumbel, et que le nombre d'excès k_n est convenablement choisi. L'ajustement de la loi Exponentielle aux excès est contrôlé graphiquement en traçant un QQ-plot. Un test d'exponentialité des excès est également proposé.

- Estimation des paramètres de la loi GPD. Sont regroupées ici plusieurs méthodes classiquement utilisées pour estimer le couple (γ, σ) , notamment les méthodes de Hill, Hill généralisé, Moments pondérés d'Hosking et Wallis, Maximum de vraisemblance et Zipf. Voir Embrechts¹ et les références indiquées.

- Estimation des quantiles extrêmes. Cette estimation s'appuie sur l'équation [2] et l'estimation des paramètres précédents.

- Procédures introduites dans la thèse de Myriam Garrido

Il s'agit de la partie la plus innovante du logiciel. Les fonctions rassemblées ici ont été intégralement développées dans le cadre d'une thèse co-financée par INRIA Rhône-Alpes et EDF.

- Test ET
- Test GPD
- Régularisation bayésienne

Le test ET et le test GPD sont 2 tests d'adéquation pour la queue de distribution. Ils sélectionnent par comparaison avec la méthode POT les modèles centraux produisant de bonnes estimations de la queue de distribution. Lorsqu'on souhaite reconstituer la loi des observations aussi bien dans la région centrale qu'extrême, on applique d'abord à un ensemble de modèles un test usuel (Anderson-Darling ou Cramer-Von Mises) puis un test d'adéquation de la queue de distribution (ET ou GPD). Si aucune loi n'est acceptée par les 2 types de tests, la procédure de régularisation bayésienne permet, à partir d'un modèle adapté aux valeurs les plus probables, d'améliorer l'adéquation extrême grâce à un avis d'expert sur la queue de distribution.

¹ Embrechts P., Klüppelberg C., Mikosch T., Modelling extremal events – Springer-Verlag, Applications of mathematics, 1997.

Dans un 1^{er} chapitre, nous expliquons la procédure d'installation du logiciel ***EXTREMES***. Le chapitre 2 est dédié au démarrage de l'application. L'interface graphique est décrite dans le chapitre 3 et enfin une dernière partie est consacrée aux exécutions manuelles (c'est-à-dire à la ligne de commande).

1. INSTALLATION DU LOGICIEL *EXTREMES*

1.1. Environnement Windows

1.1.1. Le fichier *setup.exe*

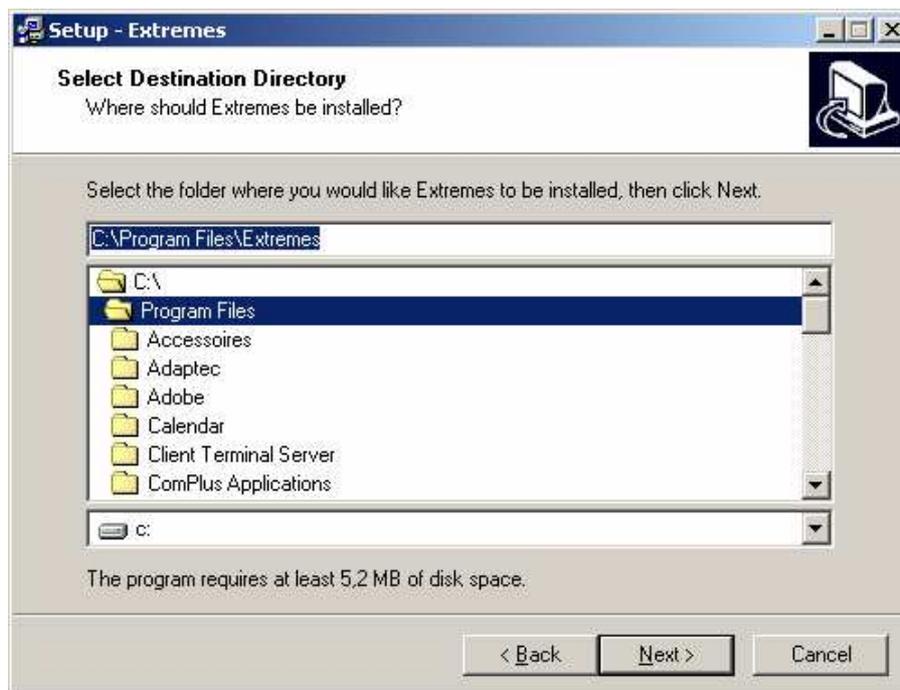
L'installation du logiciel *EXTREMES* est possible à l'aide du fichier *setup.exe*. Après l'avoir téléchargé sur l'ordinateur, il suffit de suivre les étapes suivantes :

- Exécuter *setup.exe*. Apparaît alors la **Fenêtre 1** :



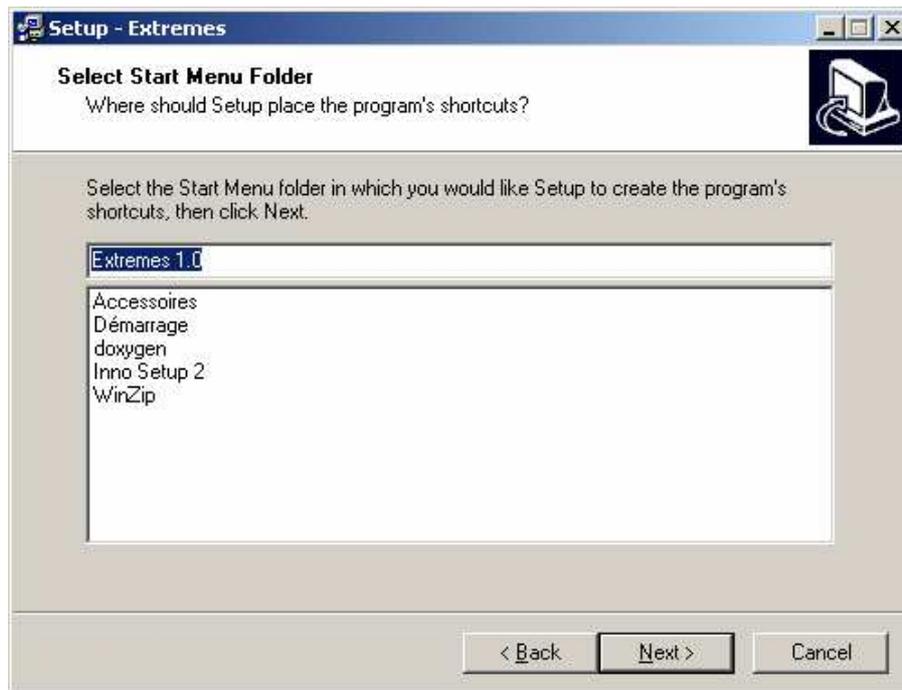
Fenêtre 1 – Ecran de début de l'installation

- Cliquer sur **Next** et choisir le répertoire d'installation (ex : c:\Program Files\Extremes)



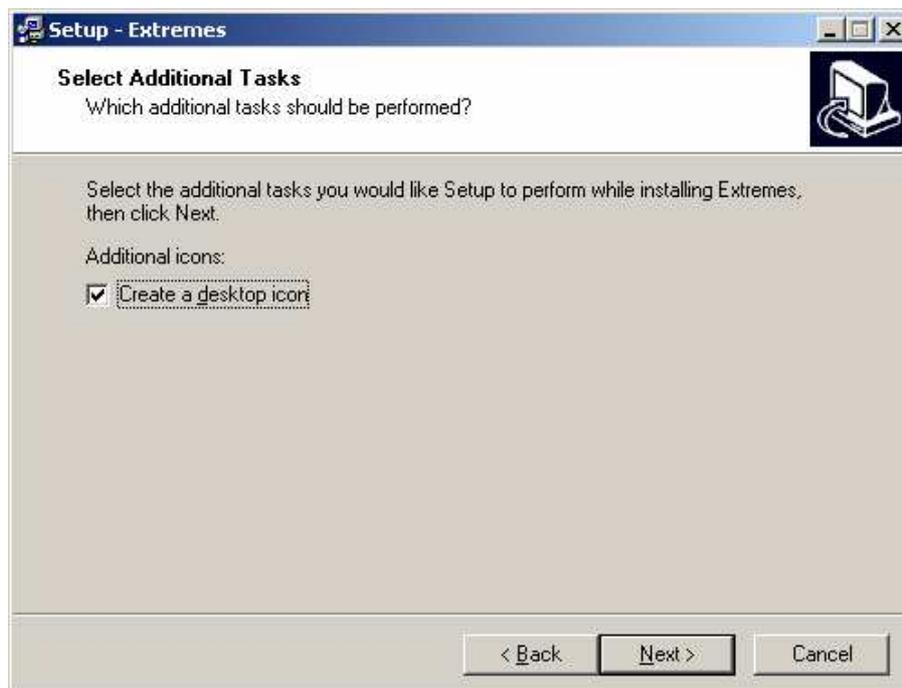
Fenêtre 2 – Sélection du répertoire d'installation

- Cliquer sur **Next** et déterminer le nom du groupe de programmes dans lequel seront installés 3 raccourcis (ex : Extremes 1.0)



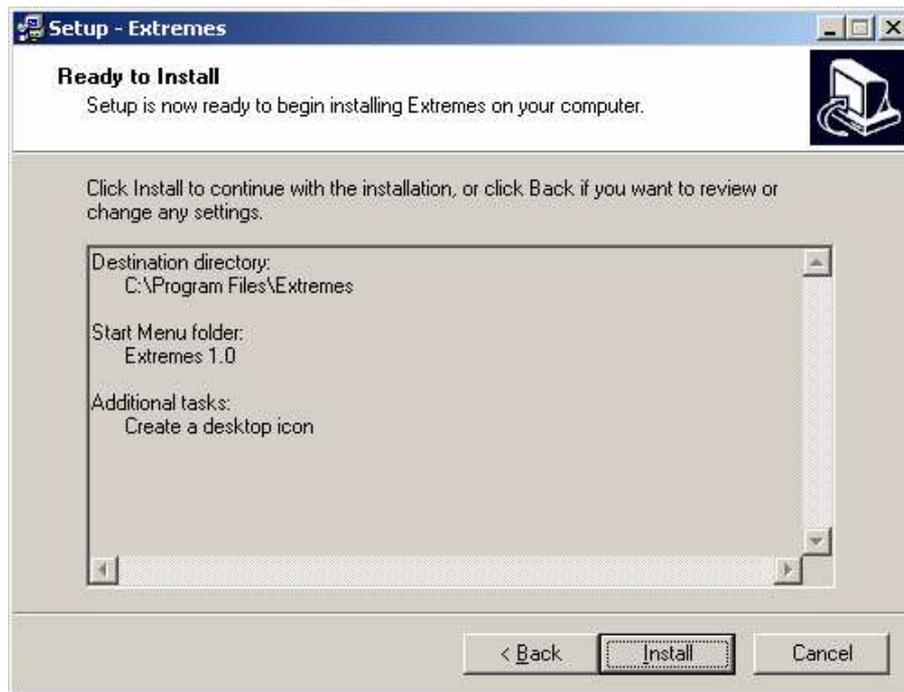
Fenêtre 3 – Sélection du groupe de programmes

- Cliquer sur **Next** et choisir la création ou non d'un icône sur le Bureau



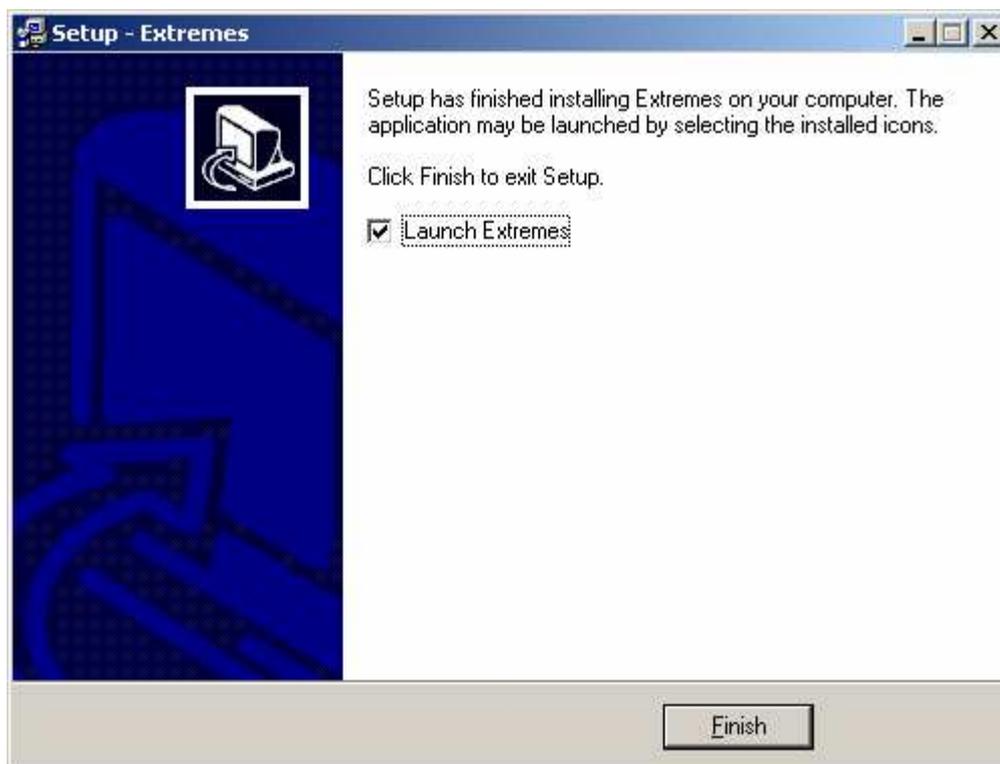
Fenêtre 4 – Création d'un icône

- Cliquer sur **Next**. Apparition de l'écran récapitulatif.



Fenêtre 5 - Récapitulatif des choix

- Cliquer sur **Install**, puis sur **Finish**



Fenêtre 6 - Ecran de fin d'installation

1.1.2. Organisation des répertoires

Si nous poursuivons notre exemple, *EXTREMES* est maintenant installé sous `c:\Program Files\Extremes`. Des sous répertoires ont été créés :

- Debug : contenant l'exécutable *liste.exe*
- doc : documentation technique en HTML (fichier principal : *index.html*) et manuel utilisateur *ManU.htm*.
- Donnees : contient des fichiers de données réelles et simulées
- Logo : image de présentation
- Matlab : renferme les fichiers d'interface Matlab
- Source : contient toutes les sources écrites en C++

1.2. Environnement Linux

1.2.1. Le fichier *extr.tar.gz*

Le fichier *extr.tar.gz* est compressé au format *zip*. Afin de le décompresser, il faut ouvrir un terminal, choisir un répertoire où installer le logiciel *EXTREMES*, puis taper les lignes de commandes suivantes :

```
gunzip extr.tar.gz  
tar -xvf extr.tar
```

1.2.2. Organisation des répertoires

Après avoir effectué la procédure précédente, un répertoire *Extremes* est maintenant créé dans lequel on trouve les sous répertoires suivants :

- Doc : documentation technique en HTML (fichier principal : *index.html*) et manuel utilisateur *ManU.htm*.
- Donnees : contient des fichiers de données réelles et simulées
- Logo : image de présentation
- Matlab : renferme les fichiers d'interface Matlab
- Source : contient toutes les sources écrites en C++
- Un fichier *runextr* : permettra l'exécution du logiciel

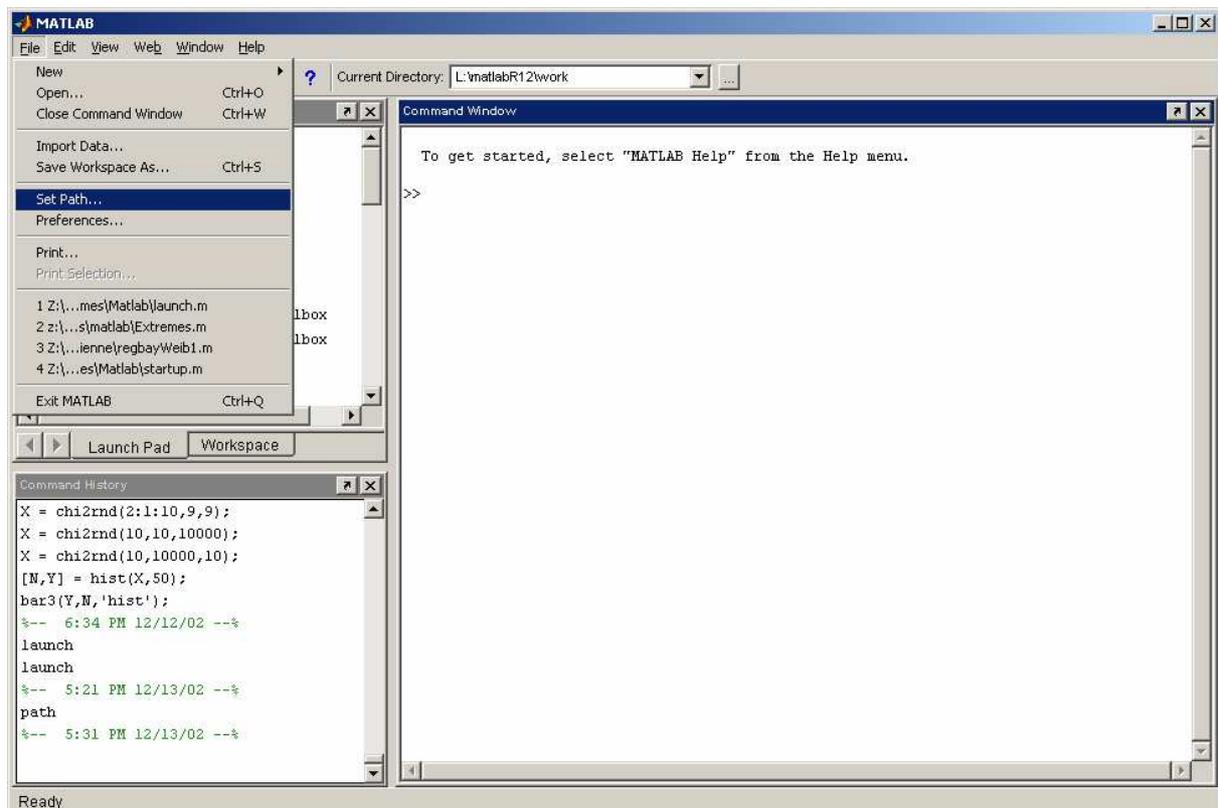
2. DEMARRAGE D'EXTREMES AVEC L'INTERFACE GRAPHIQUE MATLAB

Une fois le logiciel installé, il est possible de le lancer soit en double-cliquant sur l'icône présent sur le bureau, soit à partir du menu *Démarrer* de Windows.

2.1. Première exécution sous Windows

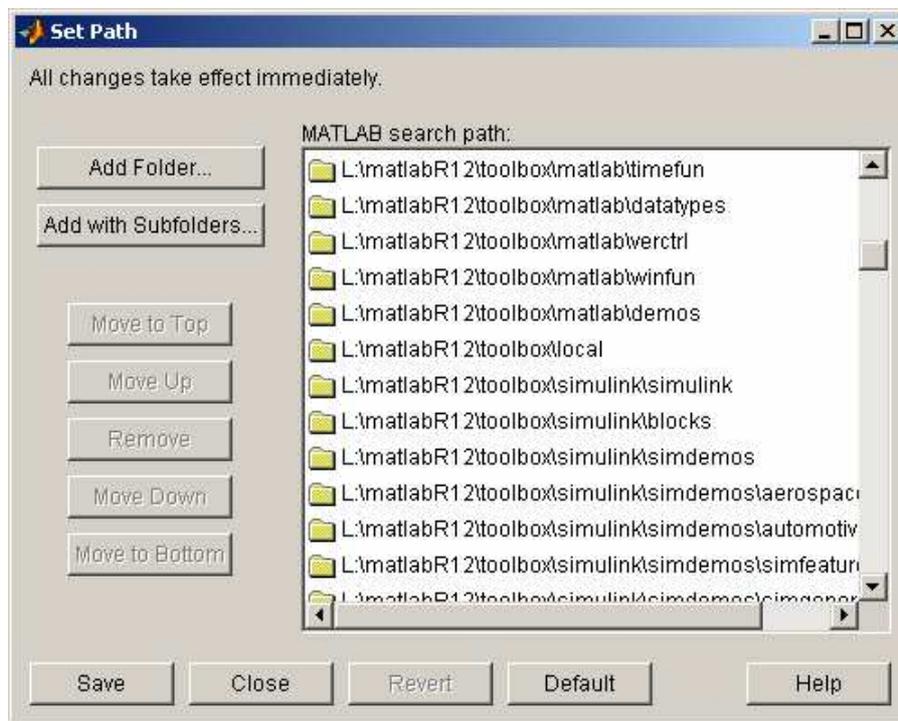
Lors de la première utilisation du logiciel, il se peut que seul *Matlab* soit lancé et qu'*EXTREMES* ne s'exécute pas. Il est alors nécessaire de configurer *Matlab*, notamment en ajoutant les chemins d'accès des fichiers d'*EXTREMES* dans la variable *path*. Pour cela, il faut procéder de la façon suivante :

- Dans Matlab, dérouler le menu *File* et cliquer sur *Set Path*



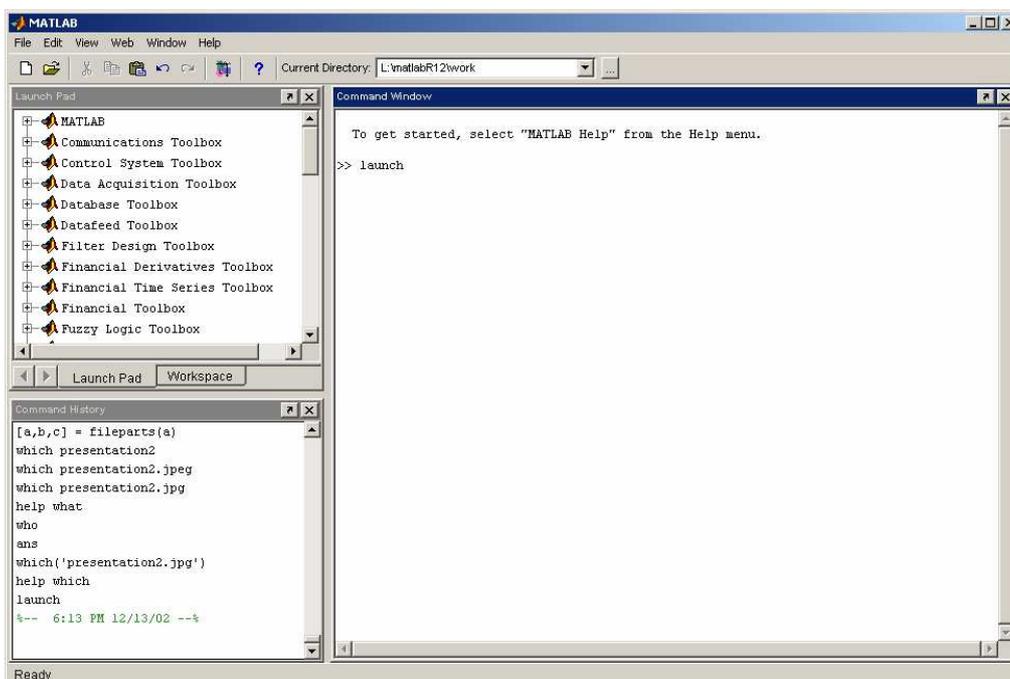
Fenêtre 7 - Ajout des chemins d'accès

- Cliquer sur *Add with Subfolders*, choisir ensuite le répertoire d'installation d'EXTREMES (ex : c:\Program Files\Extremes), puis *Save* et *Close*.



Fenêtre 8 - Ajout des chemins d'accès (2)

- Ensuite, taper *launch* à la ligne de commande. EXTREMES est alors lancé.



Fenêtre 9 - Lancement d'EXTREMES

2.2. Exécutions ultérieures sous Windows

Lancer **EXTREMES** est maintenant facile : il suffit de double-cliquer sur l'icône présent sur le bureau, ou s'il n'y est pas, de chercher **EXTREMES** dans le menu **Démarrer** de Windows. Il est à noter que **Matlab** s'exécutera toujours juste avant le logiciel **EXTREMES**.

Si des problèmes persistent, il est conseillé de placer le chemin courant (current directory) de **Matlab** sous le sous-répertoire d'**EXTREMES** nommé **Matlab** et de taper **launch** à la ligne de commande.

Exemple : si le répertoire d'installation d'**EXTREMES** est c:\Program Files\Extremes :

A la ligne de commande :

```
>> cd c:\Program Files\Extremes\Matlab  
puis  
>> launch
```

2.3. Première exécution sous Linux

ATTENTION : EXTREMES NE PEUT ETRE EXECUTE QU'AVEC MATLAB 6.5 (R13) OU MATLAB 6.

Dans le répertoire Extremes se trouve le fichier *runextr* (exécutable). Pour lancer le logiciel, il suffit de taper à la ligne de commande :

```
runextr chemin/où/se/trouve/le/logiciel/matlab
```

Exemple : runextr /softs/stow/matlab-13/bin/matlab

Le logiciel Matlab s'ouvre alors, tandis que **EXTREMES**, lui, ne s'exécute pas et provoque généralement une erreur tout à fait normale à ce stade de l'installation. Il est nécessaire d'ajouter les chemins d'accès des fichiers d'**EXTREMES** dans la variable **path**. Pour cela, il faut suivre la procédure décrite aux **Fenêtres 7,8 et 9** :

- Dans Matlab, dérouler le menu **File** et cliquer sur **Set Path**
- Cliquer sur **Add with Subfolders**, choisir ensuite le répertoire d'installation d'**EXTREMES** (ex : /home/mon_nom/Extremes)
- Puis Save. Il se peut qu'un message d'erreur apparaisse dû à une impossibilité d'écrire dans le fichier des **path** de matlab. **Pathdef.m** doit alors être sauvegardé sous le répertoire *Extremes* afin d'être utilisable pour les exécutions ultérieures.

Dès lors, sont présents sous le répertoire *Extremes* :

- les sous répertoires listés en **1.2.2**
- le fichier *runextr*
- le fichier *Pathdef.m*

- Ensuite, taper *launch* à la ligne de commande. *EXTREMES* est alors lancé.

2.4. Exécutions ultérieures sous Linux

Lancer *EXTREMES* est maintenant facile : il suffit d'ouvrir un terminal, de se placer sous le répertoire *Extremes*, puis de taper à la ligne de commande :
runextr chemin/où/se/trouve/le/logiciel/matlab

Exemple : *runextr /softs/stow/matlab-13/bin/matlab*

Il est à noter que *Matlab* s'exécutera toujours juste avant le logiciel *EXTREMES*.

Si des problèmes persistent, il est conseillé de placer le chemin courant (current directory) de *Matlab* sous le sous-répertoire d'*EXTREMES* nommé *Matlab* et de taper *launch* à la ligne de commande.

Exemple : si le répertoire d'installation d'*EXTREMES* est */home/mon_nom/Extremes* :

A la ligne de commande :

```
>> cd /home/mon_nom/Extremes/Matlab
```

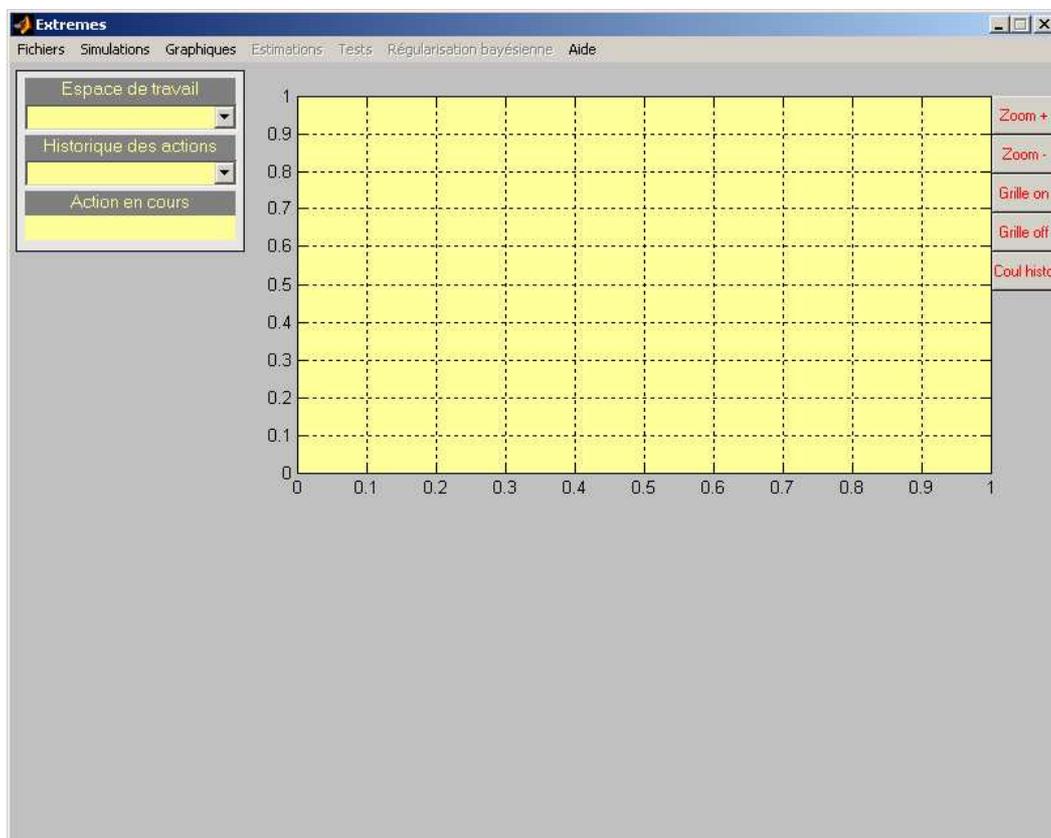
puis

```
>> launch
```

3. INTERFACE GRAPHIQUE MATLAB

Le logiciel *EXTREMES* possède 2 modes d'utilisation bien distincts : le mode jeu de données simples (lorsque l'utilisateur ne travaille que sur un seul jeu de données) et le mode jeu de données multiples ou multi données (action sur plusieurs échantillons de même taille et issus de même loi). Le logiciel détecte automatiquement si plusieurs jeux de données sont présents et demande à l'utilisateur quel mode il veut employer.

3.1. Organisation générale

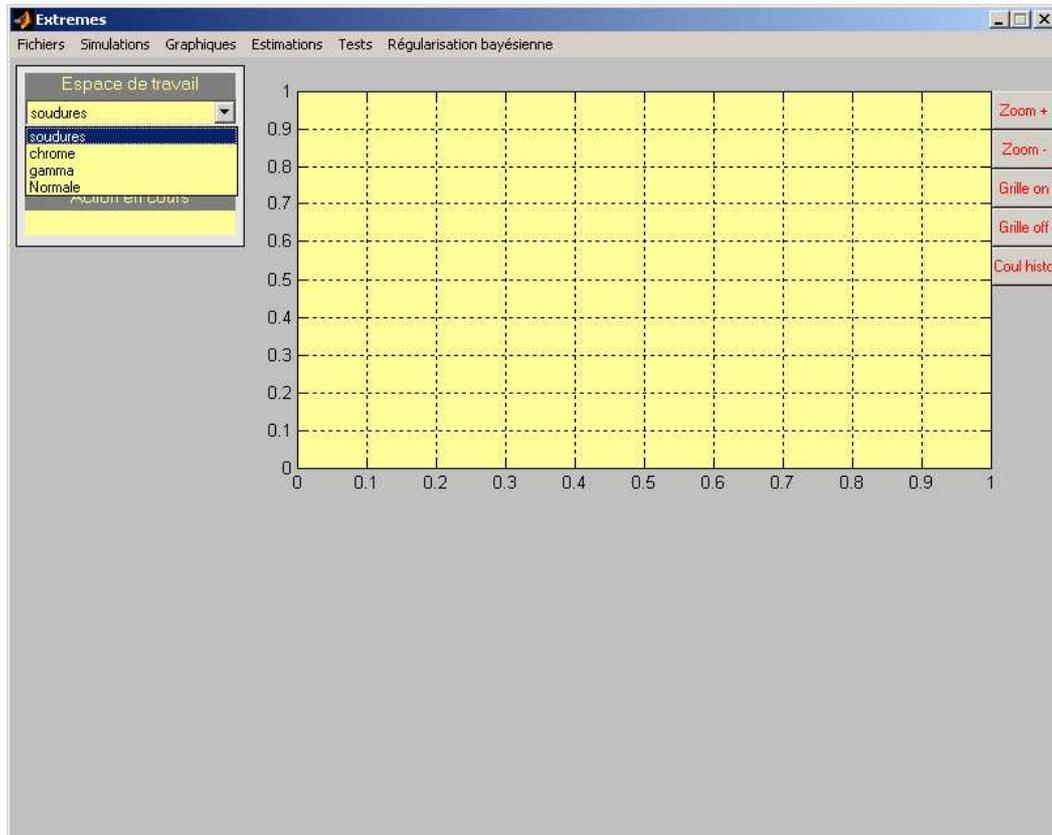


Fenêtre 10 - Ecran d'accueil d'EXTREMES

3.1.1. Espace de travail

Lorsque des données sont importées ou simulées, un *espace de travail* dont le nom est choisi par l'utilisateur est créé. Il est alors possible d'exploiter ces données à l'intérieur de celui-ci. Chaque *espace de travail* sera référencé dans la liste déroulante prévu à cet effet (Fenêtre 11). On peut passer d'un *espace de travail* à un autre simplement en cliquant sur un

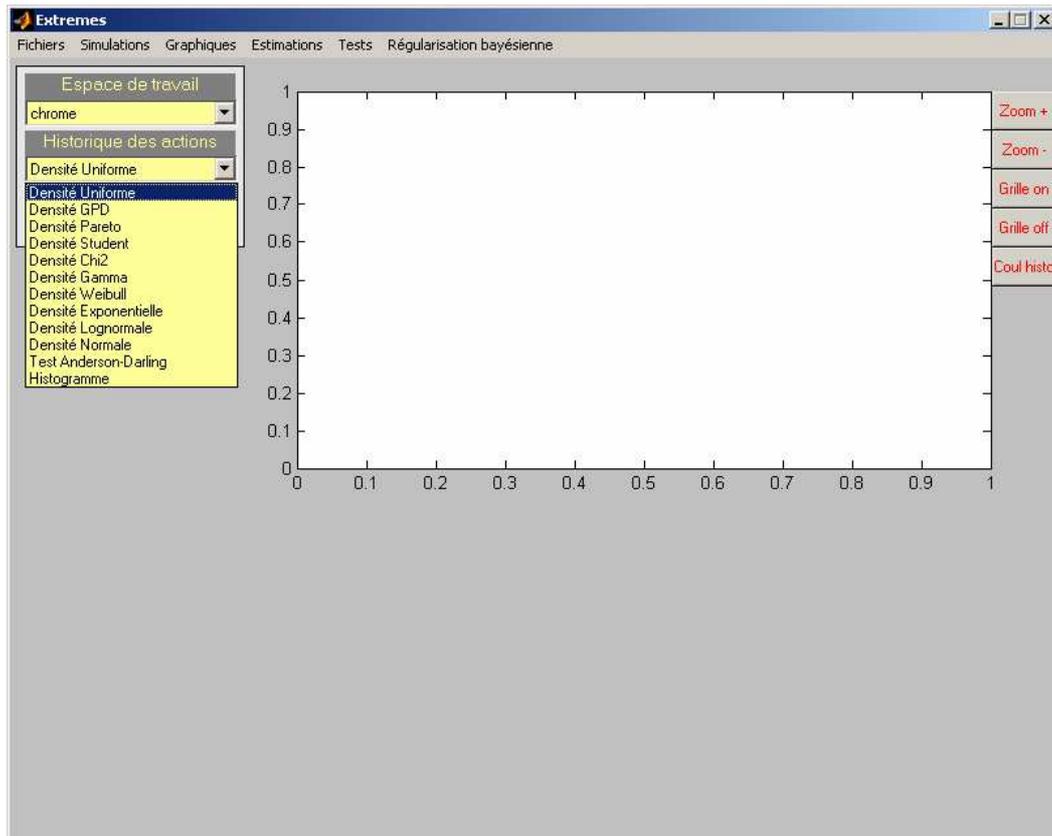
des noms de la liste déroulante. Il est donc recommandé de nommer judicieusement et de manière explicite chaque *espace de travail* afin d'éviter toute confusion préjudiciable. Un message d'erreur apparaît si l'utilisateur veut utiliser un nom déjà existant.



Fenêtre 11 - Liste des espaces de travail

3.1.2. Historique des actions

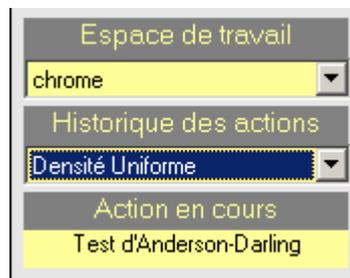
Il est possible d'effectuer diverses opérations sur chaque jeu de données (soit chaque espace de travail). Ces actions (ex : test d'Anderson-Darling, test ET, estimation des paramètres...) sont référencées dans une liste déroulante nommée *Historique des actions*. L'utilisateur peut donc à tout moment visualiser à nouveau le résultat de ses calculs précédents en cliquant sur l'action de son choix dans la liste (Fenêtre 12).



Fenêtre 12 - Liste des actions

3.1.3. Action en cours

Un champ indique l'opération dont le calcul est en cours. Il permet de se repérer et sera utile aux utilisateurs perdus dans leurs pensées et se demandant ce qu'ils avaient bien pu commencer. Le nom de l'action apparaît tant qu'elle est en cours de calcul (Fenêtre 13). Dès que notre utilisateur rêveur aura de nouveau la main, *Test d'Anderson-Darling* disparaîtra.



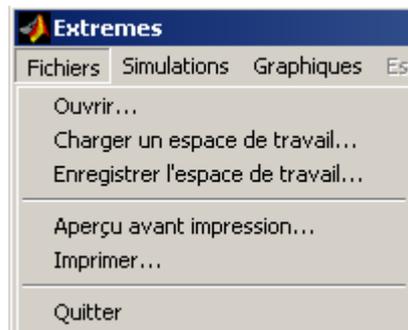
Fenêtre 13 - Action en cours

3.2. Système de menus

L'interface est organisée en menus afin d'en faciliter l'utilisation pour l'utilisateur. Ces menus sont : **Fichiers**, **Simulations**, **Graphiques**, **Estimations**, **Tests** et **Régularisation bayésienne**, les 3 derniers étant grisés (donc inaccessibles) au moment de l'écran de présentation. Ce n'est que lorsqu'un *espace de travail* est créé qu'ils peuvent être utilisés.

3.2.1. Le menu Fichiers

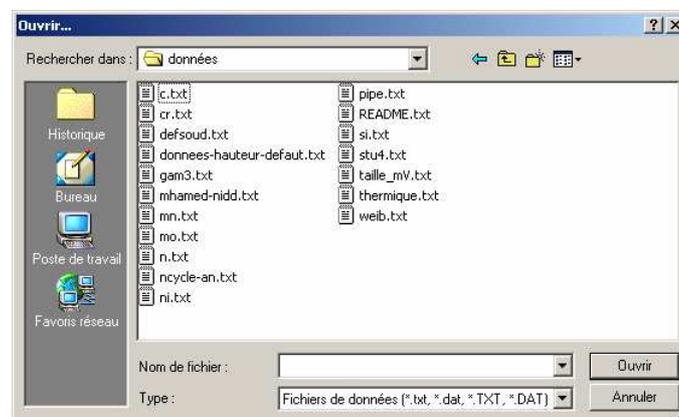
Le menu **Fichiers** contient 6 champs (Fenêtre 14) accessibles à tout moment.



Fenêtre 14 - Organisation du menu Fichiers

- Ouvrir...

Le menu **Ouvrir** s'utilise pour importer des données au format texte dans le logiciel **EXTREMES**. Ces données doivent impérativement être organisées en une seule colonne et le fichier qui les contient doit comporter l'extension *txt* ou *dat*.



Fenêtre 15 - Choix du fichier à importer

Exemple : ouverture du fichier *cr.txt* (fichier de données réelles de teneur en chrome se trouvant dans le sous-répertoire *Donnees* du répertoire d'installation)

Après avoir sélectionné le fichier voulu puis cliqué sur *Ouvrir*, l'utilisateur doit entrer un nom d'*espace de travail* (ici chrome) :



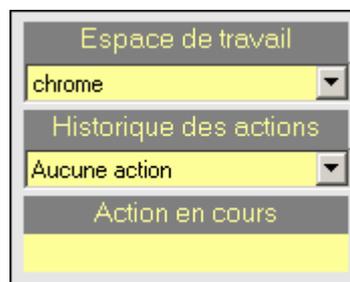
Fenêtre 16 - Saisie du nom de l'espace de travail

Le logiciel copie le fichier source dans un répertoire du même nom que l'espace de travail ainsi spécifié sous le nom de *Donnees.txt*, puis donne quelques informations à l'utilisateur (nombre de données, minimum, maximum) :



Fenêtre 17 - Informations sur le fichier de données

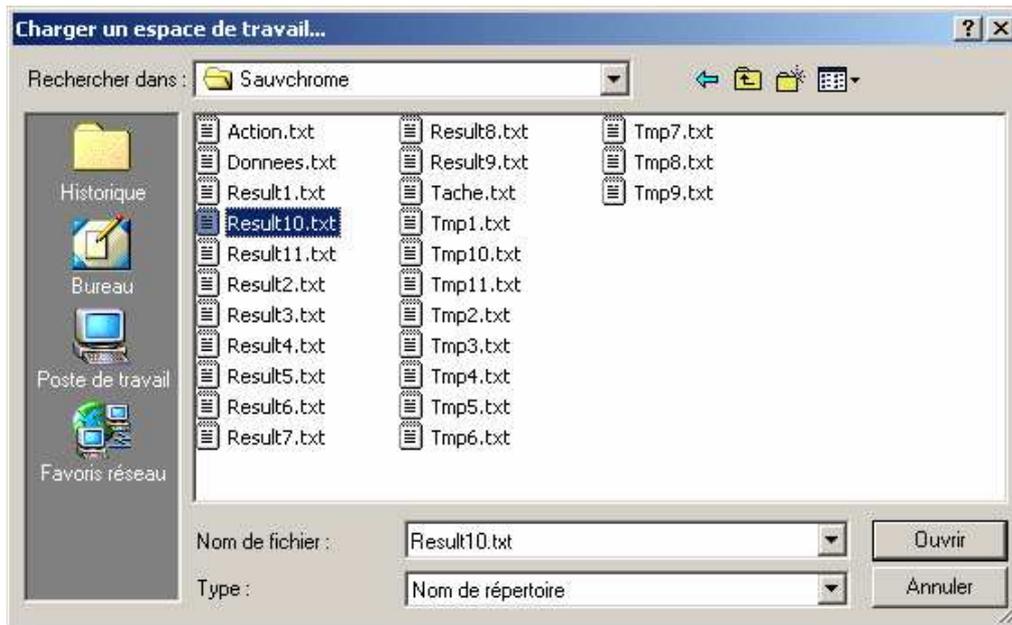
Tous les menus sont maintenant accessibles et l'interface indique maintenant qu'un *espace de travail* a été créé sous le nom de chrome et qu'aucune action n'a encore été effectuée :



Fenêtre 18 - Actualisation de la boîte de contrôle

- Charger un espace de travail...

Cette fonctionnalité peut être commandée à tout moment et charge un *espace de travail* au préalable sauvegardé dans *EXTREMES*. Il faut alors sélectionner impérativement un fichier du répertoire en question (n'importe lequel convient : ici Result10.txt du répertoire c:\TEMP\SauvChrome) :



Fenêtre 19 - Choix du répertoire à charger

Il est ensuite demandé un nom d'*espace de travail* à l'utilisateur (cf. Fenêtre 16) et la boîte de contrôle de l'historique des actions est alors actualisée. L'utilisateur peut reprendre le travail qu'il avait commencé.

- Enregistrer l'espace de travail...

L'intégralité des fichiers contenus dans l'espace de travail courant, c'est-à-dire celui qui est sélectionné à partir de la liste déroulante, peut être sauvegardée pour une utilisation ultérieure. Un emplacement et un nom de répertoire sont demandés à l'utilisateur (ici l'emplacement choisi est c:\TEMP et le nom du répertoire de sauvegarde SauvChrome) (cf. Fenêtre 20).



Fenêtre 20 - Sauvegarde d'un espace de travail

- Aperçu avant impression et Impression

Ces menus classiques permettent de configurer l'impression et d'imprimer la totalité de la fenêtre *EXTREMES*.

- Quitter

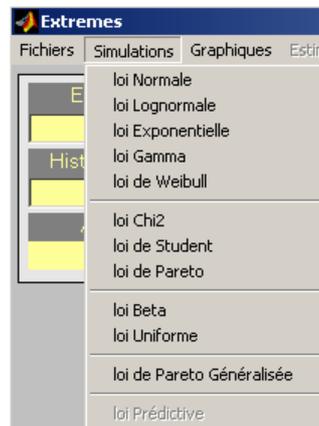
La fonction *Quitter* ferme l'application et efface le répertoire `\\.\Temp` créé par le logiciel, ainsi que tous ses sous-répertoires. Si l'utilisateur ne sauvegarde pas son travail en vue d'une prochaine manipulation, ses résultats seront alors totalement perdus.

3.2.2. Le menu Simulations

Comme son nom l'indique, cette fonctionnalité sert à simuler des variables aléatoires suivant les lois proposées. Ces lois (cf. annexe A pour la paramétrage utilisé) sont classées selon leur domaine d'attraction :

- Normale, Lognormale, Exponentielle, Gamma et Weibull appartiennent au domaine d'attraction de Gumbel (lois à queues de décroissance exponentielle)
- Chi2, Student, Pareto sont dans le domaine d'attraction de Fréchet (lois à queues lourdes, c'est-à-dire de décroissance de type puissance)
- Beta et Uniforme font partie du domaine d'attraction de Weibull (queues à point terminal fini)

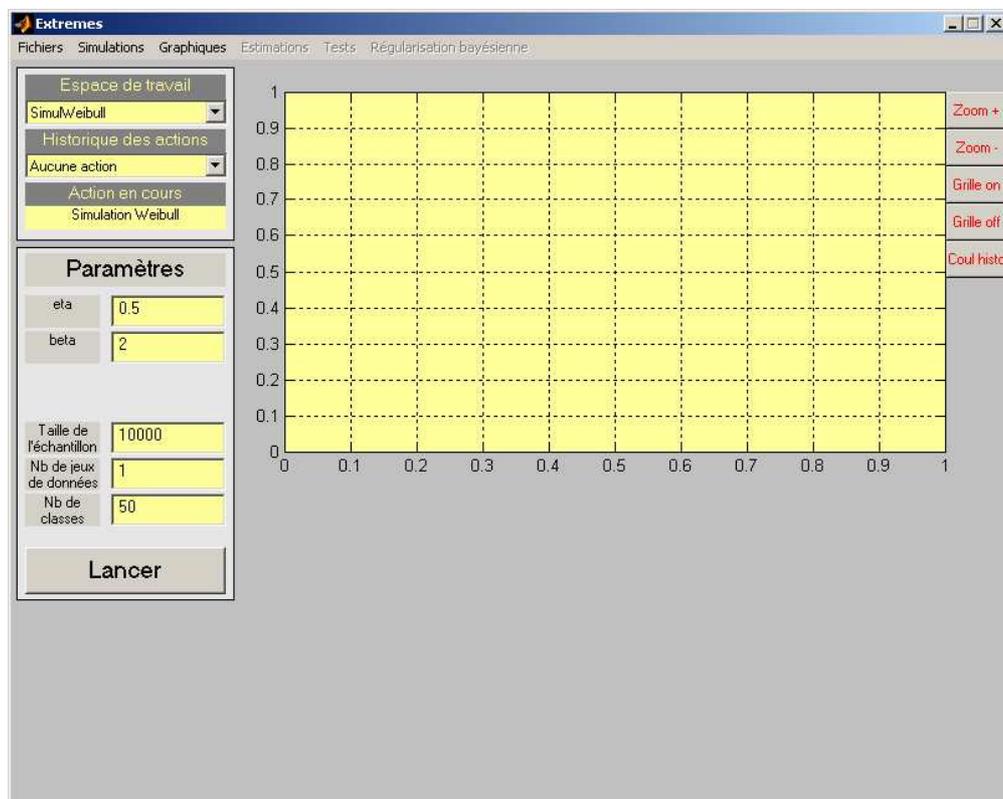
- La loi GPD (Generalized Pareto Distribution) peut être classée dans tous les domaines, suivant ses paramètres
- La mention **Loi prédictive** sera explicitée plus tard (menu régularisation bayésienne)



Fenêtre 21 - Organisation du menu Simulations

Exemple : Simulation d'une loi de Weibull

En cliquant sur **loi de Weibull**, le logiciel demande la création d'un nouvel **espace de travail** (cf. Fenêtre 16) que l'on nommera **SimulWeibull**. On obtient alors l'écran suivant :



Fenêtre 22 - Simulation d'une loi de Weibull

La boîte de dialogue *Paramètres* est apparue. Les valeurs dans les champs sont des valeurs par défaut. Il convient de les remplir correctement.

- Les paramètres de la loi choisie (cf. Annexe A – Paramétrage des lois utilisées)

Dans l'exemple proposé, il s'agit de *eta* et *beta*. Si les valeurs saisies par l'utilisateur sont invalides, le logiciel affichera un message d'erreur (Fenêtre 22).



Fenêtre 23 - Saisie de paramètres invalides pour la loi sélectionnée

- Taille de l'échantillon

La taille de l'échantillon doit être aussi renseignée. Elle est forcément supérieure ou égale à 1. Un message d'erreur est retourné si tel n'est pas le cas.

- Nb de jeux de données

Le nombre de jeux de données désiré est par défaut égal à 1. Cependant, il peut en être créé plusieurs qui sont tous issus de même loi, possédant la même taille, et qui sont tous regroupés dans le même espace de travail. Si par exemple on veut simuler 10 jeux de données, le premier jeu simulé s'appellera *Donnees.txt*, le 2^{ème} *Donnees2.txt*, le 3^{ème} *Donnees3.txt*, et ce jusqu'à 10 (ceci permet une exploration du type moyenne, intervalle de confiance, niveau et puissance). Le nombre de jeux de données doit bien sûr être supérieur strictement à 0, sinon un message d'erreur préviendra l'utilisateur.

- Nb de classes

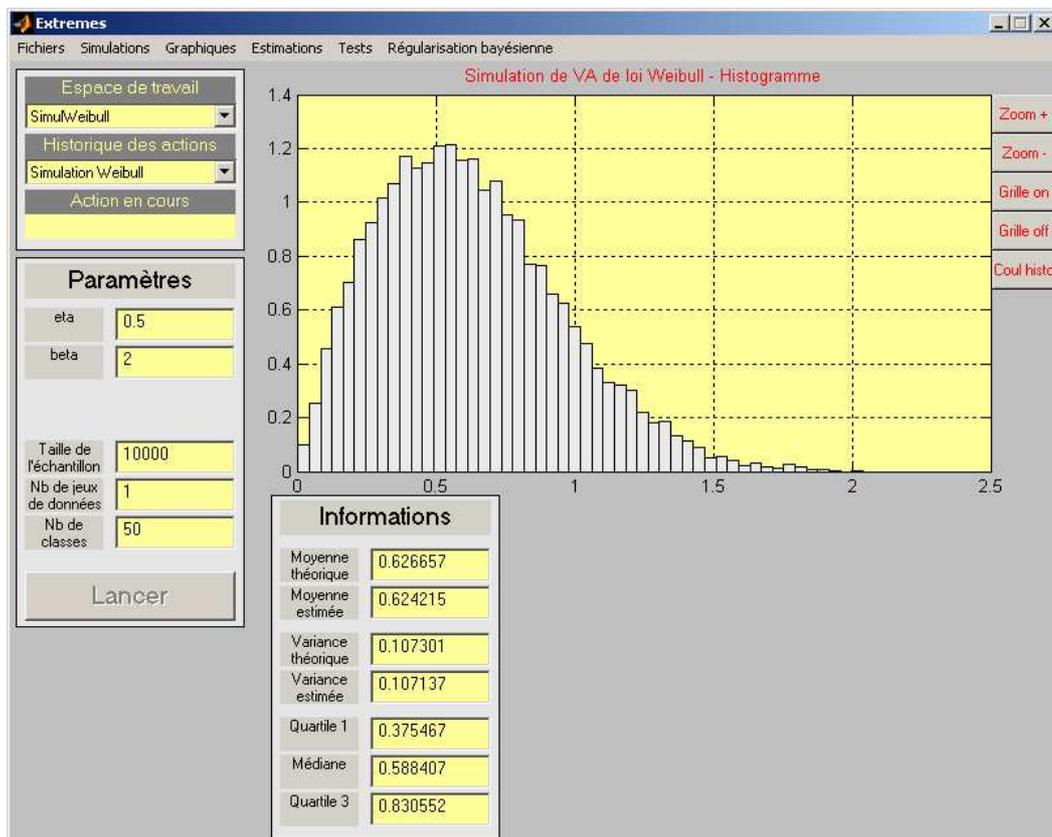
Ce champ est destiné à la représentation automatique d'un histogramme. La valeur inscrite correspond au nombre de classes de celui-ci, qui doit être plus grande que 1. Un contrôle est également prévu si cette condition n'est pas vérifiée.

Après un clic sur le bouton **Lancer**, le logiciel présente une petite boîte de dialogue demandant à l'utilisateur s'il veut superposer le graphique obtenu sur le précédent (**Non** par défaut) :



Fenêtre 24 - Superposition de graphiques

EXTREMES affiche alors les résultats suivants, en remarquant que le bouton **Lancer** devient inutilisable :



Fenêtre 25 - Résultat de la simulation

- Le graphique

L'histogramme est normalisé de manière à ce que son aire soit égale à 1, ceci afin de pouvoir lui superposer des courbes de densités.

- La boîte Informations

Cette boîte contient des informations relatives à l'échantillon simulé précédemment. Elle affiche la moyenne théorique, la moyenne estimée, la variance théorique, la variance estimée, le 1^{er} et 3^{ème} quartile ainsi que la médiane estimés. Ces éléments restent affichés constamment. Il est cependant à noter que si l'utilisateur change d'espace de travail, l'affichage de cette boîte n'est pas systématique.

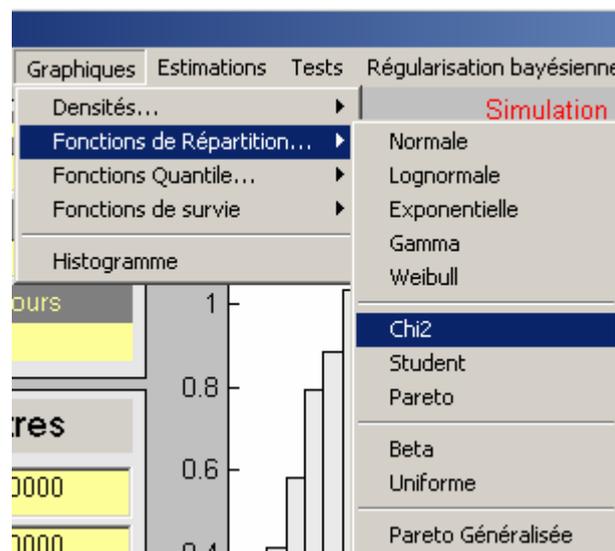
RM : Si plusieurs échantillons sont simulés, seules les informations concernant le premier jeu de données seront affichées.

- Le champ *Nb de classes*

L'utilisateur peut changer à volonté le nombre de classes de l'histogramme en remplaçant le nombre inscrit par un autre et en appuyant sur *Entrée* (ici on peut remplacer 50 par 30). L'histogramme s'actualise alors automatiquement.

3.2.3. Le menu Graphiques

A travers ce menu, il est possible de tracer les courbes de densités, de fonctions de répartition, de fonctions quantiles et de fonctions de survie du catalogue de lois (cf. annexe A pour le paramétrage utilisé), ainsi que de représenter un histogramme avec les données dont on dispose.



Fenêtre 26 – Organisation du menu Graphiques

- Densités, Fonctions de Répartition, Fonctions Quantiles, Fonctions de survie

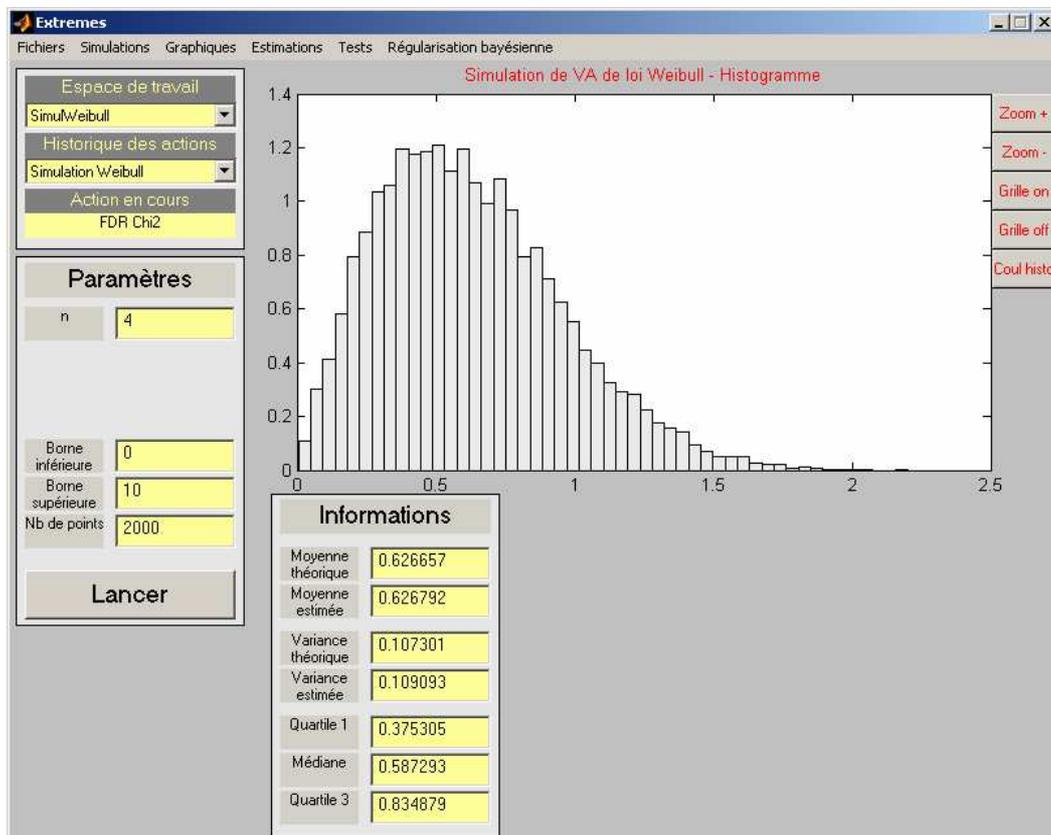
Exemple : Fonction de répartition d'une loi de Chi2

Lors de toutes ces opérations, une boîte de dialogue apparaît et demande à l'utilisateur les valeurs pour lesquelles le programme va évaluer par exemple la densité ou la fonction de répartition. Il peut choisir de rentrer lui-même ses points (*Entrées manuelles*), d'ouvrir un fichier contenant les abscisses voulues (*Fichier d'abscisses*) ou simplement d'annuler. A noter que la valeur par défaut est *Entrées manuelles*.



Fenêtre 27 - Choix des données pour la représentation graphique

Dans un premier temps, choisissons l'option *Entrées manuelles*. Apparaît cette fenêtre :

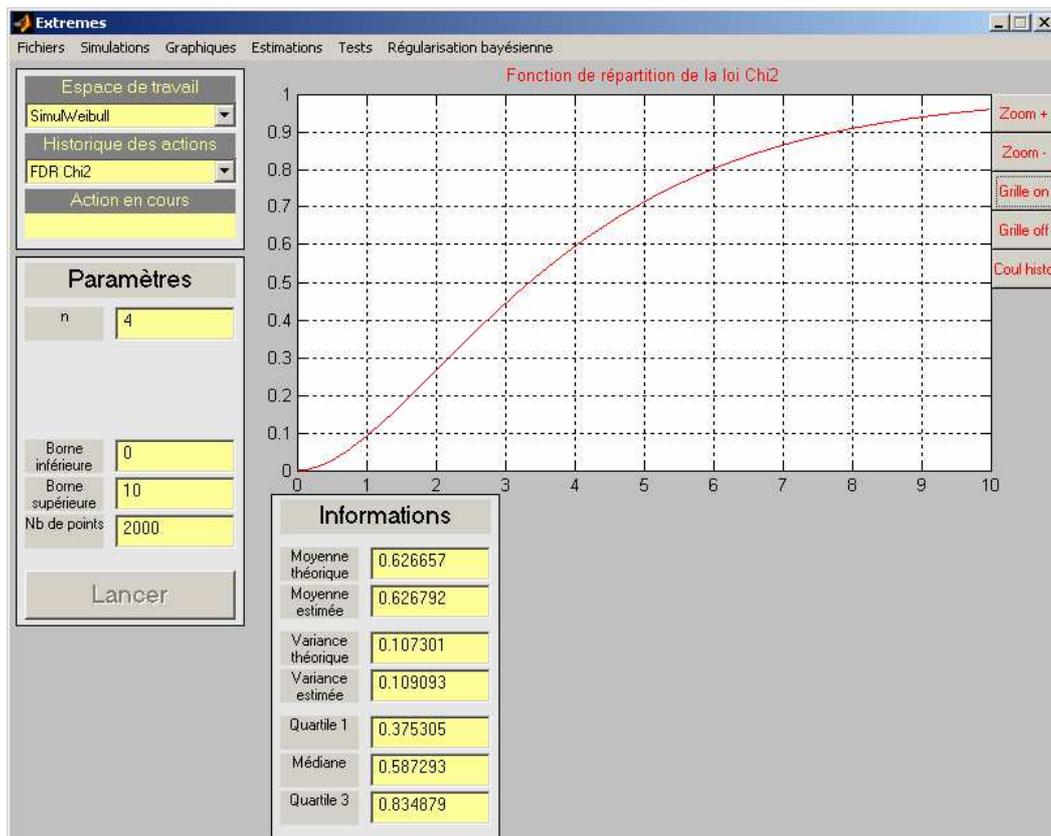


Fenêtre 28 - Ecran de saisie pour tracer une fonction de répartition d'une loi de Chi2

Des paramètres par défaut sont affichés directement. L'utilisateur peut bien sûr modifier chacun d'eux, à condition que les valeurs choisies ne soient pas invalides. Ici, le degré de liberté (4 dans l'exemple) de la loi de Chi2 ne doit pas être inférieur ou égal à 0, sinon un message d'erreur apparaît à l'écran.

Les champs *Borne inférieure* et *Borne supérieure* représente l'intervalle souhaité sur lequel on va évaluer notre fonction. Sur cet intervalle de longueur 10, on choisit de prendre 2000 points (champ *Nb de points*) répartis régulièrement. La borne supérieure de l'intervalle doit être plus grande que la borne inférieure, et le nombre de points demandés est strictement plus grand que 0. Des messages d'erreur sont prévus en cas de violation de ces critères.

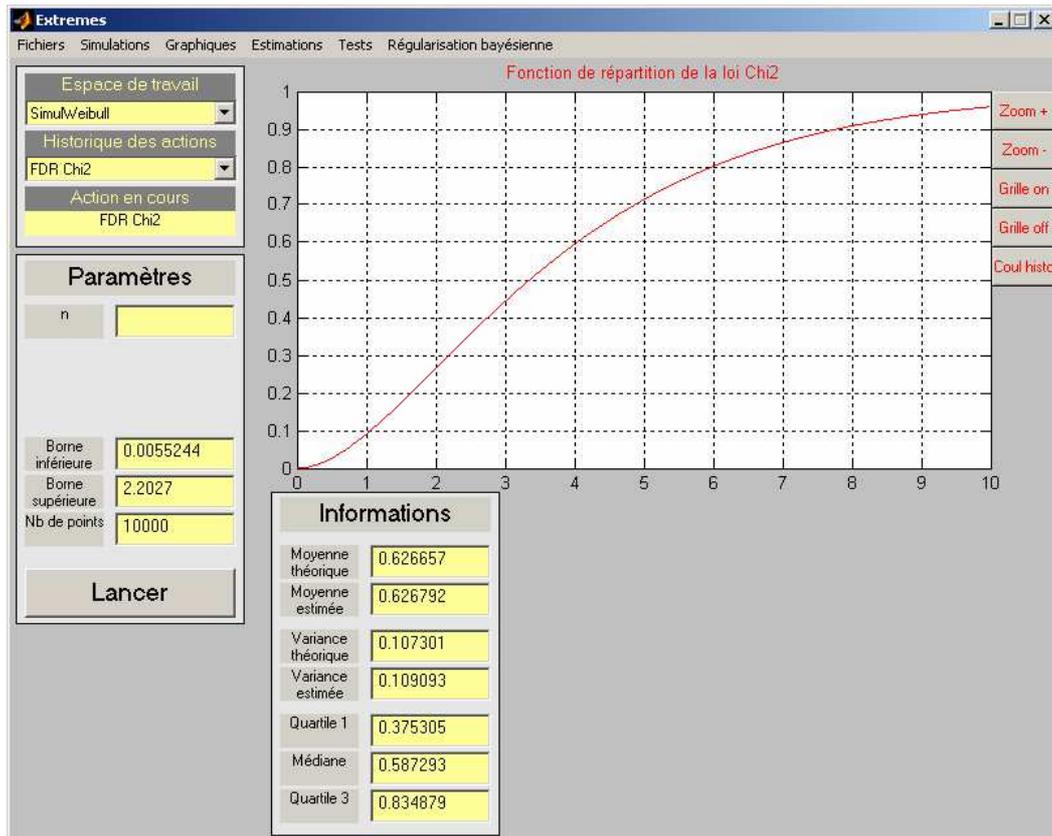
Après avoir cliqué sur *Lancer*, on obtient le dessin de la courbe en ayant au préalable répondu à la question de la superposition de l'affichage (cf. Fenêtre 24) :



Fenêtre 29 - Courbe de la fonction de répartition de la loi Chi2

Rééditons cet exemple en choisissant maintenant *Fichier d'abscisse* à la place d'*Entrées manuelles* (cf. Fenêtre 27) :

La boîte de dialogue *Ouvrir* s'affiche sur l'écran (cf. Fenêtre 15). Les abscisses des points pour lesquels l'utilisateur veut calculer la fonction doivent impérativement être organisées en colonne dans le fichier. Seuls les fichiers avec les extensions *txt* ou *dat* apparaissent. Nous choisissons par exemple le fichier *Donnees.txt* dans le répertoire *Extremes\Temp\SimulWeibull*. L'écran de saisie s'organise dorénavant comme suit :



Fenêtre 30 - Fonction de répartition d'une loi de Chi2 avec fichier d'abscisses

Le contenu du fichier est trié par le programme et sont affichés la valeur minimale (ici 0.0055244) et maximale (2.2027) ainsi que le nombre d'entrées (10000). La valeur du degré de liberté (n) de la loi de Chi2 doit être fournie par l'utilisateur. En cliquant sur **Lancer**, le programme va calculer la valeur de la fonction de répartition de la loi de Chi2 avec le paramètre choisi pour chacune des 10000 abscisses présentes dans le fichier. Une courbe du type de celle de la Fenêtre 30 sera alors dessinée.

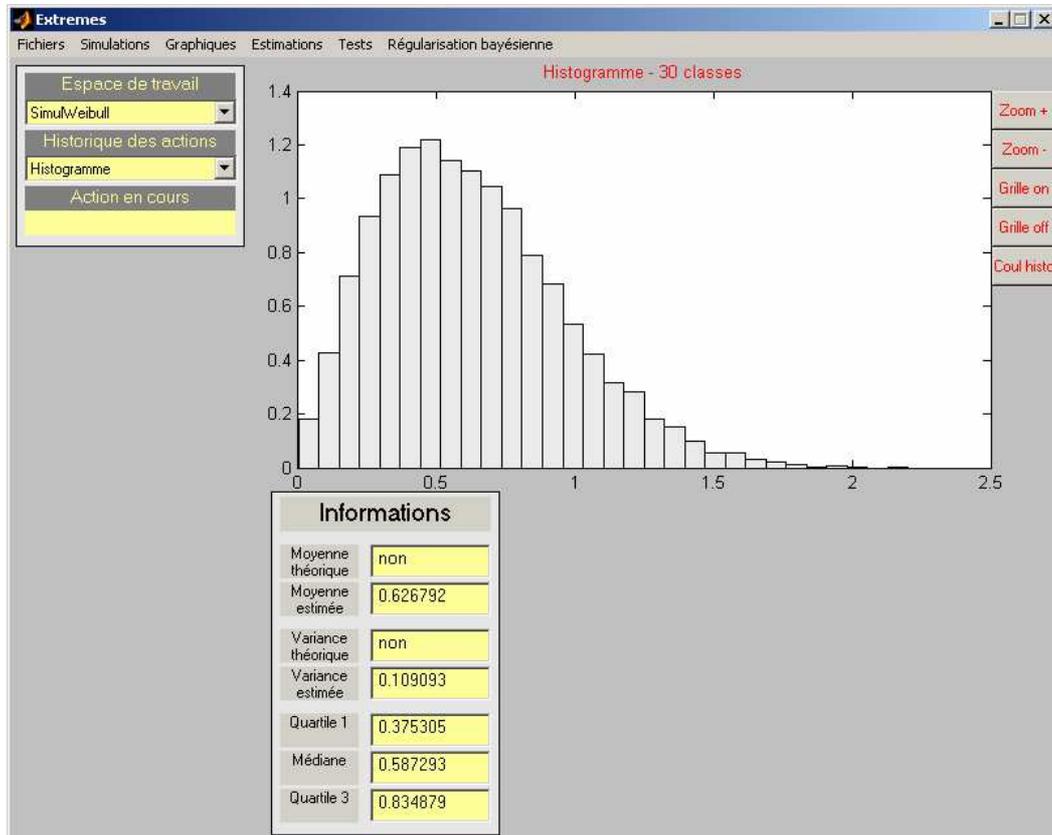
- Histogramme

Cette fonctionnalité peut-être à tout moment commandée par l'utilisateur. Elle dessine un histogramme des données en cours selon un nombre de classes à définir et affiche ou réaffiche les informations relatives à celles-ci dans la boîte **Informations**. En cliquant sur **Histogramme**, on voit apparaître cette fenêtre :



Fenêtre 31 - Nombre de classes pour un histogramme

Dans cet exemple, les données en cours sont celles simulées selon la loi de Weibull $W(0.5,2)$ et sont au nombre de 10000. On peut donc choisir un nombre de classes entre 30 et 50. Prenons par exemple 30 classes. En cliquant sur OK, on obtient :

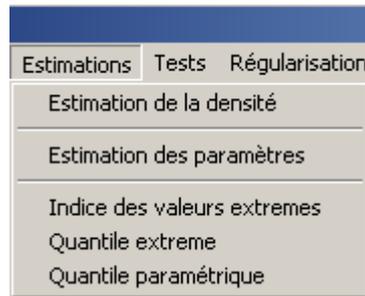


Fenêtre 32 - Représentation de l'histogramme à 30 classes

On peut remarquer que seule la boîte *Informations* s'affiche et ne présente pas de valeur pour les champs *Moyenne théorique* et *Moyenne estimée*. En effet, le programme travaille sur les données brutes et ignore leur origine.

3.2.4. Le menu Estimations

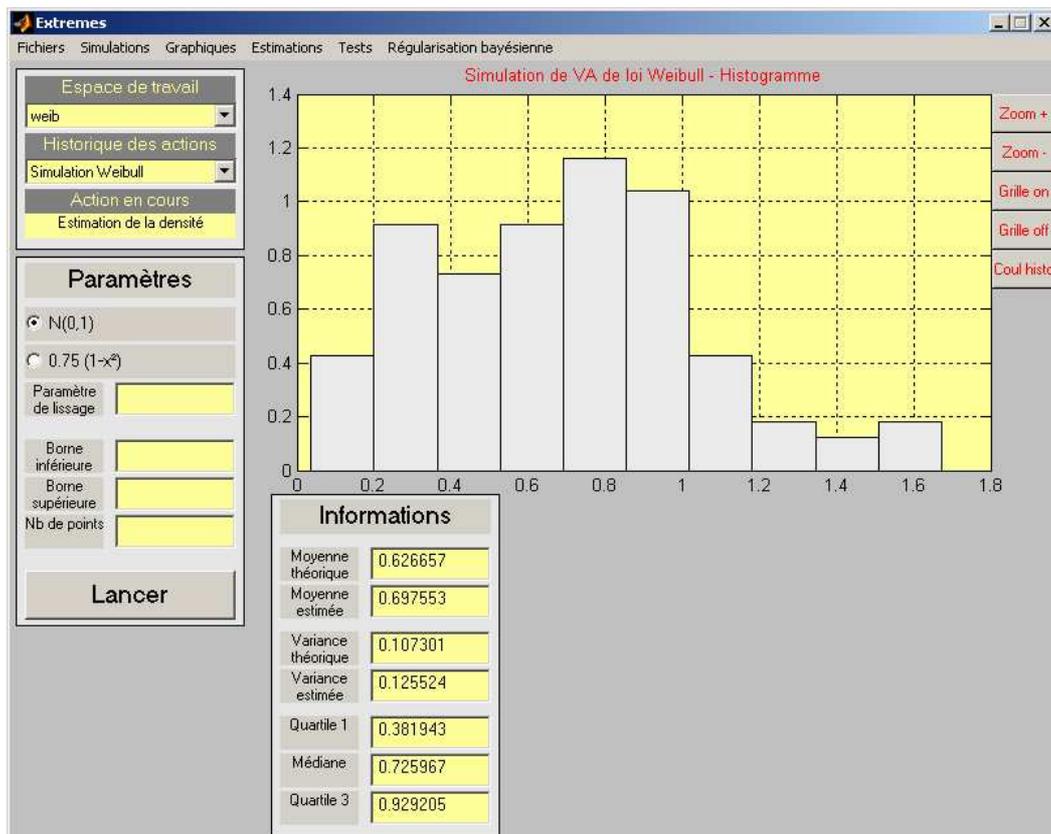
Il devient accessible dès qu'au moins un jeu de données est simulé ou importé. Les fonctionnalités de ce menu réagissent différemment en mode données multiples. *Estimation de la densité*, *Estimation des paramètres*, *Indice des valeurs extrêmes*, *Quantile extrême* et *Quantile paramétrique* sont les composantes de ce menu (cf. Fenêtre 33).



Fenêtre 33 - Organisation du menu Estimations

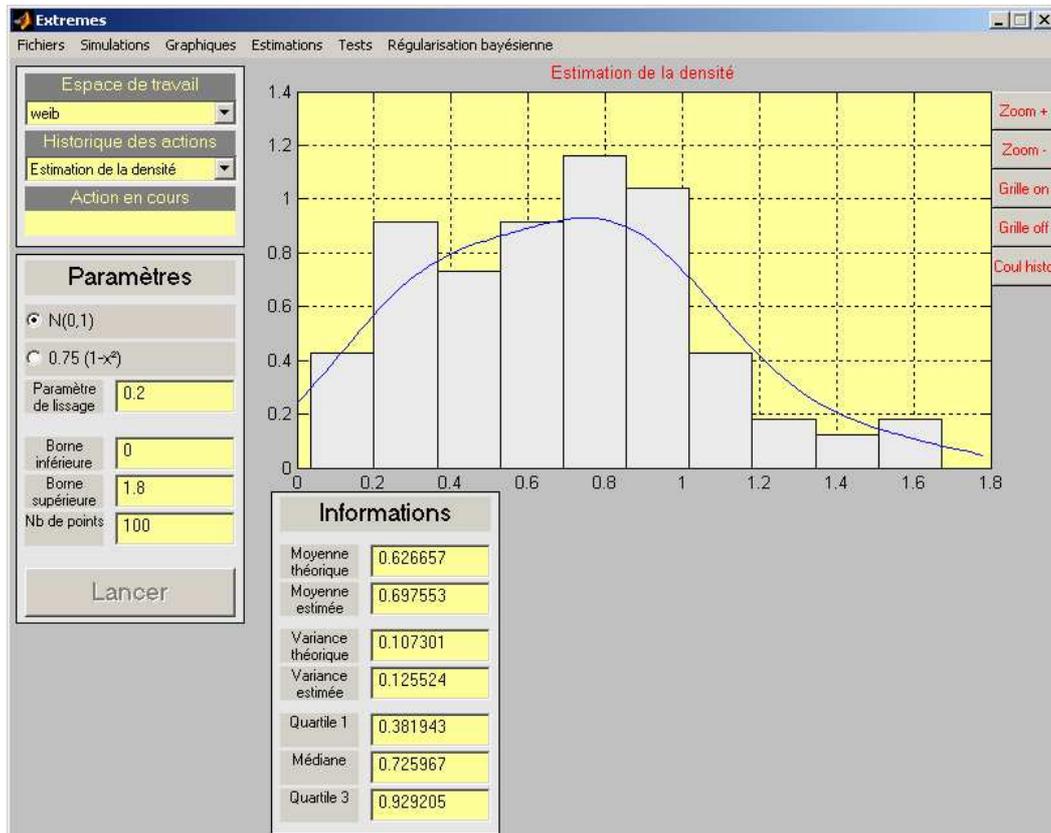
- Estimation de la densité

Cette procédure permet d'estimer non paramétriquement la densité de probabilités relative au jeu de données en cours en utilisant une méthode de noyau. L'utilisateur doit choisir les abscisses des points de calcul et comme pour le menu **Graphiques**, le choix est proposé entre des entrées manuelles et un fichier d'abscisses (cf. Fenêtre 27). Après avoir effectué ce choix, la fenêtre suivante s'affiche (exemple avec un échantillon de taille 100 issu de la loi de Weibull) :



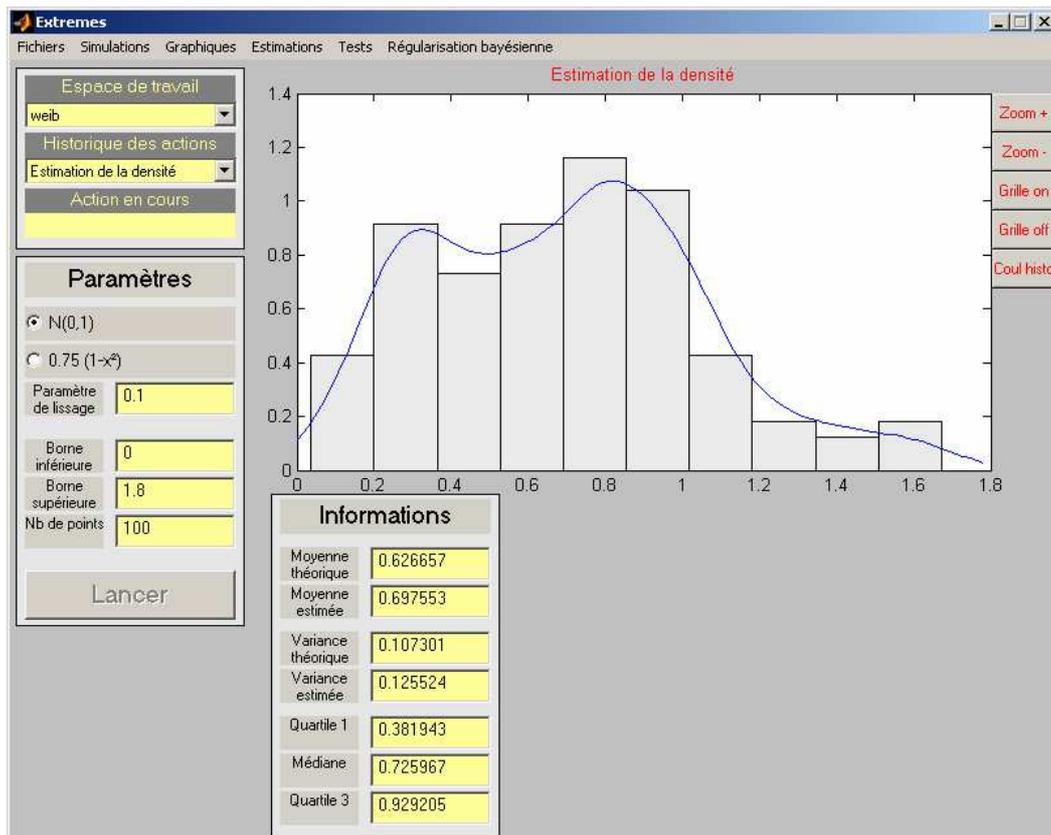
Fenêtre 34 - Ecran de saisie pour l'estimation d'une densité

L'utilisateur a le choix entre 2 noyaux : un noyau gaussien $N(0,1)$ (par défaut), et la densité $0.75(1-x^2)$. Le **paramètre de lissage** est l'élément essentiel et le plus difficile à déterminer : de grandes différences peuvent être observées pour de faibles variations de ce paramètre. Le choix du noyau influe bien évidemment aussi sur le résultat. Il n'est pas nécessaire de demander un nombre trop important de points de calcul. Cette valeur dépend de la taille de l'échantillon mais il est conseillé d'utiliser une grandeur de l'ordre de 100 ou 200.

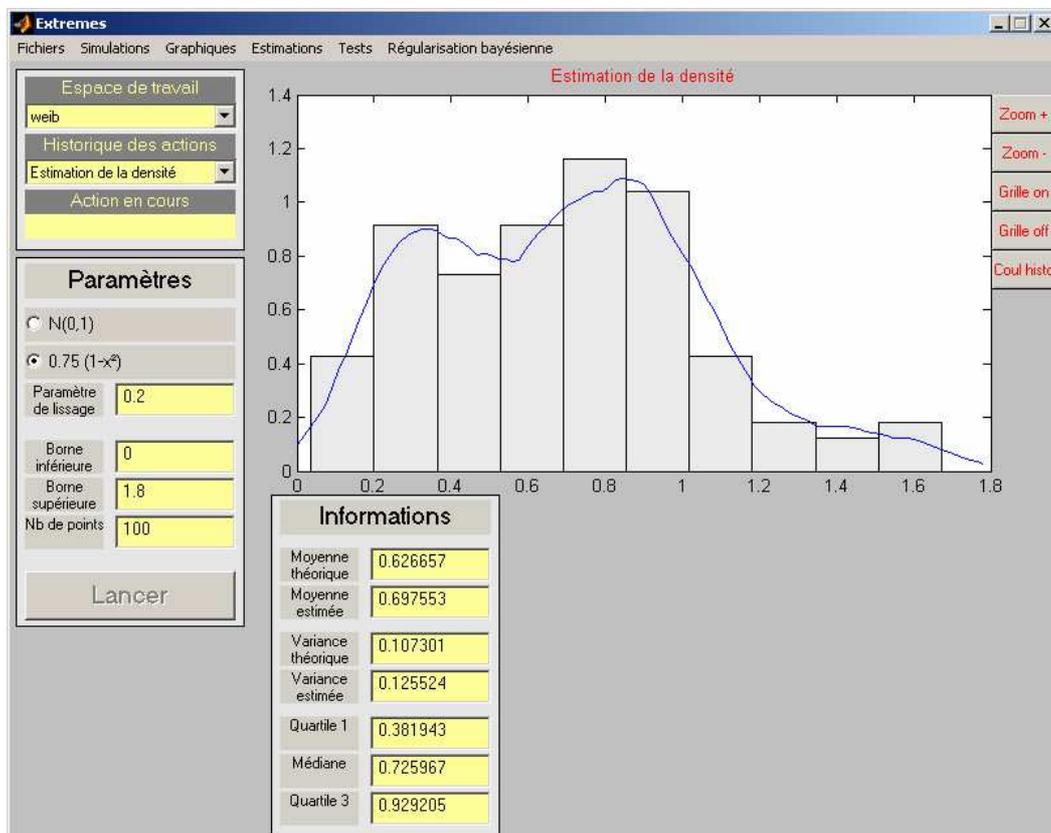


Fenêtre 35 - Estimation de la densité - Paramètre de lissage = 0.2 et Nb de points = 100

On s'aperçoit sur cet exemple que la densité estimée s'ajuste parfaitement sur l'histogramme des données. Dans les exemples suivants, on peut observer des résultats différents. Avec un **paramètre de lissage** plus faible, la courbe obtenue tient plus compte des données et elle s'en trouve moins « lissée » (cf. Fenêtre 36). Si on prend un **paramètre de lissage** plus important (de l'ordre de 1), la courbe sera très aplatie et ne s'ajustera pas correctement car le calcul accordera moins d'importance aux données. Un exemple d'utilisation du noyau $0.75(1-x^2)$ est présenté ci-dessous (cf. Fenêtre 37) qui permet de visualiser des différences entre les deux densités proposées.



Fenêtre 36 - Estimation de la densité - Paramètre de lissage = 0.1 et Nb de points = 100

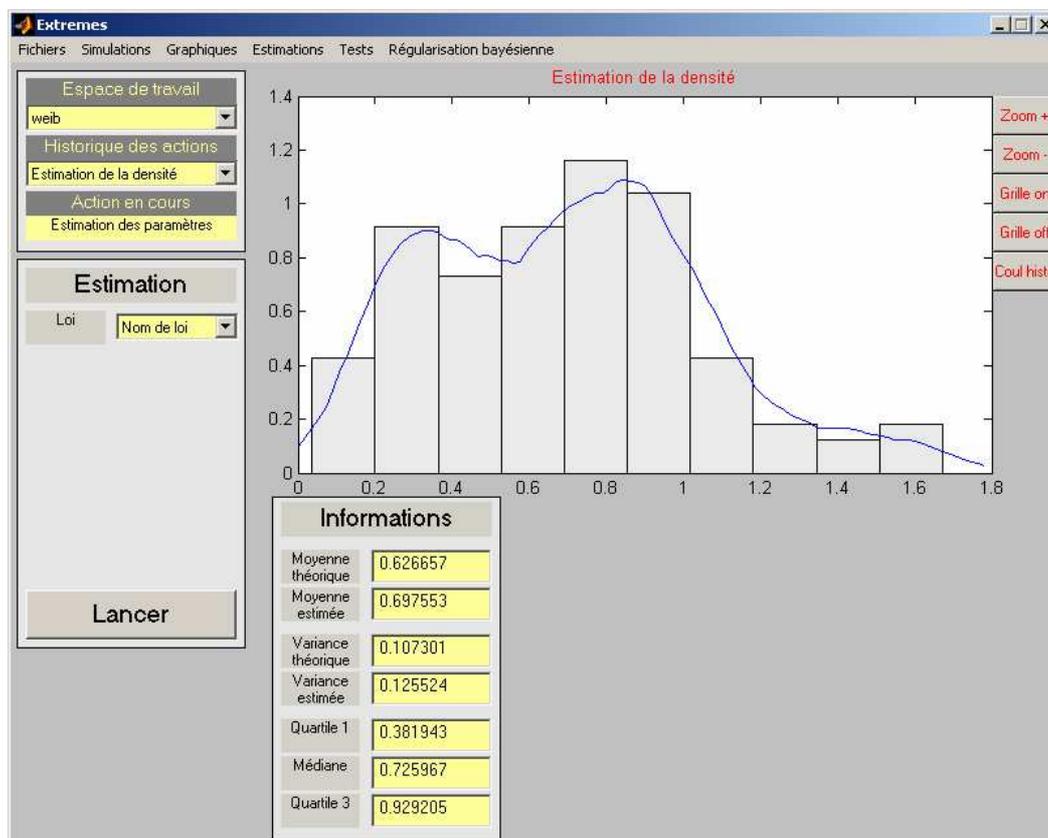


Fenêtre 37 - Estimation de la densité en changeant le noyau

Remarque 1 : Il est à noter que cette fonctionnalité ne peut s'utiliser que sur un seul échantillon. Si l'utilisateur a simulé plusieurs jeux de données, ce calcul ne sera effectué que sur le premier jeu appelé *Donnees.txt*.

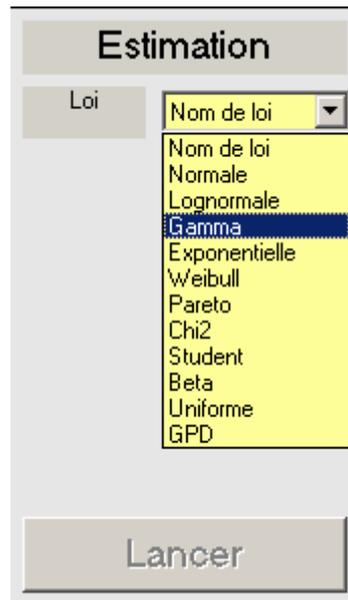
- Estimation des paramètres

Il est possible d'estimer les paramètres de toutes les lois du catalogue. En revanche, contrairement à l'estimation de la densité, cette fonction est différente selon le mode (multi données ou non) défini par l'utilisateur. On obtient l'écran suivant :



Fenêtre 38 - Ecran de saisie pour l'estimation des paramètres

On voit apparaître une liste déroulante contenant toutes les lois. Il suffit ensuite d'en choisir une (par exemple Gamma) (cf. Fenêtre 39). Les noms donnés aux paramètres apparaissent (cf. Fenêtre 40) et leur signification est présentée en annexe A.



Fenêtre 39 - Liste déroulante contenant le nom des lois



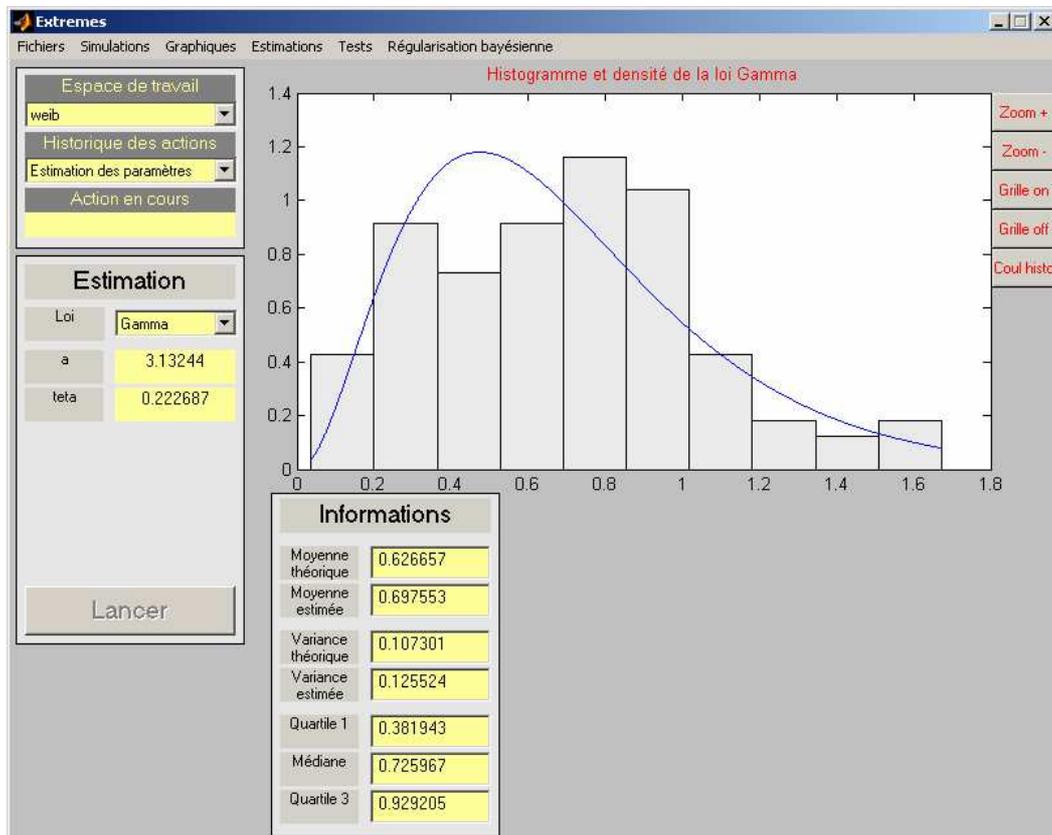
Fenêtre 40 - Apparition du nom des paramètres

Le bouton **Lancer** est alors disponible. Le résultat obtenu peut être de 2 types différents après avoir cliqué sur **Lancer** :

En mode jeu de données simple :

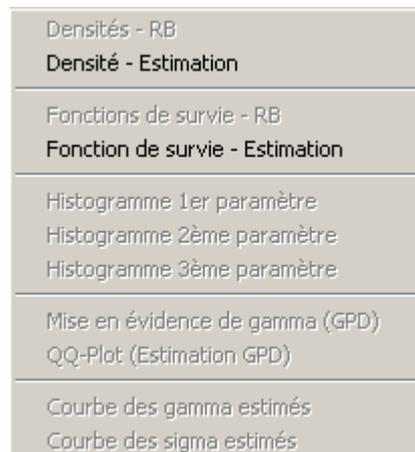
Le logiciel retourne la valeur des paramètres estimés et affiche automatiquement l'histogramme des données sur lequel il superpose la densité de la loi choisie (ici Gamma),

tracée avec les paramètres nouvellement estimés (cf. Fenêtre 41). Le programme demande au préalable un nombre de classes pour dessiner l'histogramme (cf. Fenêtre 11).



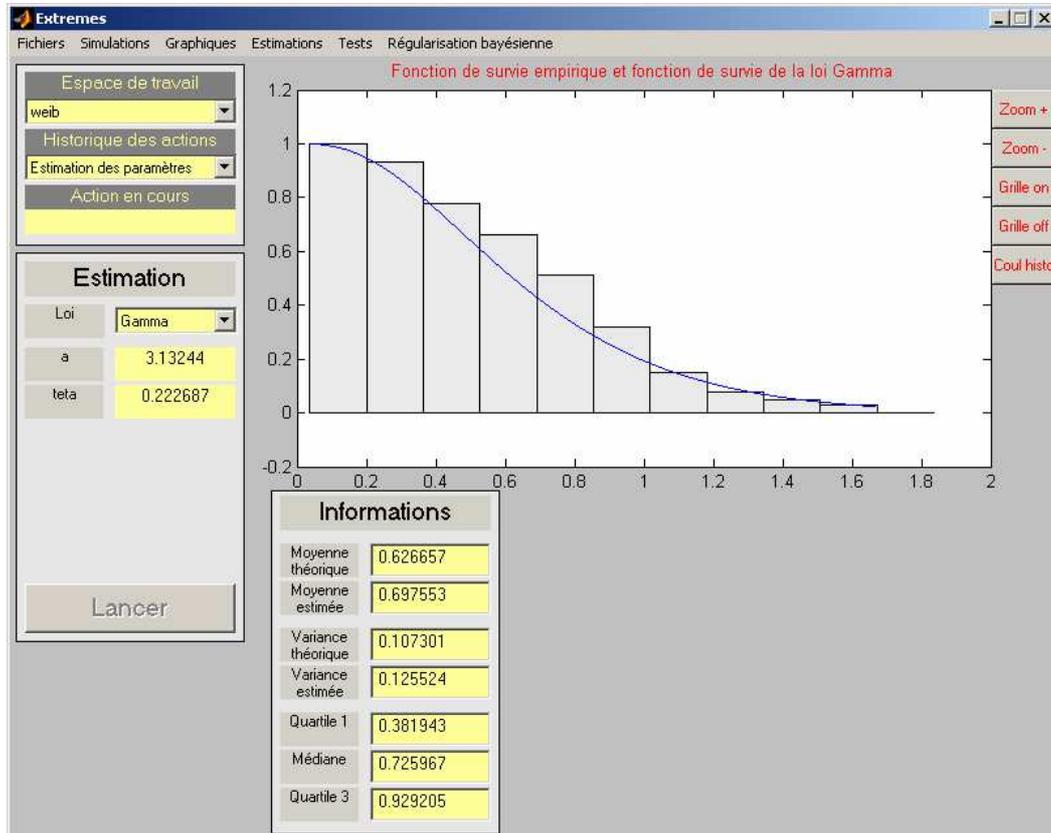
Fenêtre 41 - estimation des paramètres avec un seul jeu de données

EXTREMES a estimé les paramètres d'une loi Gamma $G(a, \theta)$ avec $a = 3.13244$ et $\theta = 0.222687$. La densité d'une loi $G(3.13244, 0.222687)$ a été dessinée sur l'histogramme. En se plaçant sur le graphique avec la souris et en cliquant sur le bouton droit, un menu contextuel apparaît à l'écran :



Fenêtre 42 - Menu contextuel du graphique

Deux rubriques peuvent être sélectionnées par l'utilisateur : **Densité – Estimation** et **Fonction de survie – Estimation**. La première affiche la même chose que la Fenêtre 41. L'autre affiche la fonction de survie empirique et la fonction de survie calculée avec les paramètres estimés (on redemande un nombre de classes à l'utilisateur) :

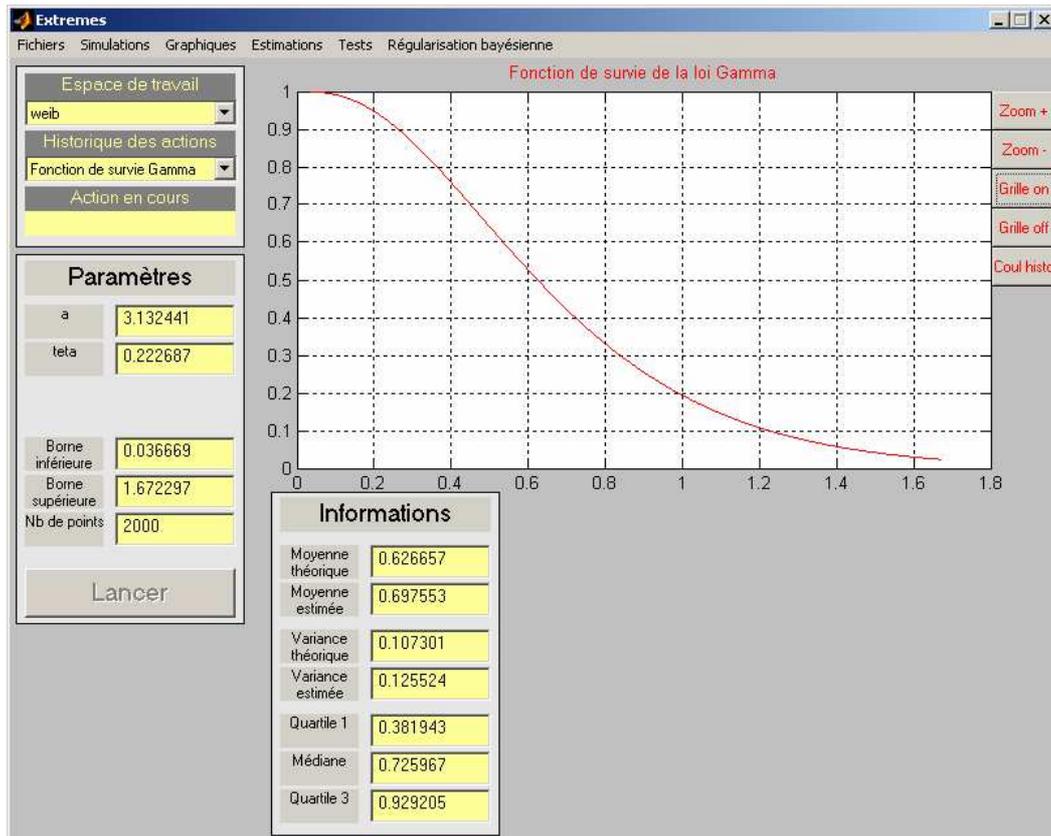


Fenêtre 43 - Fonction de survie empirique et fonction de survie calculée avec les paramètres estimés

On peut revenir à l'écran précédent de nouveau en se plaçant sur le graphique avec la souris et en cliquant sur le bouton droit. L'utilisateur choisira alors **Densité – Estimation**, un nombre de classes sera encore à fournir.

Les deux actions faites en plus (c'est-à-dire le calcul de la densité et de la fonction de survie) ont été ajoutées dans la liste déroulante de l'**historique des actions**. Ainsi, elles peuvent être visualisées à partir de cette même liste, simplement en cliquant sur leur nom, par exemple **Fonction de survie Gamma** (cf. Fenêtre 44).

Remarque 2 : Pour que le menu contextuel s'affiche, il se faut se placer sur le fond du graphique et **non pas** sur l'histogramme ou la courbe. Par ailleurs, dès que l'utilisateur cliquera sur un autre menu, sur **Historique des actions** ou changera d'**espace de travail**, les deux rubriques **Densité-Estimation** et **Fonction de survie – Estimation** ne seront plus disponibles. Pour les rendre de nouveau visibles, il suffit de dérouler la liste des actions et de choisir **Estimation des paramètres**. Le graphique s'actualisera et ces 2 opérations seront de nouveau possibles.



Fenêtre 44 - Fonction de survie réaffichée

En mode jeu de données multiples :

Simulons par exemple 100 échantillons de taille 100, tous issus d'une loi de Weibull. En cliquant sur **Estimation des paramètres**, une boîte de dialogue apparaît à l'écran et demande à l'utilisateur s'il veut travailler sur tous les jeux de données ou non (**Oui** par défaut) :



Fenêtre 45 -Demande à l'utilisateur le mode multi données ou non

Si **Non** est choisi par l'utilisateur, l'estimation des paramètres se fera uniquement sur le premier jeu de données (se reporter au début du paragraphe [Estimation des paramètres](#) p. 26). En revanche, si la réponse est affirmative, le programme va estimer les paramètres pour tous les jeux de données et rendra la moyenne et la variance empirique de ceux-ci. Les histogrammes (dont on doit choisir le nombre de classes) permettent de visualiser leur

distribution empirique (cf. Fenêtre 47). Si on choisit par exemple d'estimer les paramètres d'une loi Gamma, on obtient la boîte suivante :

Estimation

Loi: **Gamma**

a moyen:

teta moyen:

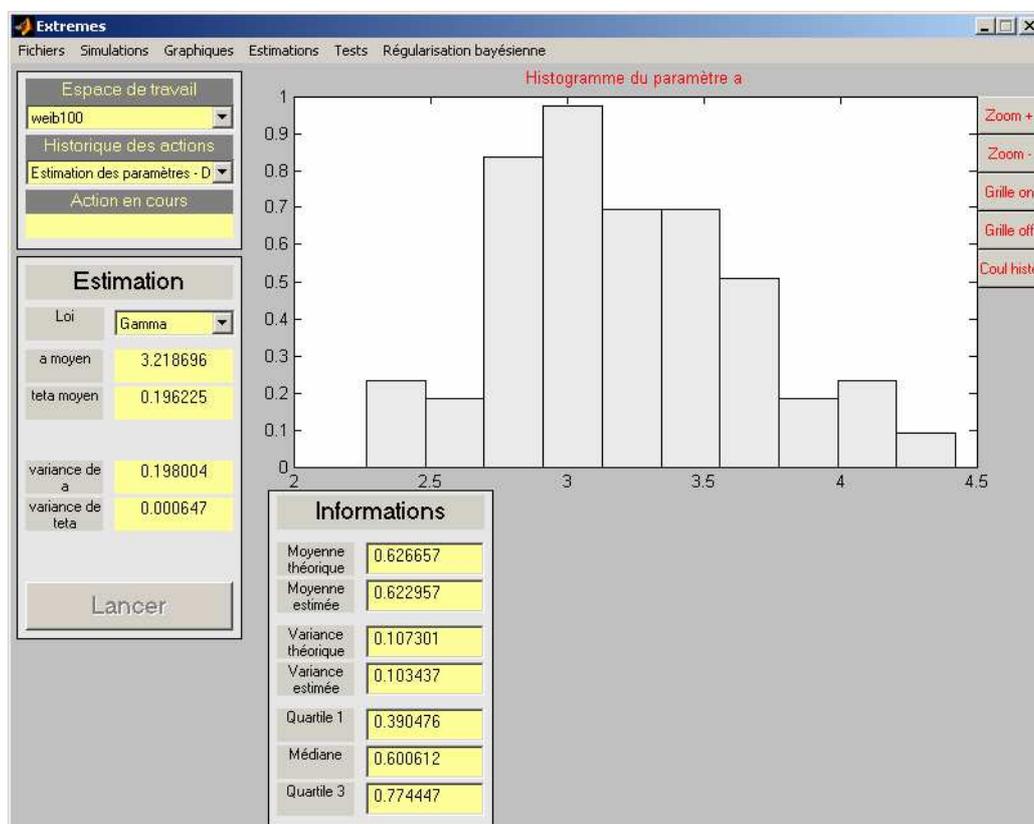
variance de a:

variance de teta:

Lancer

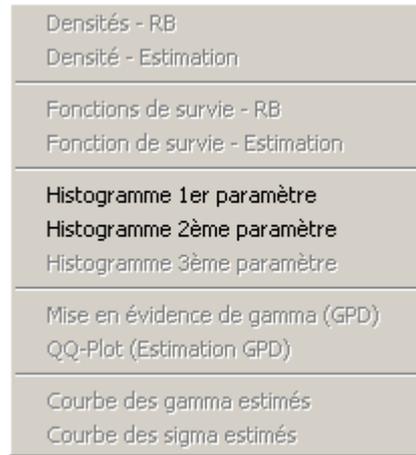
Fenêtre 46 – Estimation des paramètres en mode multi données

Un clic sur **Lancer** donne :



Fenêtre 47 - Résultat de l'estimation des paramètres en mode multi données

En se plaçant sur le graphique avec la souris et en cliquant sur le bouton droit, on obtient de nouveau un menu contextuel où apparaissent les rubriques ***Histogramme 1^{er} paramètre*** et ***Histogramme 2^{ème} paramètre*** (dans le cas d'une loi à 2 paramètres). Il suffit ensuite de choisir celui à afficher.



Fenêtre 48 - Menu contextuel du graphique

Remarque 3 : Pour que le menu contextuel s'affiche, il se faut se placer sur le fond du graphique et **non pas** sur l'histogramme ou la courbe. Par ailleurs, dès que l'utilisateur cliquera sur un autre menu, sur ***Historique des actions*** ou changera d'***espace de travail***, les deux rubriques ***Histogramme 1^{er} paramètre*** et ***Histogramme 2^{ème} paramètre*** ne seront plus disponibles. Pour les rendre de nouveau visibles, il suffit de dérouler la liste des actions et de choisir ***Estimation des paramètres DM***. Le graphique s'actualisera et ces 2 opérations seront de nouveau possibles.

- **Indice des valeurs extrêmes**

Cette fonction estime le couple (γ, σ) , paramètres de la loi GPD (cf. eq. [1] p. 4) qui approche la loi des excès. Dans le logiciel, il est possible de travailler sur un seul nombre d'excès ou alors sur un intervalle défini par l'utilisateur. Comme précédemment, cette fonctionnalité à un comportement différent selon le mode d'utilisation choisi (multi données ou non).

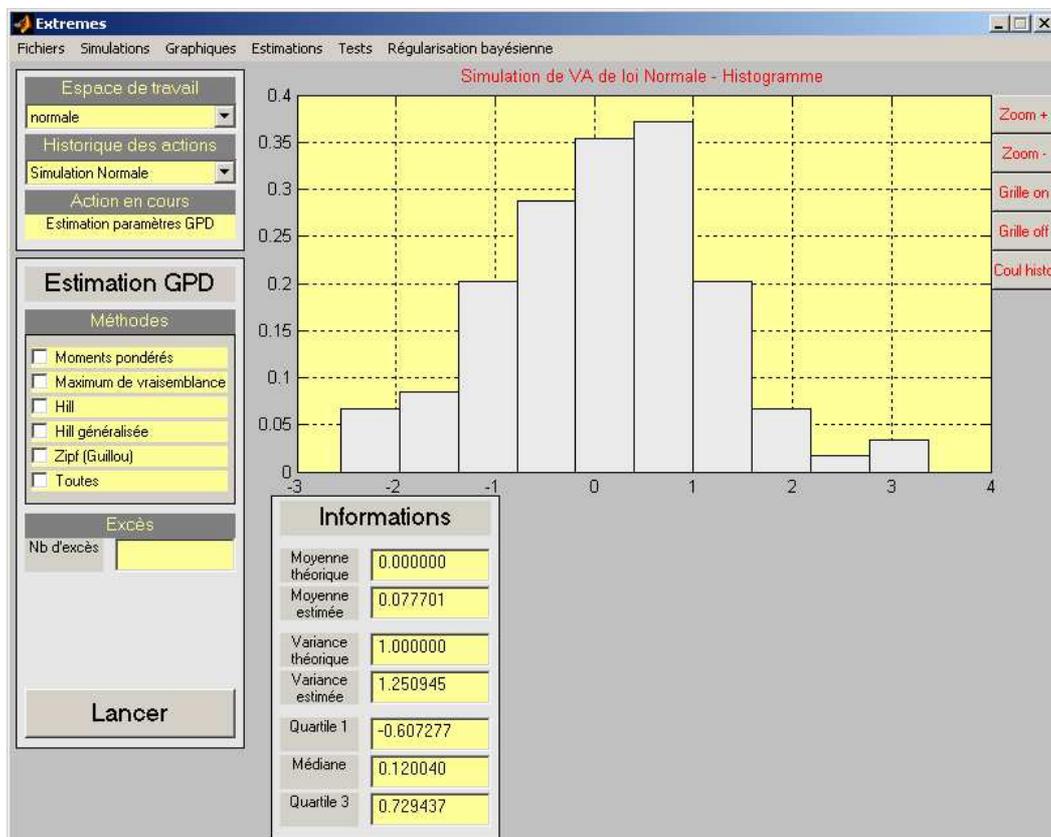
En mode jeu de données simples :

EXTREMES demande d'emblée à l'utilisateur s'il veut travailler sur un ou plusieurs nombres d'excès (***1 seul nombre d'excès*** par défaut) (cf. Fenêtre 49).



Fenêtre 49 - Choix du nombre d'excès

Prenons l'exemple de données simulées à partir d'une loi normale $N(0,1)$. Fabriquons un échantillon de taille 100. En cliquant sur 1 seul nombre d'excès, on obtient l'écran suivant :



Fenêtre 50 - Ecran de saisie pour l'estimation du couple (gamma, sigma)

Dans le cadre *Estimation GPD*, diverses méthodes sont proposées pour estimer l'indice des valeurs extrêmes : la méthode des moments pondérés d'Hosking et Wallis (HW), le maximum de vraisemblance (EMV), les méthode de Hill (Hill) et Hill généralisée (HillG), et enfin la méthode Zipf (Zipf) (cf. Embrechts¹). Un nombre d'excès doit être également stipulé.

¹ Embrechts P., Klüppelberg C., Mikosh T., Modelling extremal events – Springer-Verlag, Applications of mathematics, 1997.

Remarque 4 : Pour calculer le couple (γ, σ) avec les 3 dernières méthodes (par exemple Hill généralisée), le programme n'a pas besoin de construire la liste des excès puisque ces paramètres sont estimés à partir du logarithme des statistiques ordonnées. Or un échantillon peut comporter des données négatives. Dans de tels cas, le logiciel renvoie une valeur maximum pour le nombre d'excès afin de pouvoir évaluer l'indice des valeurs extrêmes si l'utilisateur a inscrit un nombre d'excès trop important (par exemple 90 alors que la taille de l'échantillon est 100) (cf. Fenêtre 51). En revanche, la liste des excès est formée quand la 1^{ère} ou la 2^{ème} méthode est choisie. L'utilisateur peut prendre un nombre d'excès aussi grand qu'il le veut compris entre 2 et la taille de l'échantillon. Si l'échantillon ne contient pas de données négatives, le choix du nombre d'excès ne constitue pas un problème.

Remarque 5 : Si l'utilisateur choisit l'une ou l'autre des 2 premières méthodes (Moments pondérés ou Maximum de vraisemblance) **et** un nombre d'excès égal à la taille de l'échantillon de départ, le calcul du couple (γ, σ) sera effectué à partir de l'échantillon tout entier et non pas à partir des excès.



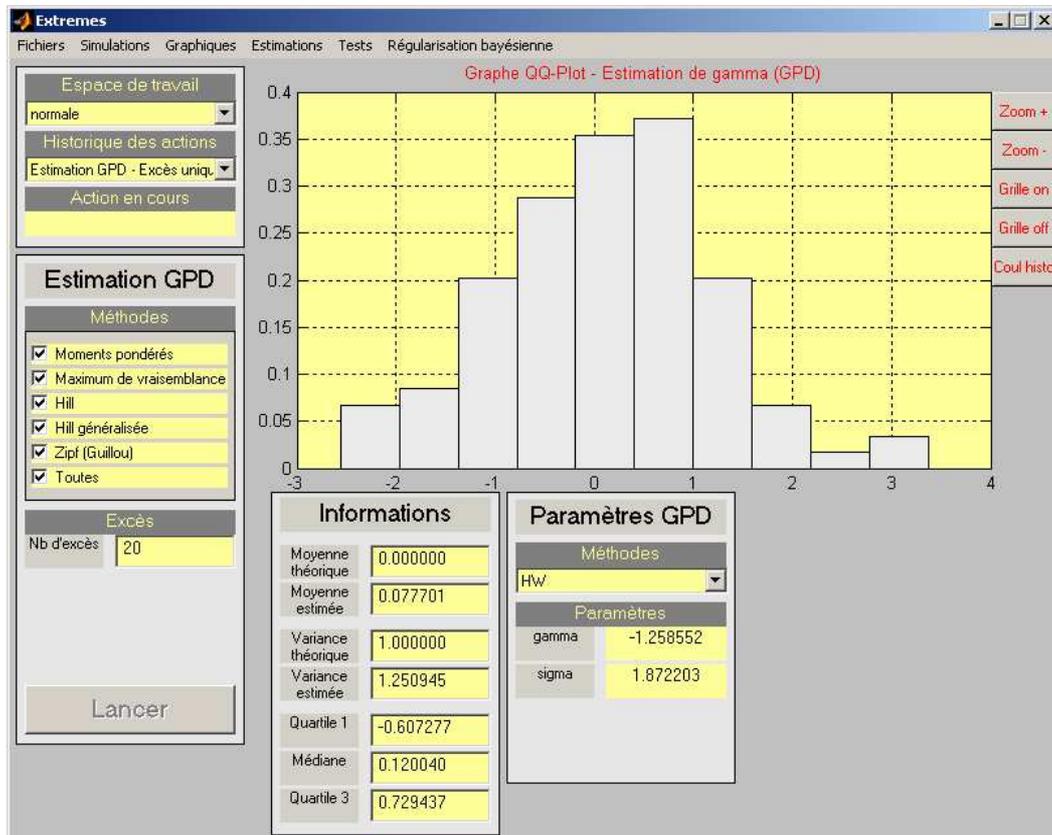
Fenêtre 51 - Nombre d'excès trop important

Il est possible de sélectionner plusieurs méthodes (le champ **Toutes** permet de cocher toutes les méthodes d'un seul clic) :



Fenêtre 52 - Sélection de toutes les méthodes

En cliquant sur Lancer, une boîte de résultat apparaît avec les valeurs de γ et σ pour les méthodes sélectionnées (il n'y a pas de sorties graphiques dans ce cas) :



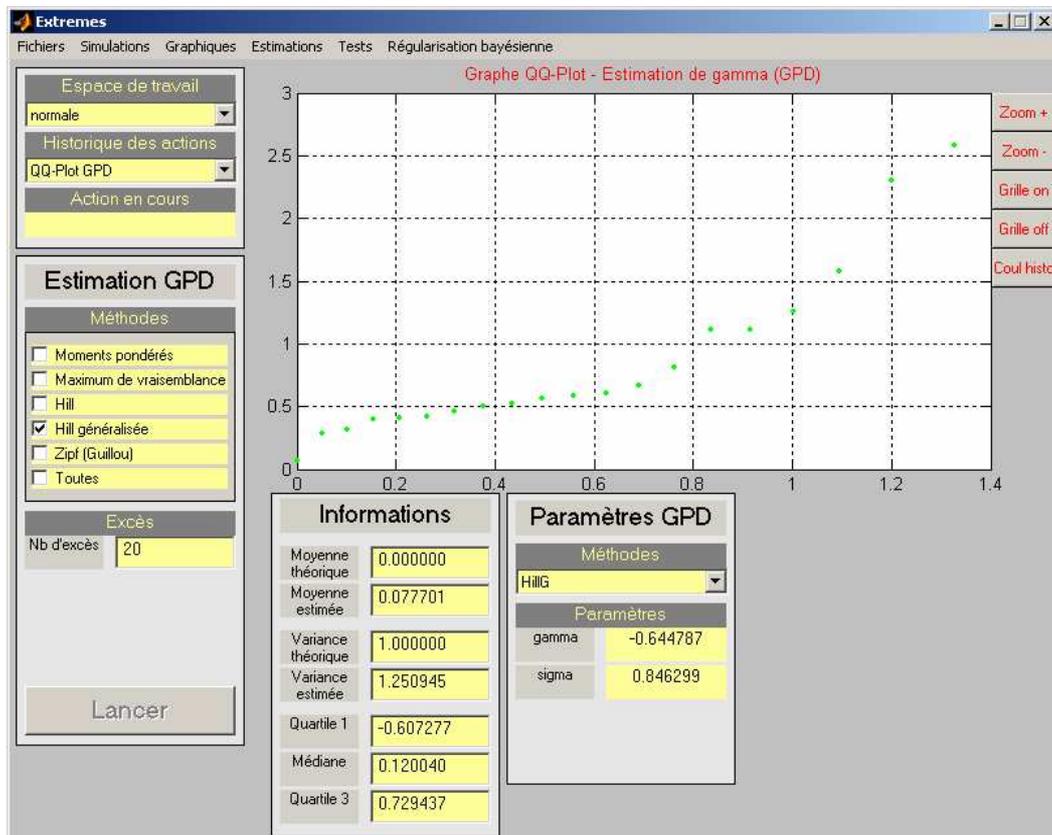
Fenêtre 53 - Boîte de résultat de l'estimation du couple (gamma, sigma)

Pour visionner les estimations fournies par les autres méthodes, il suffit de déployer la liste déroulante des *Méthodes* dans la boîte résultat *Paramètres GPD* :



Fenêtre 54 - Boîte résultat des estimations du couple (gamma, sigma)

Lorsqu'une seule méthode est cochée par l'utilisateur, des sorties graphiques s'ajoutent à la boîte de résultats (2 types de QQ-plot) :

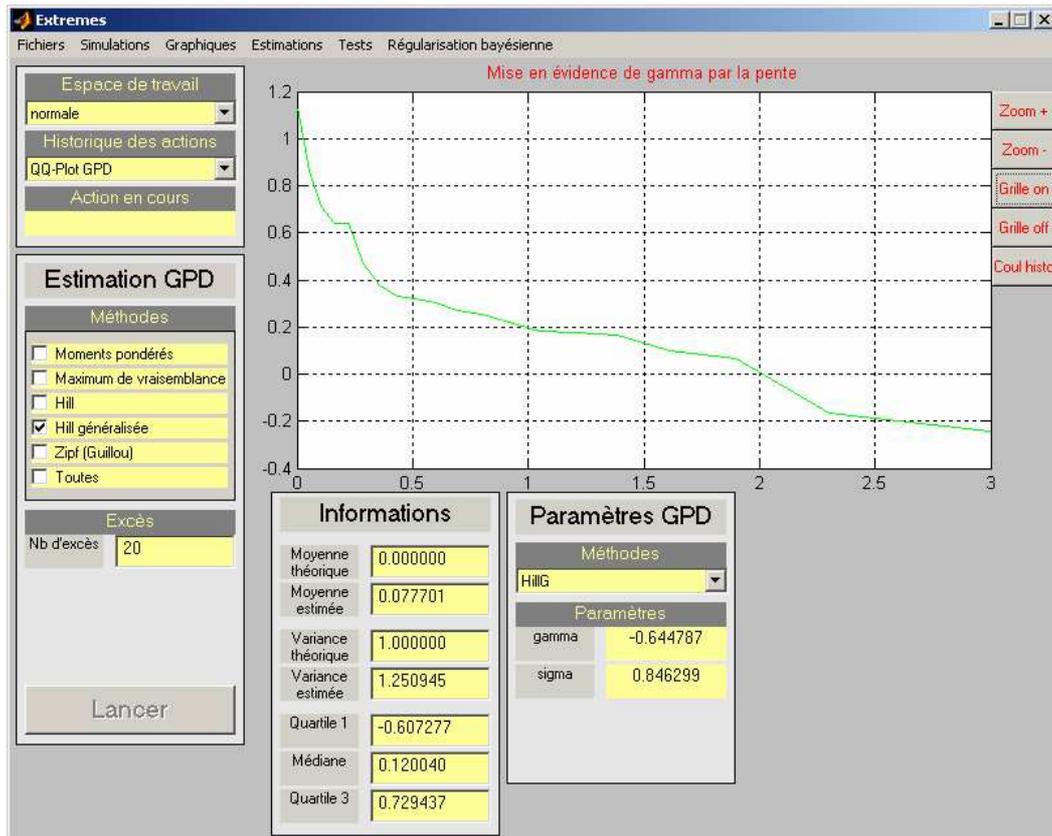


Fenêtre 55 - QQ-plot pour l'estimation de gamma

Si les excès suivent une loi GPD, alors les points du QQ-plot doivent être alignés et la pente de la courbe estimée σ . Un clic droit sur le graphique donne :



Fenêtre 56 - Menu contextuel du graphique pour l'estimation de (gamma, sigma)



Fenêtre 57 - Mise en évidence de gamma par la pente

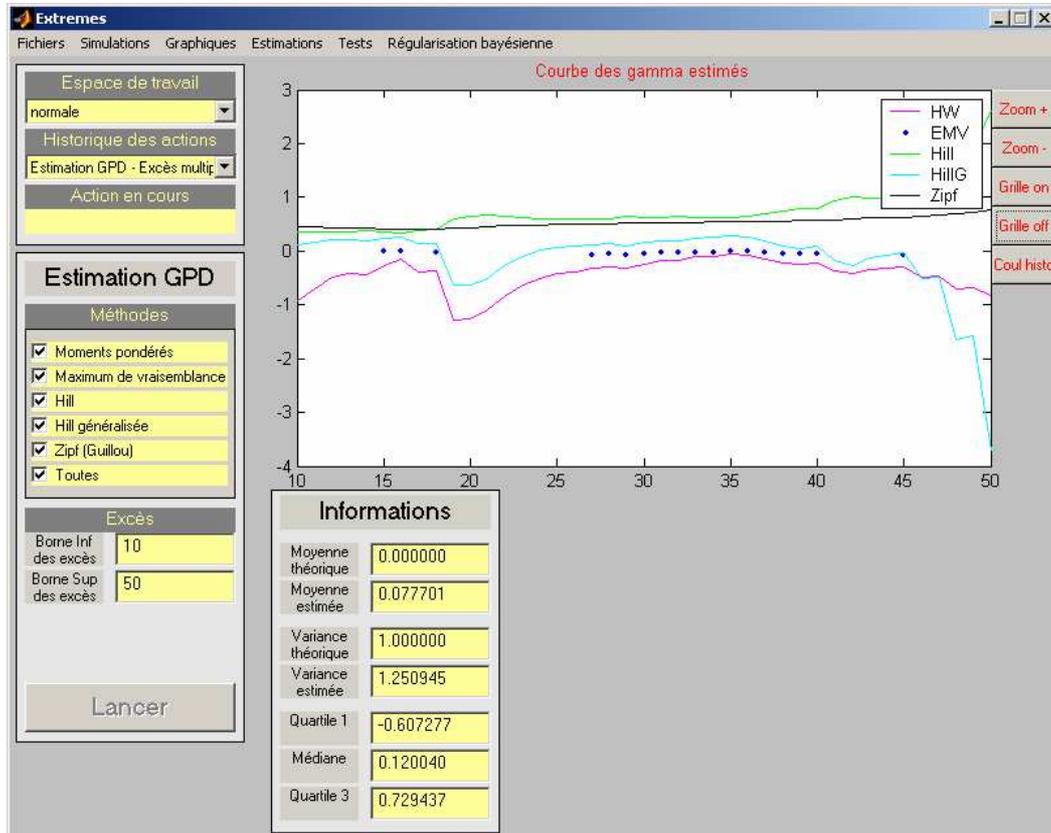
La pente de la droite estime γ . Le résultat de l'estimation par la méthode de Hill généralisée s'inscrit de nouveau dans la boîte *Paramètres GPD*.

Si maintenant nous revenons sur la **Fenêtre 49** et cliquons sur *Plusieurs nombres d'excès*. La boîte *Estimation GPD* devient :

The screenshot shows the 'Estimation GPD' dialog box. The 'Méthodes' section has 'HillG' selected. The 'Excès' section has 'Borne Inf des excès' and 'Borne Sup des excès' fields. A 'Lancer' button is at the bottom.

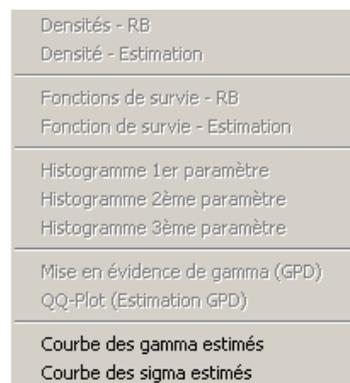
Fenêtre 58 - Cadre Estimation GPD avec plusieurs nombres d'excès

L'utilisateur doit définir un intervalle de nombres d'excès (champs *Borne Inf des excès*, *Borne sup des excès*) (par exemple ici entre 10 et 50). **EXTREMES** va calculer la valeur de l'estimation du couple (γ, σ) pour chaque nombre d'excès entre 10 et 50 compris et pour chaque méthode sélectionnée.

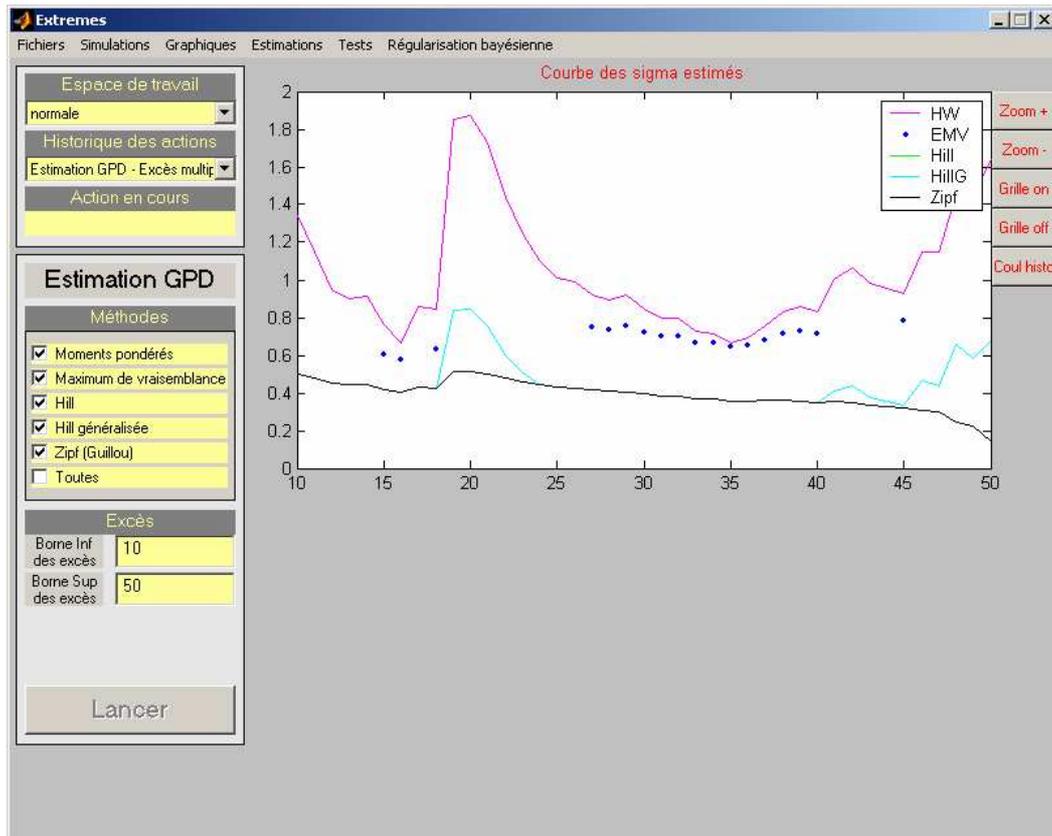


Fenêtre 59 – Courbe des estimations de gamma

Cette fenêtre ne présente que l'estimation de γ . En abscisse, on trouve les différents nombres d'excès et en ordonnées les estimations de γ pour chaque méthode. Pour observer celles de σ , il suffit de faire afficher le menu contextuel du graphique en cliquant sur le bouton droit de la souris :



Fenêtre 60 - Passage de l'estimation de gamma à celle de sigma par le menu contextuel du graphique



Fenêtre 61 - Courbes des estimations de sigma

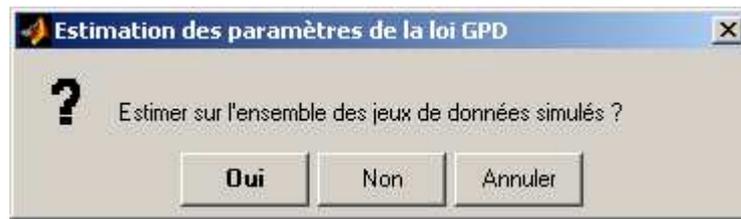
A noter que l'estimation de σ pour *l'estimateur de Hill* est la même que pour *Zipf*. Les 2 courbes sont donc superposées.

Remarque 6 : L'estimation de (γ, σ) par la méthode du *maximum de vraisemblance* (points bleus) n'apparaissent pas toujours sur le graphique. En fait, pour certaines valeurs du nombre d'excès, le calcul est impossible si une condition mathématique n'est pas vérifiée. Si *EMV* n'apparaît pas dans la légende, aucun nombre d'excès parmi la plage choisie ne satisfait la condition mathématique de calcul.

En mode jeux de données multiples :

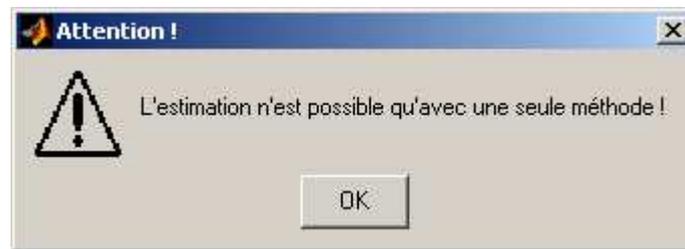
Le logiciel détecte si plusieurs jeux de données sont présents dans le même espace de travail. Si tel est le cas, le mode de fonctionnement et les résultats de la fonctionnalité *Indice des valeurs extrêmes* varient un peu. Pour cela, simulons 100 échantillons de taille 100 tous issus d'une loi Normale $N(0,1)$ dans un *espace de travail* que nous appelons *Normale100*.

En cliquant sur le menu qui nous intéresse, le programme affiche une boîte de dialogue questionnant l'utilisateur sur le mode (multi données ou non) de travail à employer (cf. Fenêtre 62) (*Oui* par défaut).

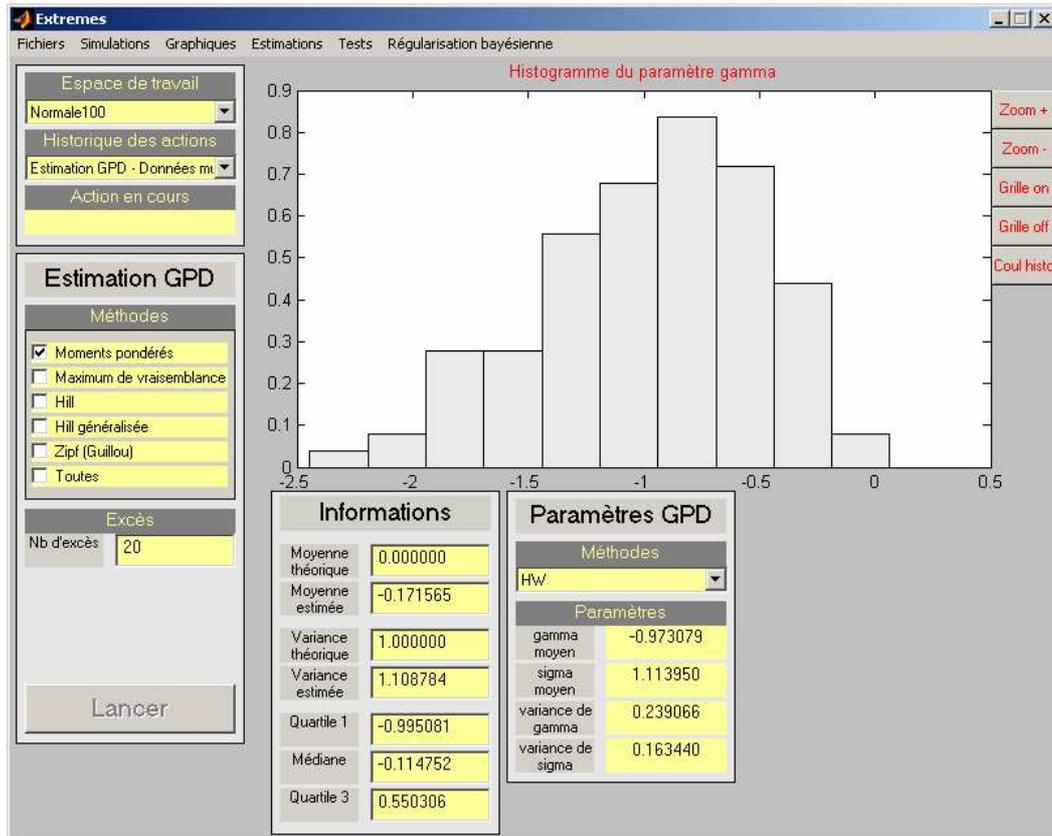


Fenêtre 62 - Choix du mode de travail

Si le mode *jeu de données simple* est sélectionné (en cliquant sur *Non*), se reporter page 32 à la rubrique *En mode jeu de données simple*. Si l'utilisateur choisit *Annuler*, il retourne à son écran précédent. En revanche, s'il choisit de travailler en mode *multi données*, un message s'inscrit pour informer l'utilisateur qu'il n'a le droit de sélectionner qu'une seule méthode parmi les 5 proposées.

Fenêtre 63 - Message d'information concernant l'estimation de (γ , σ) en mode multi données et un seul nombre d'excès

En considérant que des données peuvent être négatives, il convient de renvoyer le lecteur à la *Remarque 4* page 33. Sélectionnons par exemple la méthode *Moments pondérés* et prenons 20 comme nombre d'excès. Le programme va estimer le couple (γ, σ) par la méthode de Hosking et Wallis pour chacun des 100 échantillons. La moyenne et la variance empiriques de ces 2 estimateurs seront renvoyées à l'interface graphique respectivement dans les champs *gamma moyen / sigma moyen* et *variance de gamma / variance de sigma* de la boîte de résultat *Paramètres GPD*. L'histogramme des valeurs de γ sera affiché. Pour éditer celui de σ , il faut procéder comme expliqué pour *l'estimation des paramètres en mode multi données* page 32. A chaque fois, un nombre de classes est demandé à l'utilisateur.



Fenêtre 64 - Distribution de l'estimateur de gamma en mode multi données

- Quantile extrême

Ce menu s'utilise de manière identique au précédent, l'estimation d'un quantile extrême s'effectuant à partir de l'estimation du couple (γ, σ) selon la formule ci-dessous. Les boîtes de saisie sont analogues à celle de l'indice des valeurs extrêmes. Seul l'ordre du quantile est à renseigner en plus (cette valeur doit être strictement comprise entre 0 et 1).

$$q_{GPD} = u_n + \frac{\hat{\sigma}_n}{\hat{\gamma}_n} \left(\left(\frac{np_n}{k_n} \right)^{-\hat{\gamma}_n} - 1 \right)$$

où k_n désigne le nombre d'excès au-delà du seuil u_n et $\hat{\sigma}_n$ et $\hat{\gamma}_n$ sont des estimateurs des paramètres de la loi GPD.

En mode jeu de données simple :

La Fenêtre 49 s'affiche d'abord. L'utilisateur obtient l'une des 2 boîtes de saisie ci-dessous selon qu'il veuille travailler sur plusieurs nombres d'excès ou non.

Fenêtre 65 - Ecran de saisie pour le calcul des quantiles extrêmes

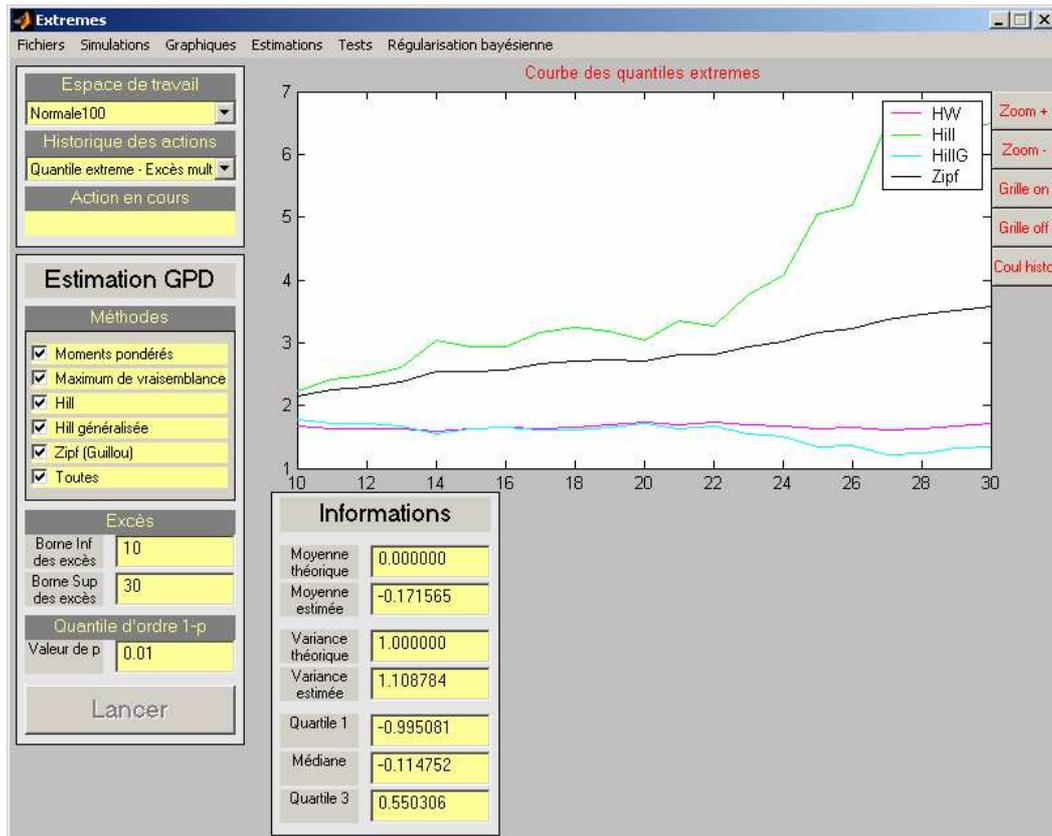
Par exemple, cochons toutes les méthodes disponibles, prenons 20 comme nombre d'excès et calculons le quantile d'ordre 1-0.01 (champ *Valeur de p* = 0.01). La boîte de résultats est de la forme :

Paramètres GPD	
Méthodes	
	HW
Paramètres	
gamma	-0.935338
sigma	0.976456
quantile	1.734350

Fenêtre 66 - Résultats de l'estimation du quantile extrême

Pour visionner les estimations fournies par les autres méthodes, il suffit de procéder comme indiqué page 35 *Fenêtre 54*.

Si plusieurs nombres d'excès ont été choisis par l'utilisateur, on obtient les courbes des valeurs des quantiles pour chaque méthode selon le nombre d'excès :



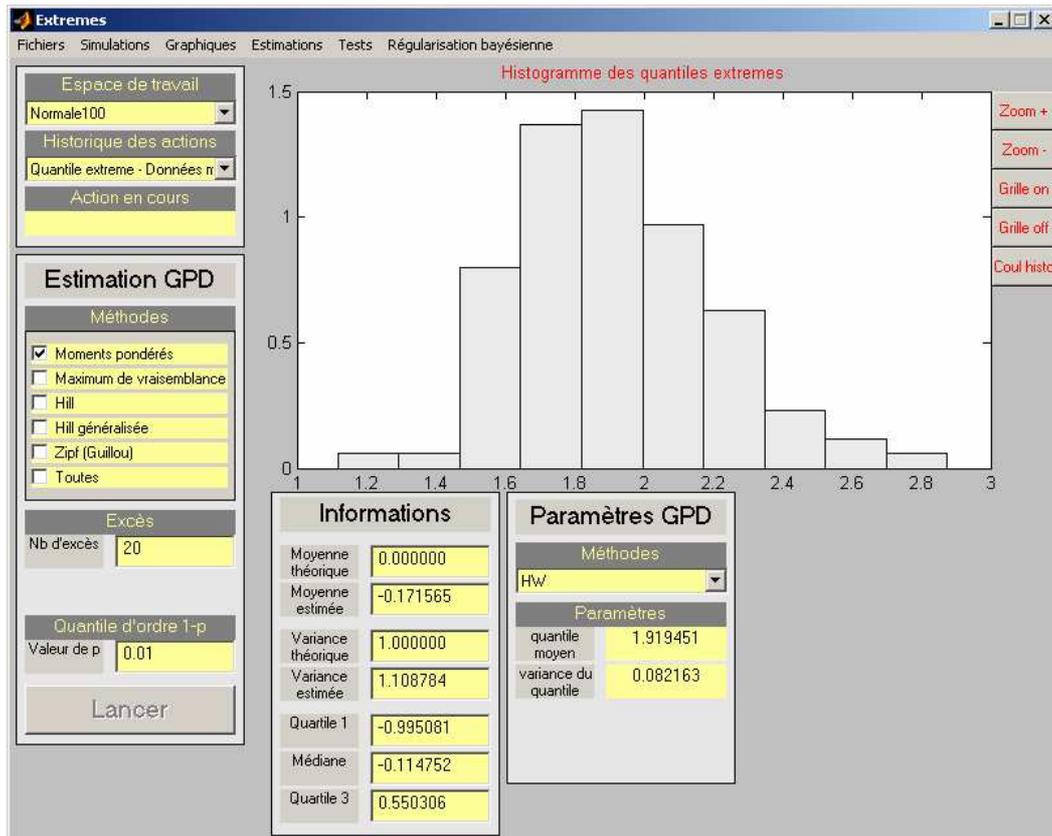
Fenêtre 67 - Courbes des quantiles extrêmes pour chaque méthode selon le nombre d'excès

En mode jeux de données multiples :

Les mêmes fenêtres que lors de l'*Indice des valeurs extrêmes* s'affichent à l'écran (cf. Fenêtre 62). L'histogramme de la distribution des quantiles extrêmes estimés est édité. Le but de cet affichage est par exemple de comparer l'estimation obtenue avec l'estimation paramétrique (cf. *Quantile paramétrique*). Dans la boîte de résultats sont affichées les valeurs du quantile extrême moyen ainsi que sa variance. Soulignons de nouveau que pour cette fonctionnalité, une seule méthode d'estimation pour le couple (γ, σ) et donc pour le quantile est permise.

Paramètres GPD	
Méthodes	
	HW
Paramètres	
quantile moyen	1.919451
variance du quantile	0.082163

Fenêtre 68 - Résultat de l'estimation des quantiles extrêmes en mode multi données



Fenêtre 69 - Histogramme de la distribution des quantiles extrêmes

- Quantile paramétrique

Cette fonction permet d'estimer un quantile d'ordre $1 - p$ de manière paramétrique avec les 11 modèles du catalogue en mode jeu de données simple. Si l'utilisateur choisit le mode jeux de données multiples, le quantile ne peut être calculé qu'avec un seul modèle. L'ordre p du quantile doit être entré par l'utilisateur. Si la présence de données négatives est révélée, le quantile sera calculé seulement pour les modèles acceptant ce genre de données (Normale, Uniforme, etc.).

En mode jeu de donnée simple :

L'écran suivant est présenté à l'utilisateur :

Q. paramétrique

Choix des lois

Quantile d'ordre 1-p

Valeur de p

Lancer

Fenêtre 70 – Boîte de saisie pour les quantiles paramétriques

Pour choisir les modèles, il suffit de cliquer sur *Choix des lois* :

Choix des lois

<input type="checkbox"/> Normale	<input type="checkbox"/> Chi2	<input type="checkbox"/> Uniforme
<input type="checkbox"/> Lognormale	<input type="checkbox"/> Student	<input type="checkbox"/> Beta
<input type="checkbox"/> Exponentielle	<input type="checkbox"/> Pareto	<input type="checkbox"/> Toutes
<input type="checkbox"/> Gamma	<input type="checkbox"/> GPD	
<input type="checkbox"/> Weibull		

OK

Fenêtre 71 - Sélection des modèles

La case *Toutes* permet de sélectionner toutes les lois simultanément. C'est ce que nous ferons dans l'exemple (le jeu de données est un échantillon de loi Normale).

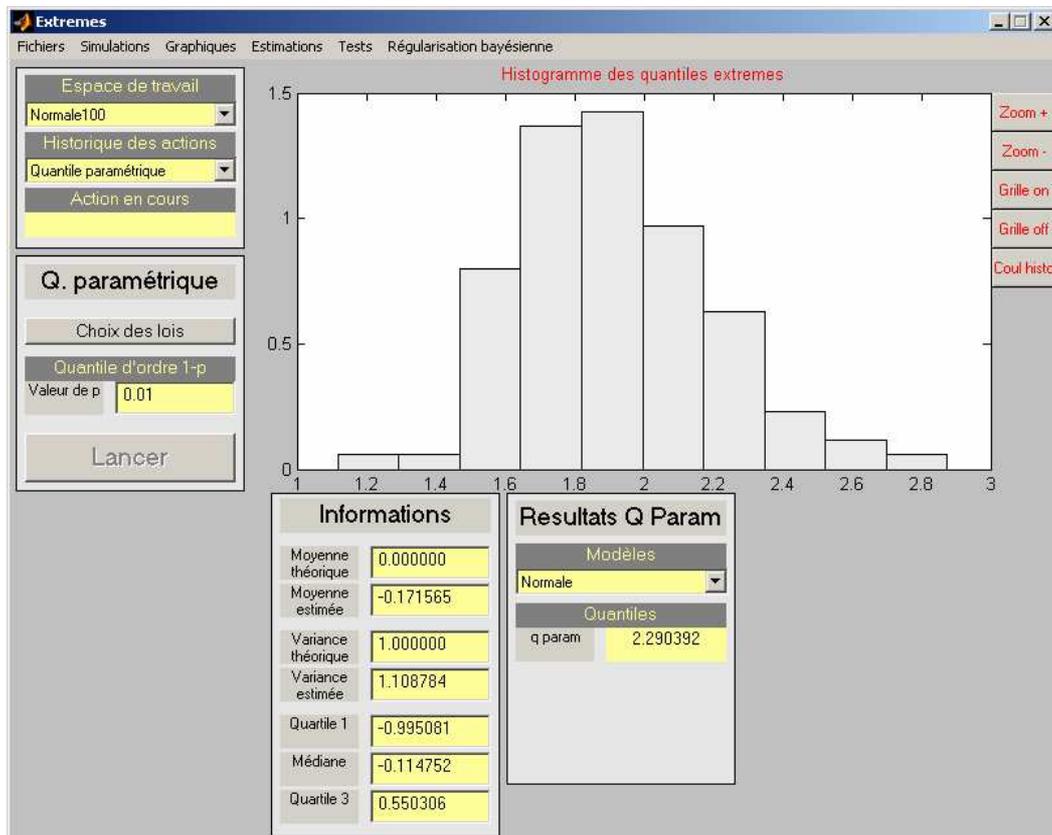
Choix des lois

<input checked="" type="checkbox"/> Normale	<input checked="" type="checkbox"/> Chi2	<input checked="" type="checkbox"/> Uniforme
<input checked="" type="checkbox"/> Lognormale	<input checked="" type="checkbox"/> Student	<input checked="" type="checkbox"/> Beta
<input checked="" type="checkbox"/> Exponentielle	<input checked="" type="checkbox"/> Pareto	<input checked="" type="checkbox"/> Toutes
<input checked="" type="checkbox"/> Gamma	<input checked="" type="checkbox"/> GPD	
<input checked="" type="checkbox"/> Weibull		

OK

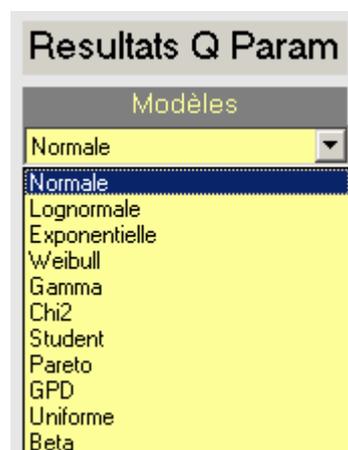
Fenêtre 72 - Sélection de modèles (2)

En cliquant sur **OK**, la boîte disparaît. Si on prend 0.01 comme *Valeur de p* et qu'on clique sur **Lancer**, l'écran ci-dessous est imprimé :



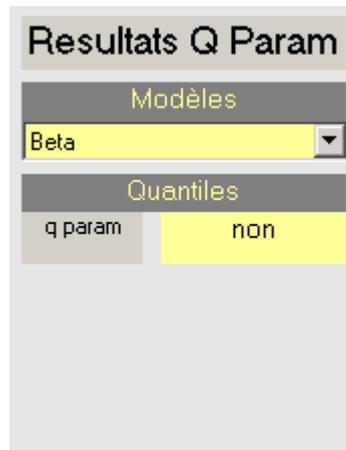
Fenêtre 73 - Résultat du calcul du quantile paramétrique

La boîte libellée *Résultats Q Param* apparaît et la valeur du quantile calculé est affichée dans le champ *q param*. En déroulant la liste des modèles, il est possible de voir les calculs des quantiles associés à toutes les lois sélectionnées.



Fenêtre 74 - Vue de la valeur du quantile pour chaque modèle sélectionné

Par exemple, le quantile paramétrique pour un modèle Beta n'a pas été calculé car le programme a détecté des données négatives :



Fenêtre 75 - Quantile paramétrique non calculé

En mode jeux de données multiples :

Quand **EXTREMES** décèle plusieurs jeux de données, il demande à l'utilisateur quel mode de travail il veut utiliser (La réponse par défaut est mentionnée en gras) :



Fenêtre 76 - Choix du mode de travail

Annuler revient à l'écran précédent. **Non** sélectionne le mode jeu de données simple et le ou les quantiles seront évalués comme précédemment (cf. page 44). Si l'utilisateur choisit le mode **données multiples**, alors il ne pourra sélectionner qu'un seul modèle. Un message s'affiche pour l'en informer :



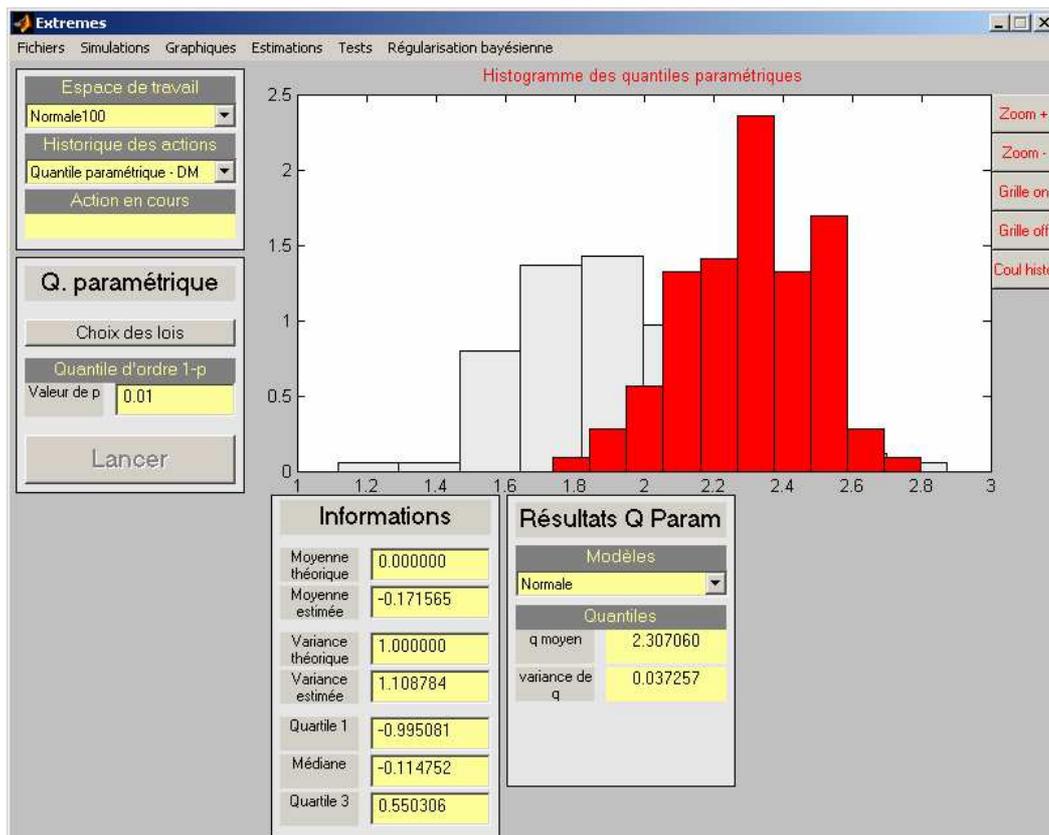
Fenêtre 77 - Sélection de plusieurs modèles impossible

La procédure pour choisir un modèle est la même qu'à la page 45, Fenêtre 71. Si la compatibilité entre le modèle choisi et les données n'est pas assurée (par exemple un modèle Exponentielle et des données négatives), un avertissement est affiché à l'écran :



Fenêtre 78 - Modèle incompatible

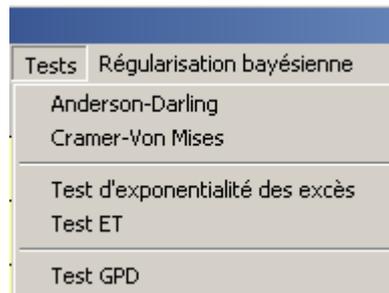
Sélectionnons un modèle Normale et prenons 0.01 comme valeur de p . Le quantile paramétrique est calculé pour chaque jeu de données et la moyenne et la variance empirique sont affichées dans la boîte **Résultats Q Param**. L'histogramme de la distribution des quantiles estimés est affiché (superposée à l'histogramme des quantiles extrêmes) :



Fenêtre 79 - Histogramme et résultat du calcul des quantiles

3.2.5. Le menu Tests

Ce menu rassemble les différentes procédures de tests d'adéquation centrale, d'exponentialité des excès, de test ET (Exponential Tail) et de test GPD (Generalized Pareto Distribution) (cf. Fenêtre 80). L'utilisation de ce menu est différente selon le mode de travail choisi par l'utilisateur.



Fenêtre 80 - Menu Tests

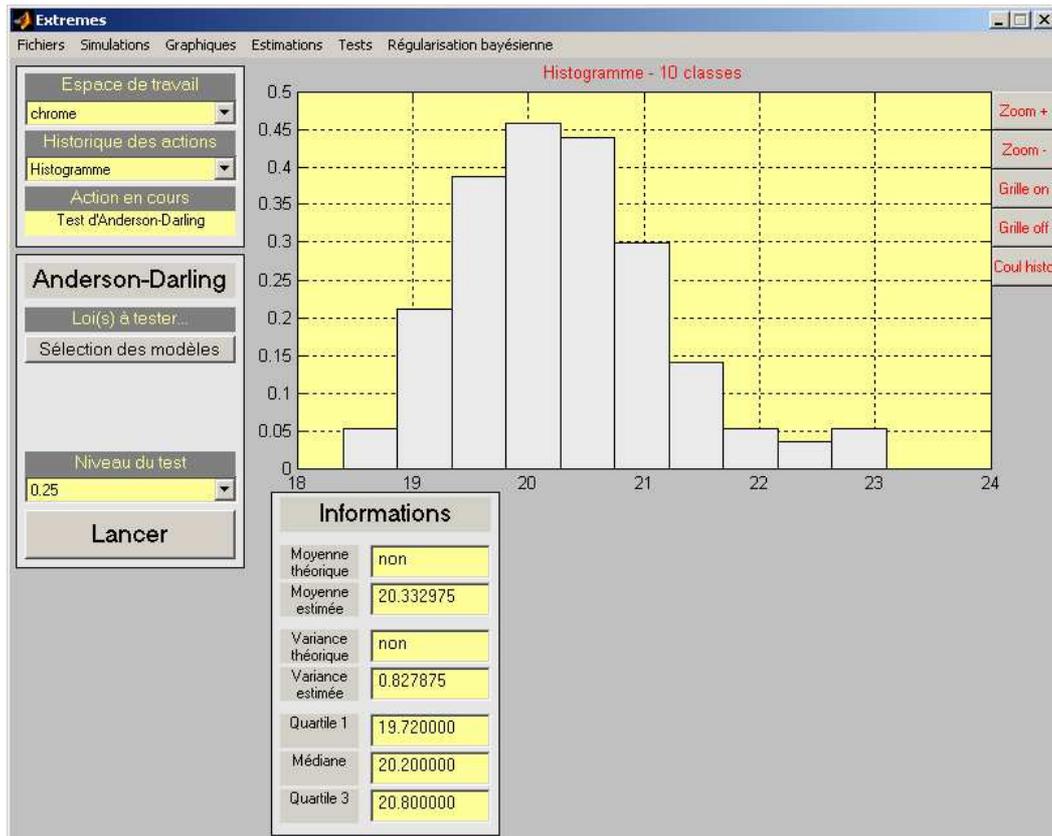
Les tests centraux (*Anderson-Darling* et *Cramer-Von Mises*) sont regroupés dans une sous catégorie. Il en est de même pour *Test d'exponentialité des excès* et *Test ET* qui est construit pour les modèles appartenant au domaine d'attraction de Gumbel. Enfin, le *Test GPD*, défini pour les lois appartenant à tous les domaines d'attraction, est lui aussi à part.

- [Anderson-Darling](#)

Le test d'Anderson-Darling est un test classique d'adéquation à la partie centrale de la distribution. Cette fonction peut être appelée dès qu'un ou plusieurs jeux de données ont été simulés ou importés. Les résultats renvoyés par *EXTREMES* sont de nature différente selon le mode de travail (multi données ou données simples).

En mode jeu de données simple :

Par exemple, ouvrons le fichier nommé *cr.txt* (fichier de données réelles de teneur en chrome). Ce jeu de données compte 121 éléments. Pour se représenter ces données, construisons d'abord un histogramme à 10 classes pour avoir plus de renseignements. Appliquons alors le test d'Anderson-Darling. L'écran suivant apparaît à l'utilisateur :



Fenêtre 81 - Ecran d'accueil pour le test d'Anderson-Darling

Pour choisir les modèles à tester, il faut cliquer sur Sélection de modèles :

The 'Choix des lois' dialog box contains a list of probability distributions with checkboxes:

- Normale
- Lognormale
- Exponentielle
- Gamma
- Weibull
- Chi2
- Student
- Pareto
- GPD
- Uniforme
- Beta
- Toutes

An 'OK' button is located at the bottom center of the dialog.

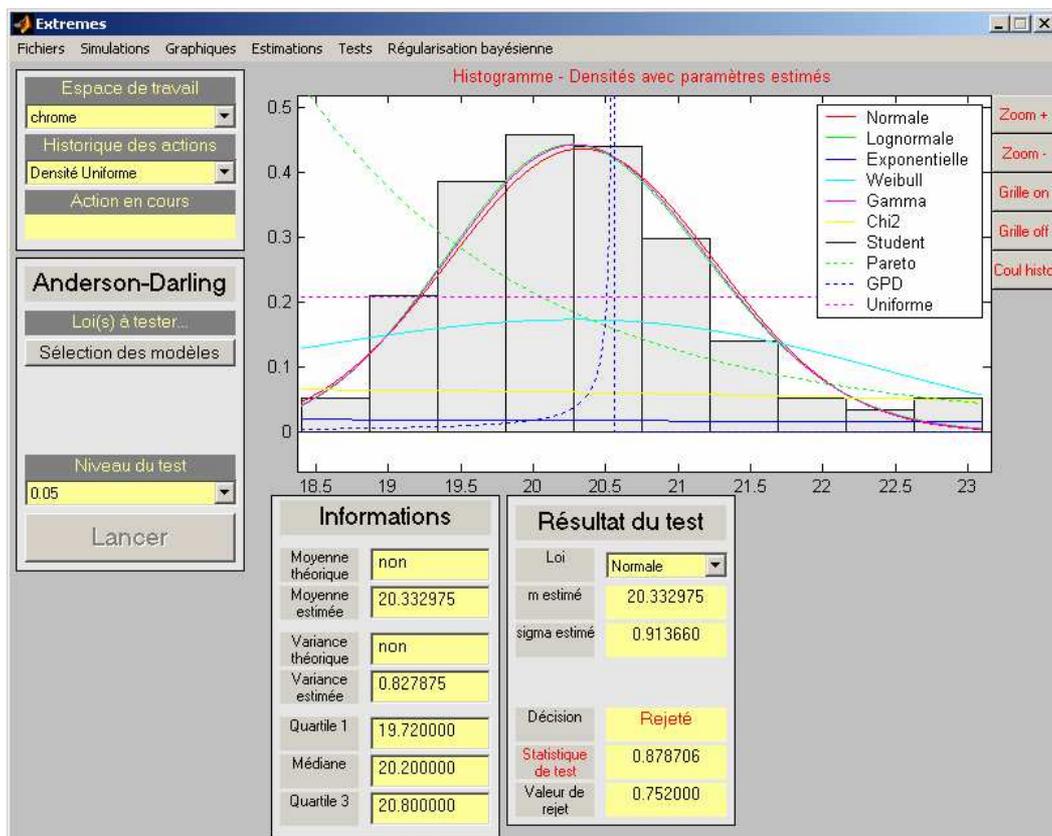
Fenêtre 82 - Choix des modèles

Les différentes lois sont regroupées par domaine d'attraction. La case **Toutes** permet de sélectionner simultanément tous les modèles (ce que nous ferons dans l'exemple). Un clic sur **OK** fait disparaître la boîte. L'utilisateur doit ensuite choisir le niveau du test. 5 valeurs sont proposées (cf. Fenêtre 83) : 0.25, 0.1, 0.05, 0.025 et 0.01.



Fenêtre 83 - Choix du niveau du test

En cliquant sur **Lancer**, le logiciel va tester tous les modèles sélectionnés, afficher les paramètres estimés pour chaque modèle et dessiner l'histogramme des données sur lequel seront superposées les densités de probabilités ajustées.



Fenêtre 84 - Résultats du test d'Anderson-Darling

Les densités des lois dont les noms se trouvent dans la légende ont été tracées. Il se peut que cette liste soit différente de celle commandée au départ. Le programme n'affiche pas les densités des modèles dont les paramètres n'ont pas pu être estimés. Les paramètres estimés sont affichés, ainsi la statistique de test et la valeur de rejet. Si la statistique de test est supérieure à la valeur de rejet, le test est rejeté. Il est accepté sinon.

Pour visualiser le résultat du test pour les autres modèles, il faut cliquer sur la liste déroulante dans la boîte résultat *Résultat du test*.

Résultat du test	
Loi	Normale
m estimé	Normale
sigma estimé	Lognormale
	Exponentielle
	Weibull
	Gamma
	Chi2
	Student
Décision	Pareto
	GPD
	Uniforme
	Beta
Statistique de test	
Valeur de rejet	0.752000

Fenêtre 85 - Liste déroulante contenant tous les modèles testés

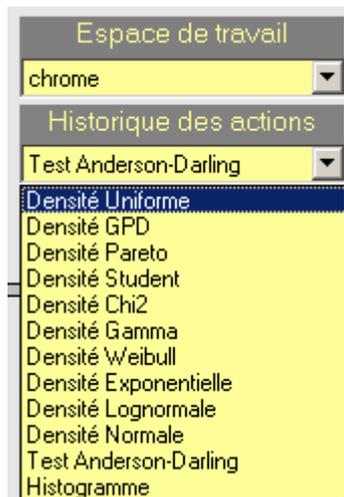
Résultat du test	
Loi	Exponentielle
alpha estimé	20.332975
Décision	Rejeté
Statistique de test	51.150073
Valeur de rejet	1.321000

Résultat du test	
Loi	Beta
a estimé	non
b estimé	non
Décision	Rejeté
Statistique de test	non
Valeur de rejet	non

Fenêtre 86 - Visualisation du résultat du test pour d'autres modèles

Quand le programme n'a pas pu estimer les paramètres de la loi ou testé le modèle (cas de rejet flagrant : support du modèle incompatible avec les données, ...), *non* est affiché.

La liste *Historique des actions* est actualisée et contient désormais toutes les densités tracées (cf. Fenêtre 87).



Fenêtre 87 - Actualisation de la liste de l'historique des actions après le test central

En mode jeux de données multiples :

Simulons par exemple 100 échantillons de taille 100 de loi Gamma $G(2,0.5)$ dans l'espace de travail *gamma100*. Lorsqu'on veut exécuter le test d'Anderson-Darling, **EXTREMES** détecte la présence de plusieurs jeux de données et demande à l'utilisateur de choisir entre le mode multi données ou jeu de données simple :



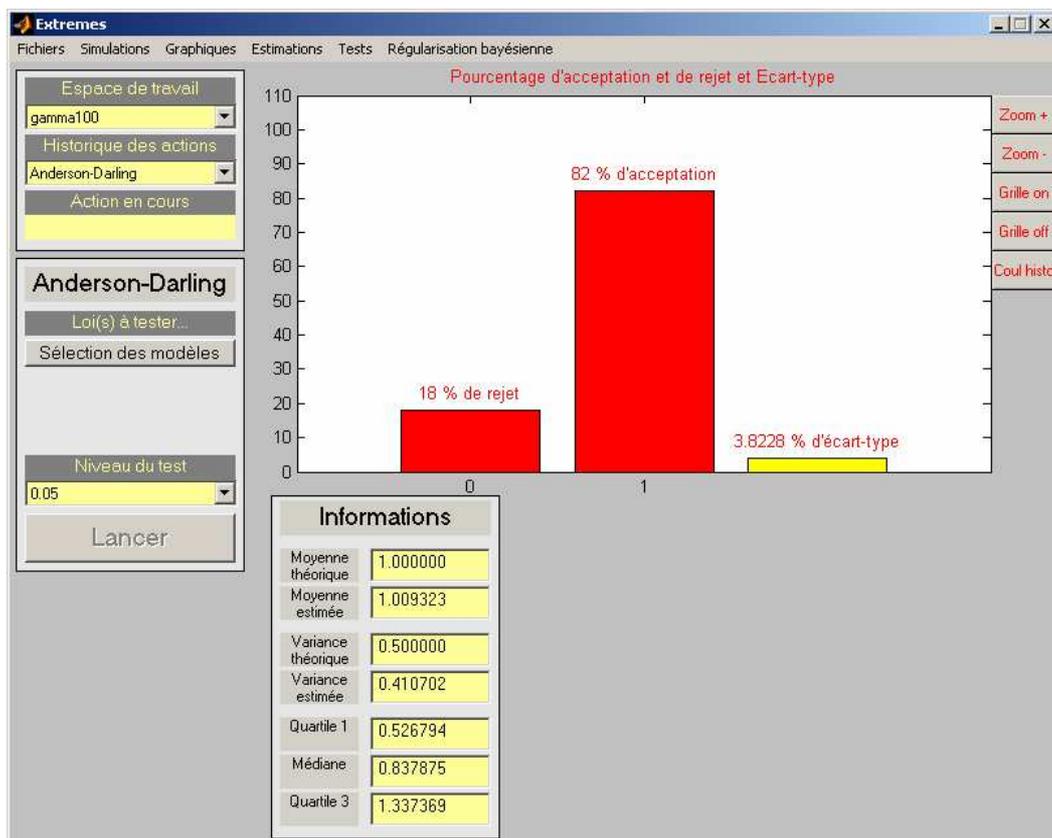
Fenêtre 88 - Choix du mode de travail pour le test

S'il choisit *Non*, le programme utilisera seulement le premier jeu de données nommé *Donnees.txt* et affichera les mêmes résultats que la rubrique précédente. *Annuler* revient à l'écran précédent. Enfin, s'il adopte le mode multi données en cliquant sur *Oui*, une petite boîte d'information est éditée afin de prévenir l'utilisateur qu'un seul modèle peut être testé.



Fenêtre 89 - Message de mise en garde pour le test

En cliquant sur **Sélection de modèles**, la Fenêtre 82 s'affiche et l'utilisateur peut alors choisir l'unique loi à tester. S'il en choisit quand même plusieurs, un message d'erreur sera renvoyé à son intention. Par exemple, choisissons de tester nos échantillons de loi Gamma contre la loi de Weibull avec un risque de 5%. Le résultat de cette fonctionnalité sera uniquement graphique. Le logiciel teste le modèle Weibull sur chaque jeu de données et présente le pourcentage de rejet et d'acceptation de l'ensemble, ainsi que l'écart type de ces résultats.



Fenêtre 90 - Test d'Anderson-Darling en mode jeux de données multiples

On peut constater que le test est peu puissant sur cet exemple. En effet, la réponse a été affirmative sur 82 jeux de données (100 jeux avaient été simulés), alors que la vraie loi est Gamma. Ce mode de fonctionnement est un bon moyen pour étudier le niveau et la puissance des tests.

- Test de Cramer-Von Mises

Cette fonctionnalité s'utilise en tout point comme la précédente. Les sorties graphiques et numériques sont de même nature que le test d'Anderson-Darling.

- Test d'exponentialité des excès

Cette fonction s'utilise en amont du *test ET*. Elle sert à s'assurer que la distribution des excès est bien de forme exponentielle. Plus concrètement, après avoir formé la liste des excès, *EXTREMES* applique le test d'Anderson-Darling avec le modèle exponentiel sur celle-ci.

On peut employer ce menu de 2 façons différentes : soit en spécifiant un seul nombre d'excès, soit avec tous les nombres d'excès possibles (cf. Fenêtre 91). Comme les 2 tests centraux et comme tous les autres tests présents dans le logiciel, le test d'exponentialité des excès est différent selon le mode de travail choisi par l'utilisateur (multi données ou jeu de données simple).

En mode jeu de données simple :

La boîte de dialogue suivante apparaît :

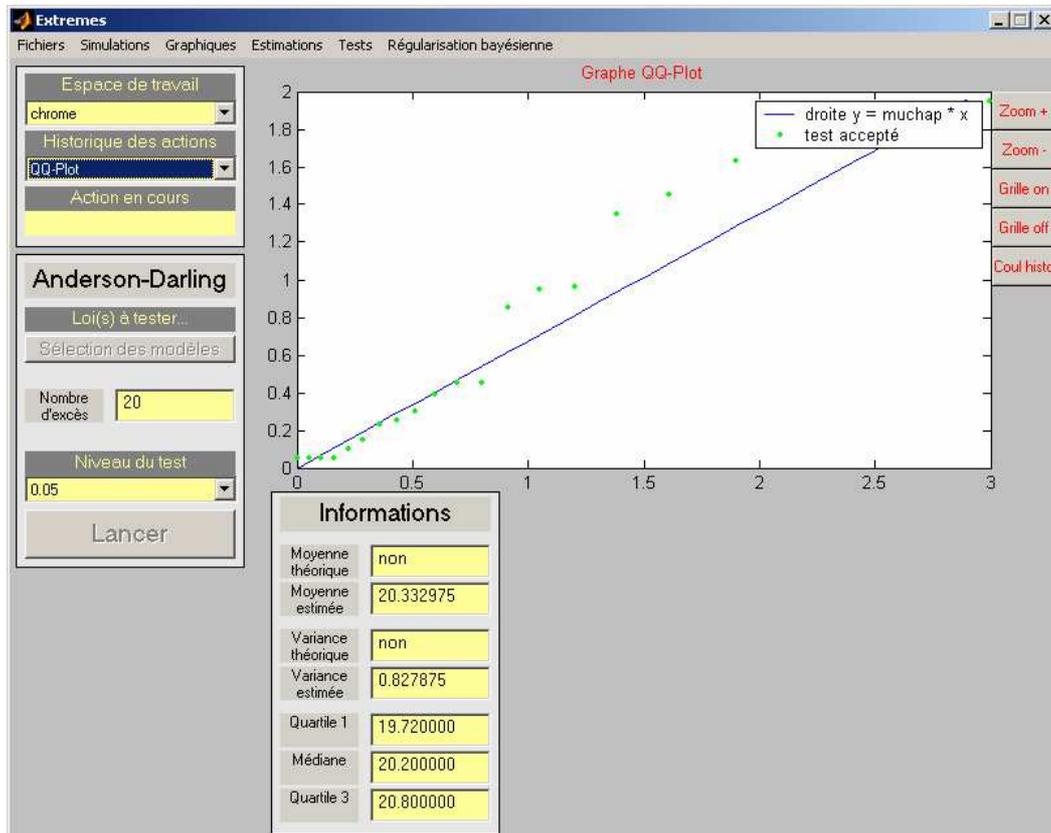


Fenêtre 91 - Choix du nombre d'excès

Si l'utilisateur clique sur *1 seul nombre d'excès*, il verra s'afficher le cadre ci-dessous :

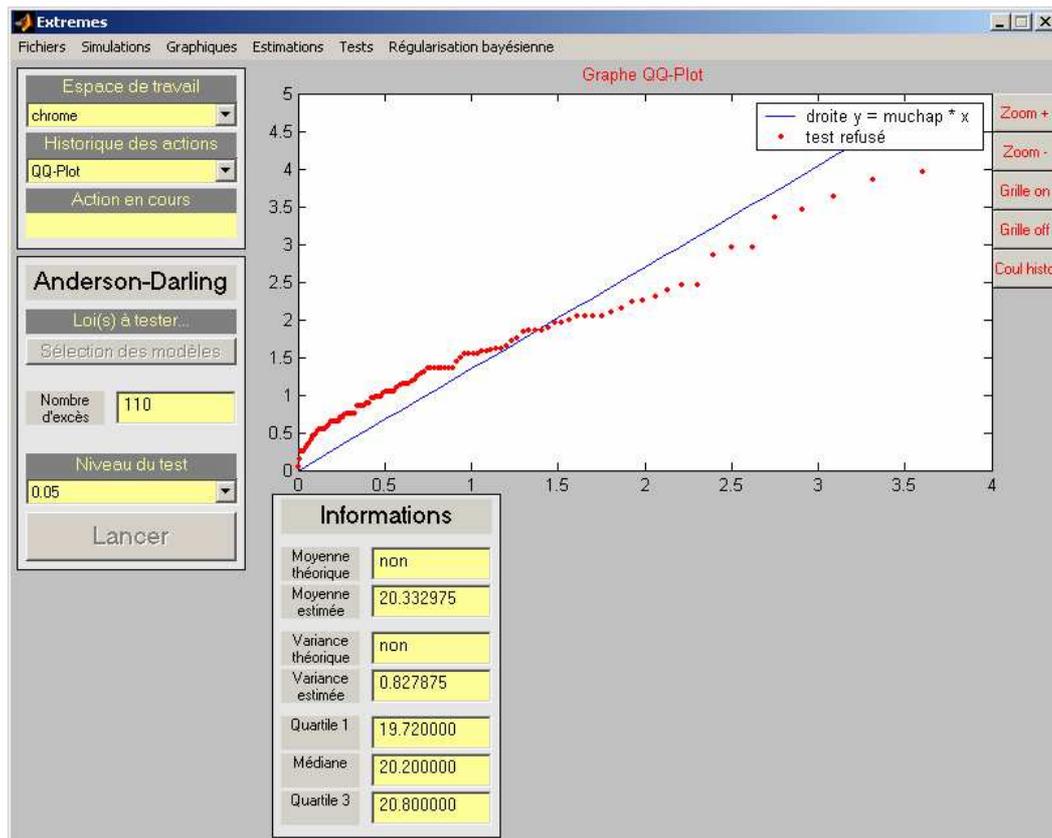
Fenêtre 92 - Cadre de saisie pour le test d'exponentialité à un seul nombre d'excès

Le libellé **Anderson-Darling** souligne que le logiciel utilise ce test d'adéquation centrale pour vérifier l'exponentialité des excès. Le bouton **Sélection de modèles** est grisé car seul le modèle exponentiel est testé. Un nombre d'excès compris entre 2 et la taille de l'échantillon de départ est demandé à l'utilisateur (par exemple 20), ainsi que le niveau du test. En cliquant sur **Lancer**, on obtient le QQ-plot (en abscisse : les quantiles de la loi Exponentielle de paramètre 1. En ordonnée : les quantiles empiriques) ci dessous :



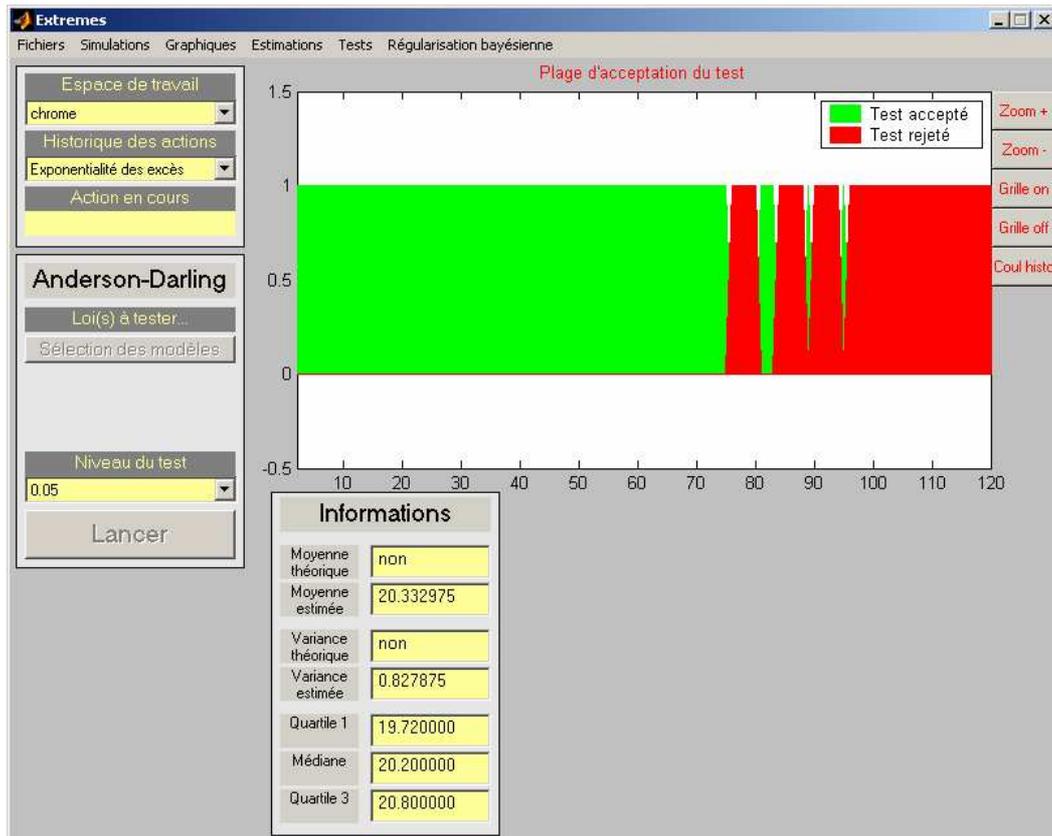
Fenêtre 93 - Résultat du test d'exponentialité avec un seul nombre d'excès

Si les points sont coloriés en vert, alors le test est accepté. S'ils sont en rouge, le test est refusé (cf. Fenêtre 94). Le bon alignement des points avec la droite bleue dont la pente représente le paramètre de la loi Exponentielle donne aussi une indication sur l'exponentialité de la distribution des excès. Il est clair que le résultat de ce test dépend du nombre d'excès saisi par l'utilisateur. Si par exemple on choisit un nombre d'excès égal à 110 (sur cet exemple), on s'aperçoit que la distribution composée par les 110 excès ne correspond pas à une loi exponentielle (cf. Fenêtre 94). Si les excès ne sont pas de forme exponentielle, on ne pourra pas appliquer le **test ET** par la suite.



Fenêtre 94 - Refus de la loi exponentielle pour la distribution des excès

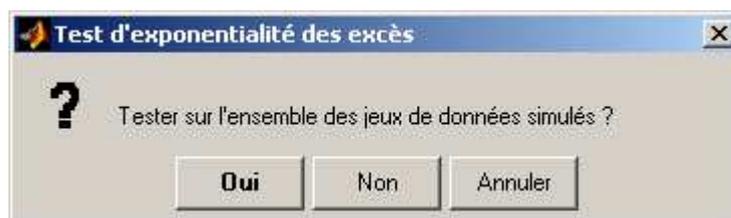
Revenons à la Fenêtre 91. L'utilisateur choisit maintenant *Plusieurs nombres d'excès*. Pour chaque valeur du nombre d'excès allant de 2 à la taille totale de l'échantillon, le programme va effectuer le test d'exponentialité. L'utilisateur aura alors un aperçu global des résultats du test suivant le nombre d'excès. Il pourra alors choisir d'une manière plus confortable un nombre d'excès lui permettant d'appliquer le *test ET* (cf. Fenêtre 95). La zone colorisée en vert correspond à la plage des nombres d'excès pour lesquels le test a été accepté. En rouge, le test est refusé. Sur cet exemple, l'utilisateur peut choisir à peu près n'importe quel nombre d'excès entre 10 et 70 (il est conseillé de ne pas choisir un nombre d'excès trop petit dans les procédures de test ou d'estimation).



Fenêtre 95 - Plage des nombres d'excès acceptant le test d'exponentialité

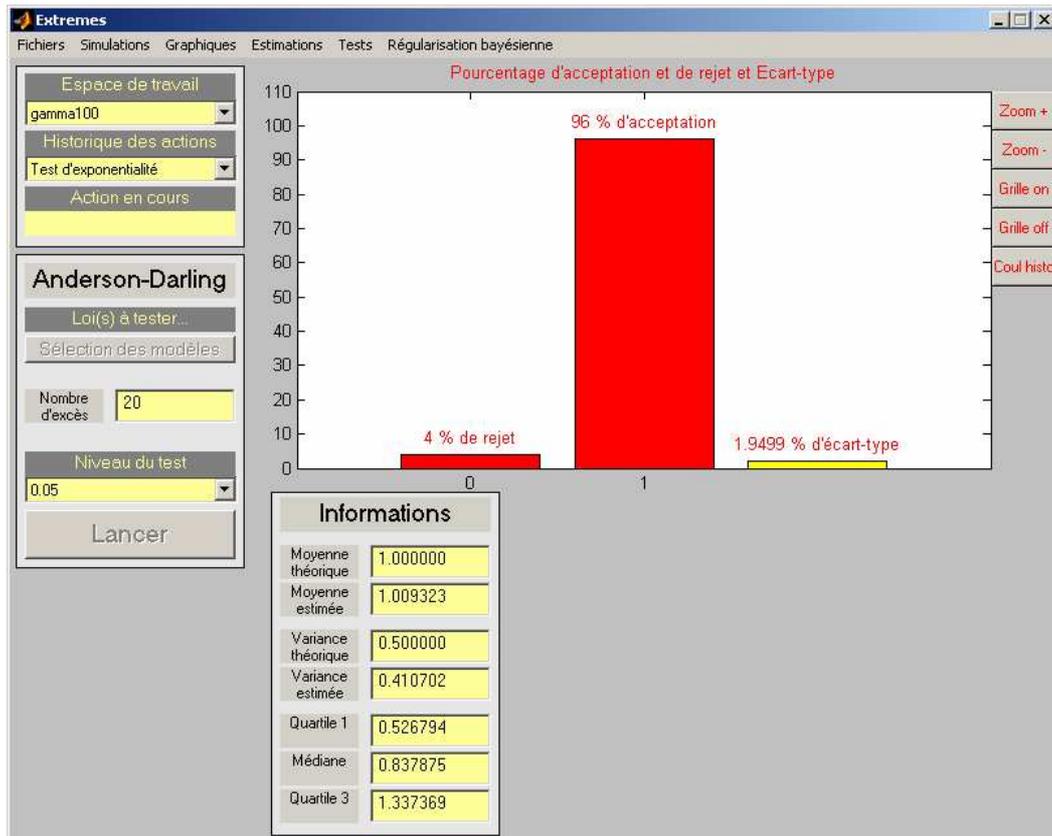
En mode jeux de données multiples :

Retournons à l'espace de travail *gamma100* où 100 échantillons de loi Gamma ont été simulés. En cliquant sur *Test d'exponentialité des excès*, on trouve la fenêtre suivante :



Fenêtre 96 - Choix du mode de travail

En mode multi données (réponse *Oui*), le programme forme la liste des excès pour chaque jeu de données avec le nombre d'excès fourni par l'utilisateur (cf. Fenêtre 92). Il applique pour chacune d'elle le test d'exponentialité et affiche comme pour les tests centraux le pourcentage d'acceptation et de rejet du test, ainsi que l'écart type de ce résultat (cf. Fenêtre 97).



Fenêtre 97 - Test d'exponentialité des excès en mode multi données

- Test ET

Le **test ET** (Exponential Tail) a été développé et étudié dans la thèse de Myriam Garrido¹ (chapitre 1 de la thèse) et est un test d'adéquation pour la queue de distribution. Le but est de sélectionner un modèle qui soit accepté en partie centrale et en queue de distribution. Pour cela, le cheminement conseillé est d'appliquer un test central sur les données, d'effectuer le test d'exponentialité des excès puis, si ce dernier est accepté, lancer un **test ET**. Le test est basé sur une comparaison de \hat{q}_{ET} (estimateur **ET** d'un quantile extrême, consistant à choisir $\gamma = 0$ dans \hat{q}_{GPD} p.64) et \hat{q}_{param} (estimateur paramétrique). Il y a 3 versions disponibles de ce test dans **EXTREMES** : asymptotique (c'est-à-dire basée sur la loi asymptotique de \hat{q}_{ET}) (version la moins puissante), bootstrap paramétrique complet (version la plus puissante basée sur la distribution empirique de $\hat{q}_{ET} - \hat{q}_{param}$) et bootstrap paramétrique simplifié (basée sur la distribution empirique de \hat{q}_{ET}). Comme tous les autres tests, les résultats sont différents selon le mode opératoire choisi par l'utilisateur (multi données ou non).

¹ Garrido M., Modélisation des événements rares et estimation des quantiles extrêmes, Méthodes de sélection de modèles pour les queues de distribution, Thèse de doctorat, Université Grenoble 1, 2002.

En mode jeu de données simple :

Poursuivons l'exemple de données réelles (espace de travail *chrome*) où les lois Lognormale et Gamma avaient été acceptées en partie centrale (cf. Fenêtre 84), et où le nombre d'excès pouvait être choisi entre 10 et 70 (cf. Fenêtre 95). Le logiciel propose la fenêtre suivante :



Fenêtre 98 - Choix du nombre d'excès

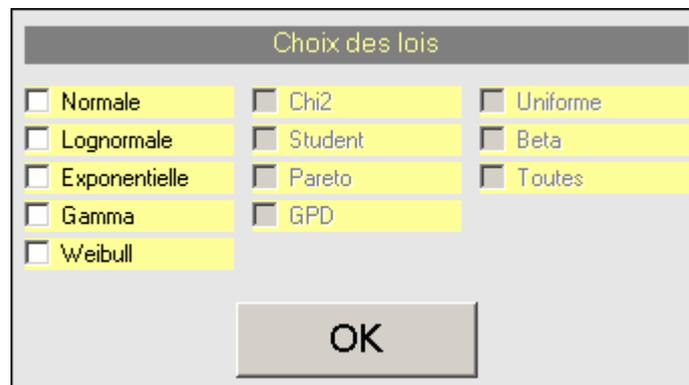
Selon le choix de l'utilisateur, les 2 cadres ci-dessous peuvent apparaître :

Test ET	
Loi(s) à tester...	
Sélection de modèles	
Version du test	
Asymptotique	
Niveau du test	
0.25	
Nb d'excès	
Quantile d'ordre 1-p	
Valeur de p	
Lancer	

Test ET	
Loi(s) à tester...	
Sélection de modèles	
Version du test	
Asymptotique	
Niveau du test	
0.25	
Borne Inf des excès	
Borne Sup des excès	
Quantile d'ordre 1-p	
Valeur de p	
Lancer	

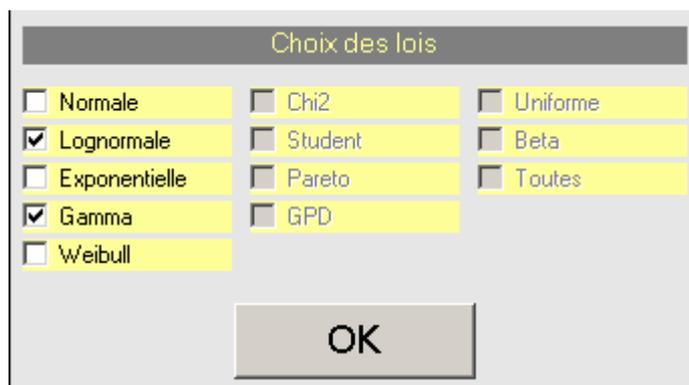
Fenêtre 99 - Cadre de saisie pour le test ET

Le bouton **Sélection de modèle** fait apparaître la boîte des lois où ne sont disponibles que celles appartenant au domaine d'attraction de Gumbel :



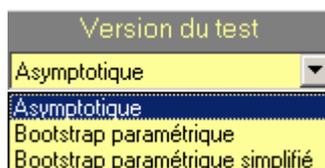
Fenêtre 100 - Choix des modèles appartenant au domaine d'attraction de Gumbel

On coche donc les modèles acceptés par le test d'Anderson-Darling (Lognormale et Gamma) puis **OK** :



Fenêtre 101 - Choix des modèles acceptés en région centrale

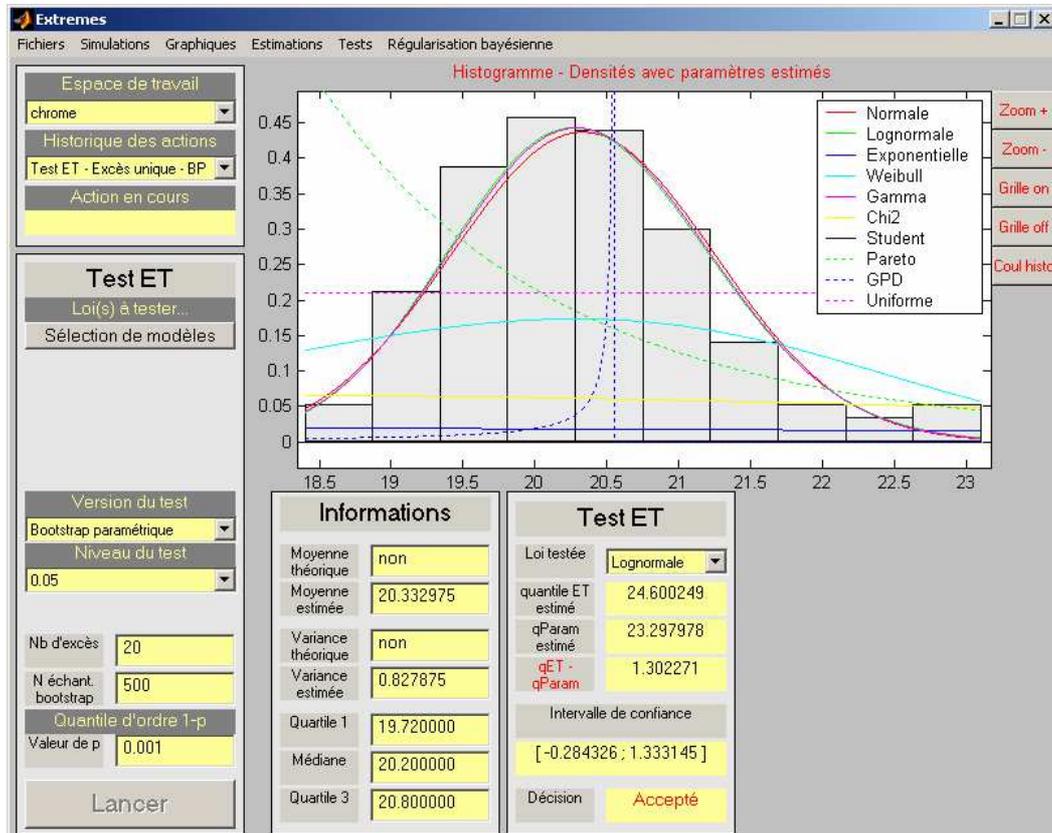
Les différentes versions du **test ET** apparaissent dans la liste déroulante **Version du test** :



Fenêtre 102 - Versions du test ET

On choisira **bootstrap paramétrique** sur cet exemple. Prenons un seul nombre d'excès égal à 20. Le champ **N échant. bootstrap** demande un nombre d'échantillons de simulation de

bootstrap pour le test. Des valeurs autour de 500 sont conseillées. La valeur de p du quantile d'ordre $1 - p$ doit également être renseignée. On veut estimer un quantile extrême situé au-delà de l'observation maximale, donc on doit avoir $p \leq 1/n$, n désignant la taille de l'échantillon. Choisissons par exemple $p = 0.001$. Un clic sur **Lancer** nous donne :



Fenêtre 103 - Résultats du test ET avec un seul nombre d'excès

Les résultats sont concentrés dans la boîte intitulée **Test ET** en bas du graphique (il n'y a pas de sorties graphiques pour cette fonctionnalité). Les lois testées apparaissent dans la liste déroulante prévue à cet effet. Pour cette version du test, la statistique de test (qui est toujours colorisée en rouge) est $\hat{q}_{ET} - \hat{q}_{param}$. Si cette valeur appartient à l'intervalle de confiance, alors le test est accepté. Sinon, on rejette le modèle proposé. La statistique de test n'est pas la même selon la version du test choisie. Pour visualiser les résultats concernant la loi Gamma, il faut dérouler la liste et cliquer sur Gamma :

Test ET	
Loi testée	Gamma
quantile ET estimé	24.600249
qParam estimé	23.234127
qET - qParam	1.366122
Intervalle de confiance	
[-0.233667 ; 1.368709]	
Décision	Accepté

Fenêtre 104 - Résultat du test ET pour la loi Gamma

On peut remarquer que pour cette valeur du nombre d'excès, le *test ET* version *bootstrap paramétrique* ne peut départager le modèle lognormal du modèle gamma. Il convient d'utiliser les conseils prodigués dans la thèse de Myriam Garrido¹ (chapitre 1.3.3, p.44 de la thèse) pour décider correctement selon le nombre d'excès.

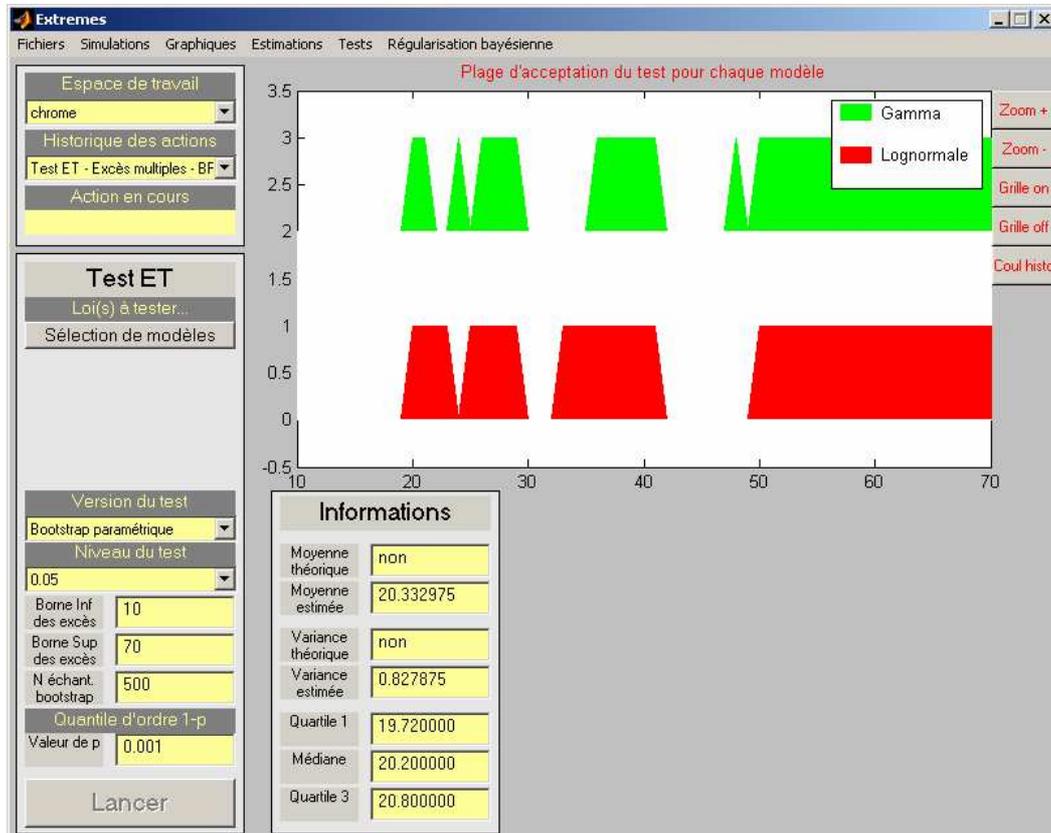
Si l'utilisateur choisit plusieurs nombres d'excès, on pourrait remplir le cadre de saisie comme suit après avoir sélectionné les modèles lognormal et gamma :

Test ET	
Loi(s) à tester...	
Sélection de modèles	
Version du test	
Bootstrap paramétrique	
Niveau du test	
0.05	
Borne Inf des excès	10
Borne Sup des excès	70
N échant. bootstrap	500
Quantile d'ordre 1-p	
Valeur de p	0.01
Lancer	

Fenêtre 105 - Cadre du test ET pour plusieurs nombres d'excès

¹ Garrido M., Modélisation des événements rares et estimation des quantiles extrêmes, Méthodes de sélection de modèles pour les queues de distribution, Thèse de doctorat, Université Grenoble 1, 2002.

On peut ensuite cliquer sur **Lancer**. Cette opération peut prendre plusieurs minutes. En effet, la taille de l'échantillon de départ est d'environ 120. On en simule 500 pour chaque modèle (dont il faut estimer les paramètres) et pour chaque nombre d'excès ($70 - 10 = 61$ nombres d'excès). Cela conduit à près de 3.6 millions de simples itérations. On obtient les sorties graphiques suivantes :



Fenêtre 106 - Test ET avec plusieurs nombres d'excès

Les nombres d'excès sont représentés en abscisse du graphique (de 10 à 70). Les zones colorées indiquent les régions d'acceptation du test (par exemple de 50 à 70 excès) et les zones blanches les régions de rejet (par exemple de 42 à 49 excès). Encore une fois, la décision finale d'accepter un modèle plutôt qu'un autre est délicate. Ces situations sont étudiées dans la thèse de Myriam Garrido¹ (chapitre 1.2 de la thèse).

En mode jeux de données multiples :

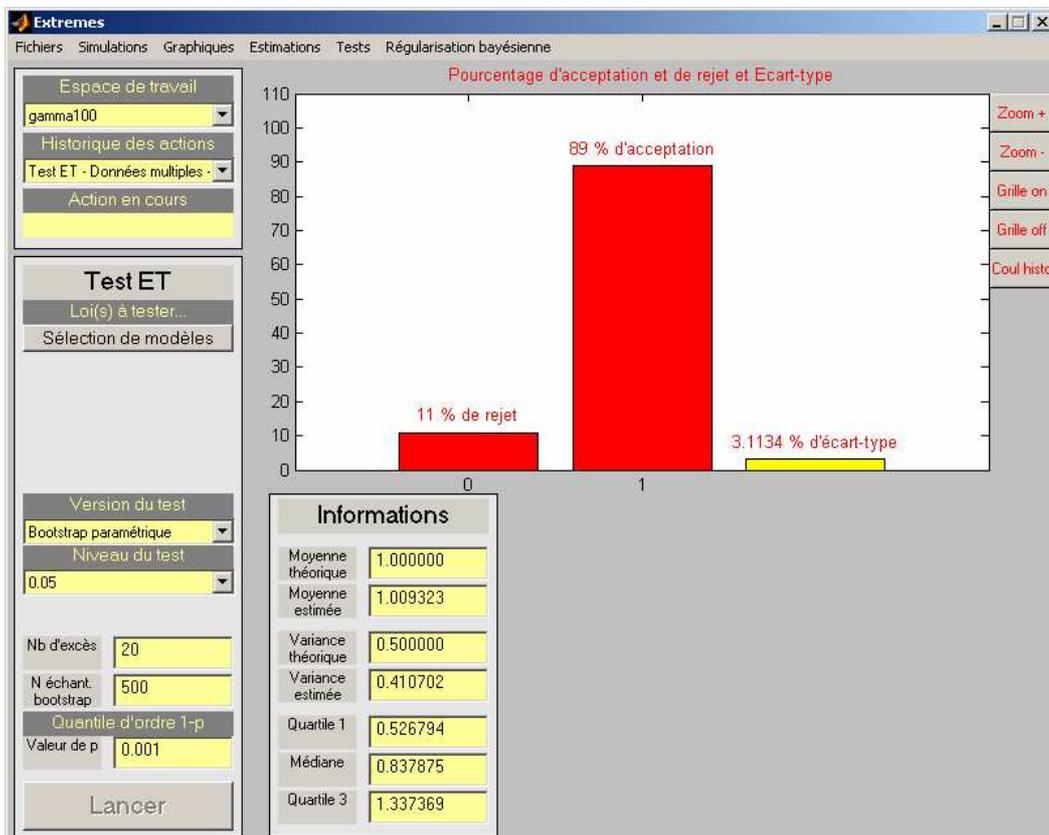
Le **test ET** admet exactement la même façon de procéder que les autres tests en mode multi données. Replaçons-nous dans l'espace de travail **gamma100**. Tout d'abord, le logiciel demande à l'utilisateur le mode d'utilisation de la fonction :

¹ Garrido M., Modélisation des événements rares et estimation des quantiles extrêmes, Méthodes de sélection de modèles pour les queues de distribution, Thèse de doctorat, Université Grenoble 1, 2002.



Fenêtre 107 - Choix du mode de travail

En cliquant sur **Oui**, l'utilisateur choisit le mode multi données. A partir de ce point, il ne sera possible de tester qu'un seul modèle comme le rappelle **EXTREMES**. On retrouve alors le même cadre que la Fenêtre 99. Le **test ET** n'est disponible qu'avec un seul nombre d'excès. Testons par exemple la loi de Weibull à 5% avec 500 échantillons de bootstrap et $p = 0.001$ pour la quantile d'ordre $1 - p$. Les sorties sont semblables à celles rencontrées lors des tests centraux (cf. Fenêtre 97).



Fenêtre 108 - Test ET en mode multi données

- Test GPD

Le **test GPD** est un test d'adéquation pour la queue de distribution. Il est incorporé dans le logiciel **EXTREMES** de manière expérimentale. En effet, aucune étude n'a été pratiquée sur ce test au préalable. Aucun conseil ne sera fourni à l'utilisateur pour choisir les

différentes valeurs à entrer. Le **test GPD** ne se limite pas à un domaine d'attraction mais est utilisable avec toutes les lois du catalogue. Son fonctionnement est différent selon le mode de travail choisi par l'utilisateur. Le test est basé sur une comparaison entre \hat{q}_{GPD} et \hat{q}_{param} . Pour effectuer ce test, on a besoin d'estimer les paramètres de la loi GPD que suit en théorie la distribution des excès. Pour cela, on applique les méthodes décrites p. 33 de ce manuel.

En mode jeu de données simple :

Reprenons l'exemple avec l'espace de travail **chrome** (cf. p. 49). Comme le **test ET**, on ne retient que les modèles ayant été acceptés en partie centrale (ici lognormal et gamma). Le programme présente la boîte suivante :



Fenêtre 109 - Choix du nombre d'excès

Ce qui a pour conséquence l'affichage de l'un des 2 cadres ci-dessous :

Test GPD	
Loi(s) à tester...	
Sélection de modèles	
Méthode d'estimation	
Sélection de la méthode	
Version du test	
Asymptotique	
Niveau du test	
0.05	
Nb d'excès	
Quantile d'ordre 1-p	
Valeur de p	
Lancer	

Test GPD	
Loi(s) à tester...	
Sélection de modèles	
Méthode d'estimation	
Sélection de la méthode	
Version du test	
Asymptotique	
Niveau du test	
0.05	
Borne Inf des excès	
Borne Sup des excès	
Quantile d'ordre 1-p	
Valeur de p	
Lancer	

Fenêtre 110 - Cadre de présentation du test GPD

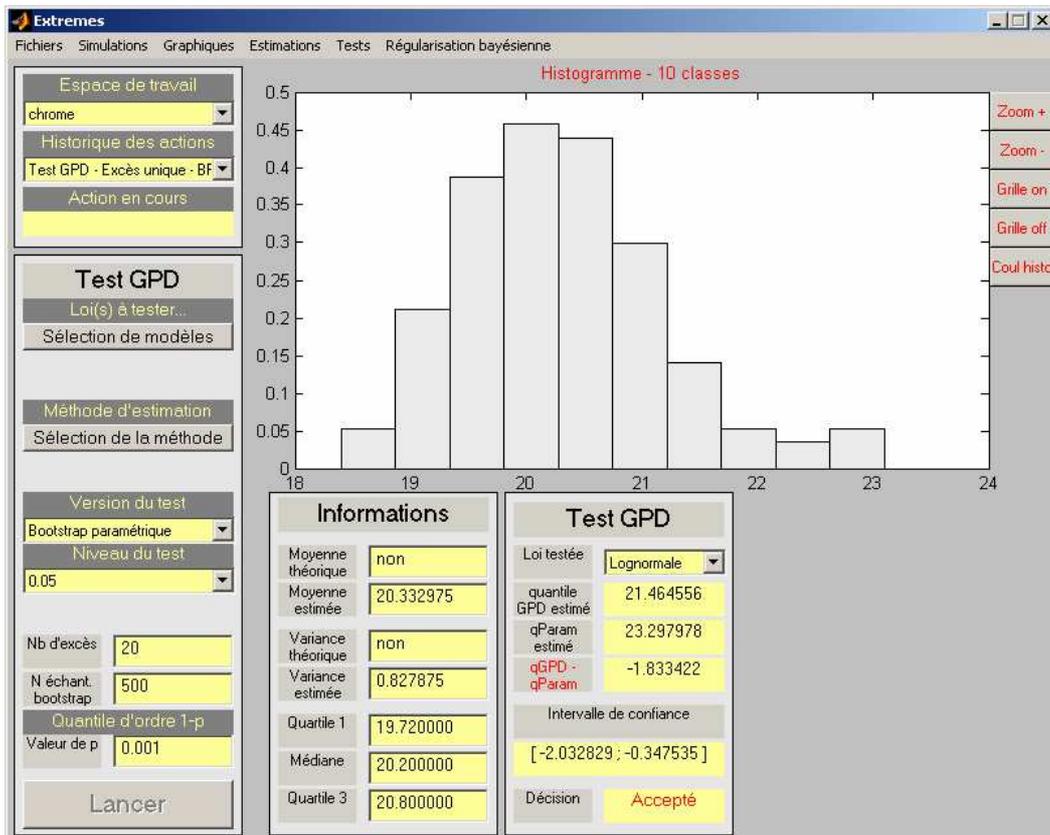
Si *1 seul nombre d'excès* est choisi, une boîte résultat identique à celle du *test ET* sera présentée. En cliquant sur *Sélection de modèles*, la Fenêtre 82 apparaît à l'écran et l'utilisateur n'a plus qu'à cocher les modèles qui conviennent. En cliquant sur *Sélection de la méthode*, on voit s'afficher le cadre suivant :

Fenêtre 111 - Choix des méthodes d'estimation des paramètres de la loi GPD

L'utilisateur ne peut employer qu'une seule méthode d'estimation. Pour l'exemple, choisissons *Moments pondérés*. Il faut ensuite déterminer une version du test : bootstrap complet ou bootstrap paramétrique simplifié (la version asymptotique n'est pas disponible pour le *test GPD*). Prenons 20 comme nombre d'excès, 500 échantillons de bootstrap et 0.001 comme valeur de p pour le quantile d'ordre $1-p$ (p doit être inférieur ou égal à $1/n$ où n est la taille de l'échantillon étudié) :

Fenêtre 112 - Boîte de saisie du test GPD

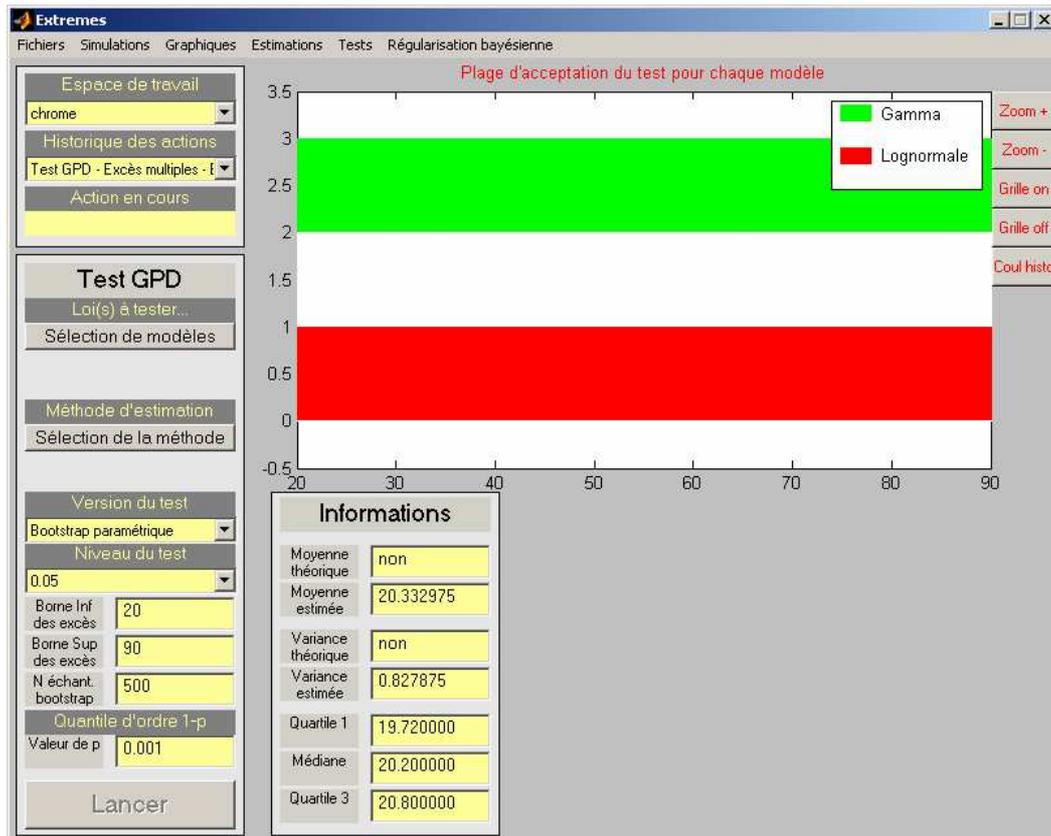
En cliquant sur **Lancer**, on obtient :



Fenêtre 113 - Résultat du test GPD pour un seul nombre d'excès

La statistique de test est indiquée en rouge. Si elle se situe à l'intérieur de l'intervalle de confiance, alors le test est accepté, sinon il est refusé. Pour voir les résultats du test concernant la loi Gamma, il suffit de dérouler la liste intitulée **Loi testée** (cf. p 62-63).

Si l'utilisateur choisit **Plusieurs nombres d'excès**, il doit fournir un intervalle comme présenté par la Fenêtre 110 (cadre de droite). Choisissons par exemple $[20,90]$ comme intervalle pour les nombres d'excès et gardons les mêmes valeurs que précédemment pour les autres champs. Les sorties seront uniquement graphiques et permettront à l'utilisateur d'avoir en un seul coup d'œil un aperçu du résultat du **test GPD** en fonction du nombre d'excès (cf. Fenêtre 114).



Fenêtre 114 - Résultats du test GPD avec plusieurs nombres d'excès

Les nombres d'excès sont représentés en abscisse. Quand la zone est coloriée, le test est accepté. Quand elle n'y est pas, le test est refusé. On peut s'apercevoir ici que quel que soit le nombre d'excès, la réponse au test est affirmative. Sur cet exemple précis et avec cette méthode d'estimation (*Moments pondérés*) pour les paramètres de la loi GPD, il n'y a aucune valeur du nombre d'excès pour laquelle le test soit refusé.

Remarque 7 : Pour le *test GPD* et ce quel que soit le mode de travail choisi par l'utilisateur, une seule méthode d'estimation des paramètres de la loi GPD peut être cochée. Il n'est pas permis d'en sélectionner plusieurs en même temps. Un message d'erreur informera l'utilisateur en cas de non respect de ce critère.

Remarque 8 : Comme pour le *test ET*, cette procédure peut prendre quelques minutes (suivant le choix de l'intervalle des nombres d'excès et du nombre d'échantillons de bootstrap par exemple).

En mode jeux de données multiples :

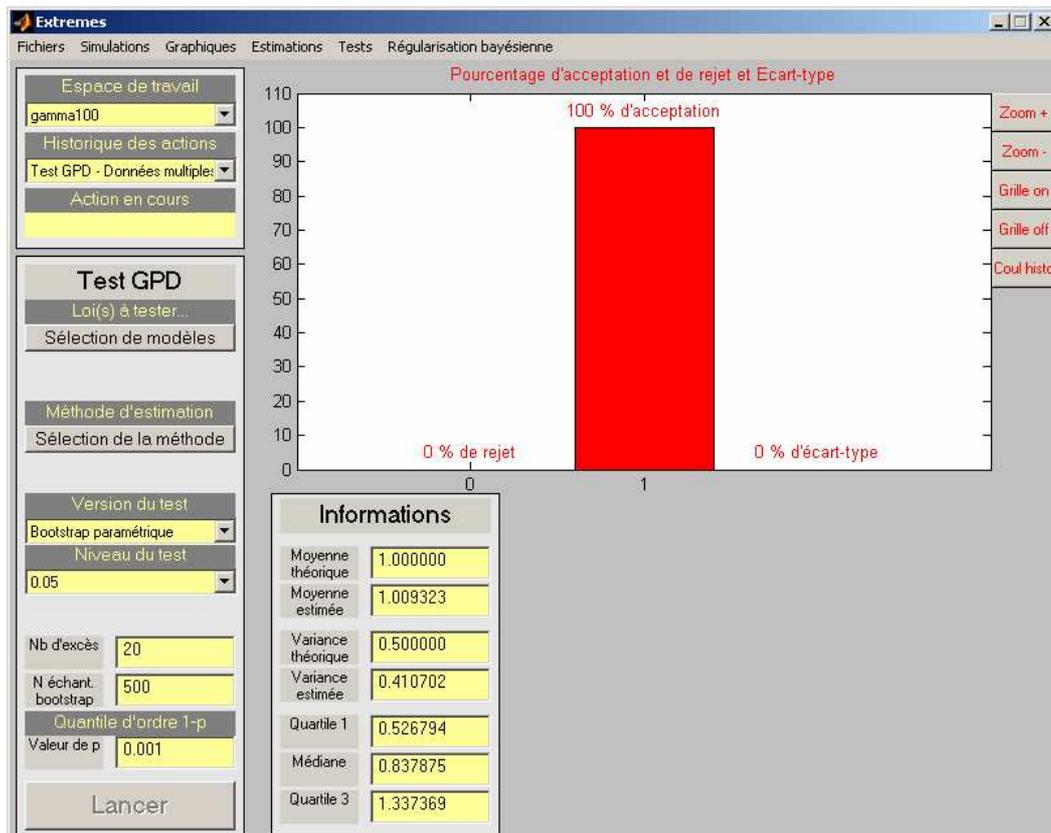
La méthode est identique à tous les tests du logiciel dans le cas de jeux de données multiples. Le programme applique le test à chaque jeu de données et présente le pourcentage de rejet et d'acceptation, ainsi que l'écart type. Continuons à travailler dans l'espace de travail

gamma100 (100 échantillons de taille 100 issus d'une loi Gamma). Une première fenêtre est proposée à l'utilisateur :



Fenêtre 115 - Choix du mode de travail

En sélectionnant le mode jeux de données multiples (un clic sur **Oui**), l'utilisateur se voit prévenir que le test ne peut s'appliquer que sur un seul modèle à la fois. Il convient ensuite de remplir convenablement les champs demandés. Sélectionnons le modèle Weibull, la méthode d'estimation des **Moments pondérés**, la version bootstrap paramétrique complète du **test GPD**, un risque de 5%, 20 excès, 500 échantillons de bootstrap et un quantile d'ordre 1-0.001. On obtient :



Fenêtre 116 - Résultats du test GPD en mode multi données

Le **test GPD** a accepté dans 100% des cas le modèle Weibull. Rappelons une nouvelle fois le caractère expérimental de ce test.

3.2.6. Le menu Régularisation bayésienne

Ce menu ne contient qu'une fonctionnalité. Cette fonction s'utilise après avoir appliqué au jeu de données dont on dispose un test central et un *test ET*. Si aucune loi n'est acceptée par les 2 types de tests, la procédure de régularisation bayésienne permet, à partir d'un modèle adapté aux valeurs les plus probables, d'améliorer l'adéquation extrême grâce à un avis d'expert sur la queue de distribution. Elle peut être aussi utilisée lorsque la distribution a été acceptée à la fois par un test usuel et un test pour la queue de distribution afin de construire un meilleur modèle. Si plusieurs lois sont acceptées en partie centrale et en queue de distribution, la *régularisation bayésienne* fournit des indications pour sélectionner le meilleur modèle. Elle a été développée par Myriam Garrido¹ (chapitre 2 de la thèse) et il est nécessaire de se référer à son travail pour remplir correctement les informations demandées par le logiciel. Cette fonctionnalité ne peut être appliquée que sur un seul jeu de données, le mode multi données n'étant pas disponible.

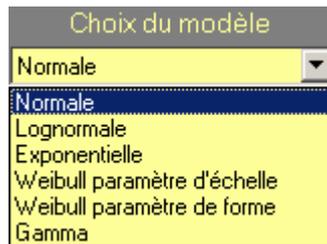
En cliquant sur *Régularisation bayésienne*, le cadre suivant s'affiche :



Fenêtre 117 - Présentation de la fenêtre de saisie pour la régularisation bayésienne

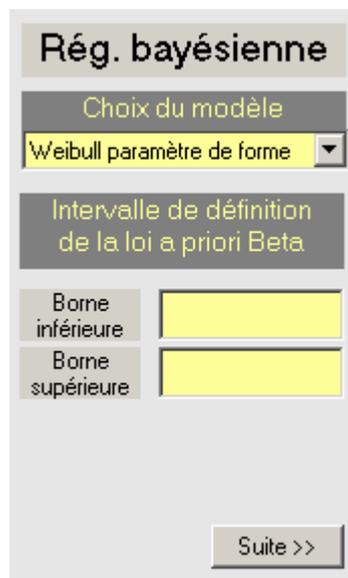
6 modèles sont présents : Normale, Lognormale, Exponentielle, Weibull paramètre d'échelle, Weibull paramètre de forme et Gamma :

¹ Garrido M., Modélisation des événements rares et estimation des quantiles extrêmes, Méthodes de sélection de modèles pour les queues de distribution, Thèse de doctorat, Université Grenoble 1, 2002.



Fenêtre 118 - Choix du modèle pour la régularisation bayésienne

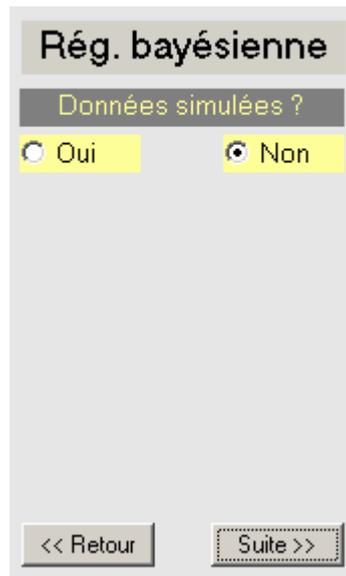
Si on choisit de faire une régularisation bayésienne sur le paramètre de forme de la loi de Weibull, on obtient l'écran suivant :



Fenêtre 119 - Cadre pour le modèle Weibull paramètre de forme

Typiquement, on prendrait $[0,1]$ ou $[1,5]$ comme intervalle de définition pour la loi Beta a priori.

En cliquant sur *Suite*, on obtient l'écran ci-dessous :



Fenêtre 120 - Cadre de régularisation bayésienne

Par défaut, le programme affiche que les données sont réelles. Si elles ont été simulées, un clic sur *Oui* donne :



Fenêtre 121 - Nom de la loi dont sont issues les données simulées

Dans ce cas, la densité du modèle avec les paramètres estimés serait tracée. La valeur des quantiles de la loi de simulation avec les vrais paramètres et de la loi de simulation avec les paramètres estimés serait affichée en sortie. Si par exemple, on clique sur Lognormale, *EXTREMES* demande la valeur des paramètres de la loi de simulation (cf. annexe A) :

The screenshot shows a dialog box titled "Rég. bayésienne". Below the title is a section labeled "Données simulées ?" with two radio buttons: "Oui" (selected) and "Non". Below this is a section labeled "Vraie loi" with a dropdown menu showing "Lognormale". Underneath are two input fields labeled "m" and "sigma". At the bottom are two buttons: "<< Retour" and "Suite >>".

Fenêtre 122 - Valeur des paramètres de la loi de simulation

L'écran suivant s'obtient en cliquant sur *Suite* :

The screenshot shows the same dialog box "Rég. bayésienne". The section "Avis d'expert ?" has two radio buttons: "Oui" and "Non" (selected). Below this is an input field labeled "Degré de défiance". At the bottom are two buttons: "<< Retour" and "Suite >>".

Fenêtre 123 - Avis d'expert

Par défaut, il n'y a pas d'avis d'expert. Si on en possède un, il faut cocher *Oui* et d'autres champs apparaissent :

The screenshot shows a dialog box titled "Rég. bayésienne" with a sub-section "Avis d'expert ?". It contains two radio buttons: "Oui" (selected) and "Non". Below are four input fields: "Quantile maximum", "Borne inf p1 du risque", "Borne sup p2 du risque", and "Degré de défiance". At the bottom are two buttons: "<< Retour" and "Suite >>".

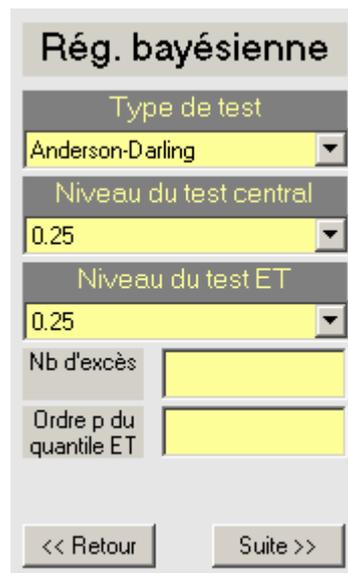
Fenêtre 124 - Avis d'expert (2)

L'utilisateur doit fournir un quantile extrême donné par l'expert, ainsi qu'un encadrement du risque associé à ce quantile. Le degré de défiance quantifie en quelque sorte le doute en l'expert (des valeurs de l'ordre de 0.01 ou 0.05 sont conseillées). Si au préalable on avait choisi le modèle Weibull paramètre de forme, on aurait obtenu le cadre suivant :

This screenshot is similar to the previous one but includes a dropdown menu at the bottom. The dropdown is titled "Confiance en l'information en queue de distribution" and currently shows "Faible". The "Suite >>" button is highlighted.

Fenêtre 125 - Avis d'expert et modèle Weibull paramètre de forme

L'utilisateur a le choix entre 3 propositions pour indiquer la confiance en l'information en queue de distribution : Faible, Moyenne ou Forte. Si *Suite* est cliqué, l'écran ci-dessous est affiché :

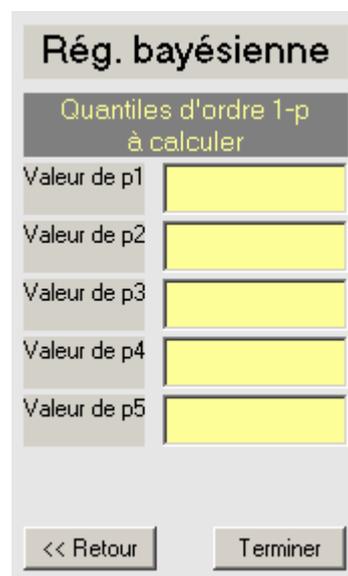


The screenshot shows a dialog box titled "Rég. bayésienne". It contains several configuration options:

- Type de test:** A dropdown menu with "Anderson-Darling" selected.
- Niveau du test central:** A dropdown menu with "0.25" selected.
- Niveau du test ET:** A dropdown menu with "0.25" selected.
- Nb d'excès:** An empty text input field.
- Ordre p du quantile ET:** An empty text input field.
- At the bottom, there are two buttons: "<< Retour" and "Suite >>".

Fenêtre 126 - Saisie des entrées pour les tests

Le programme va appliquer un test central et le *test ET* version bootstrap paramétrique simplifiée au jeu de données. Les niveaux de ces 2 tests doivent être désignés. Un nombre d'excès et l'ordre du quantile ET doivent être renseignés. Ne reste à saisir que les ordres des quantiles à estimer en cliquant sur *Suite* :



The screenshot shows a dialog box titled "Rég. bayésienne". It contains several input fields for quantile orders:

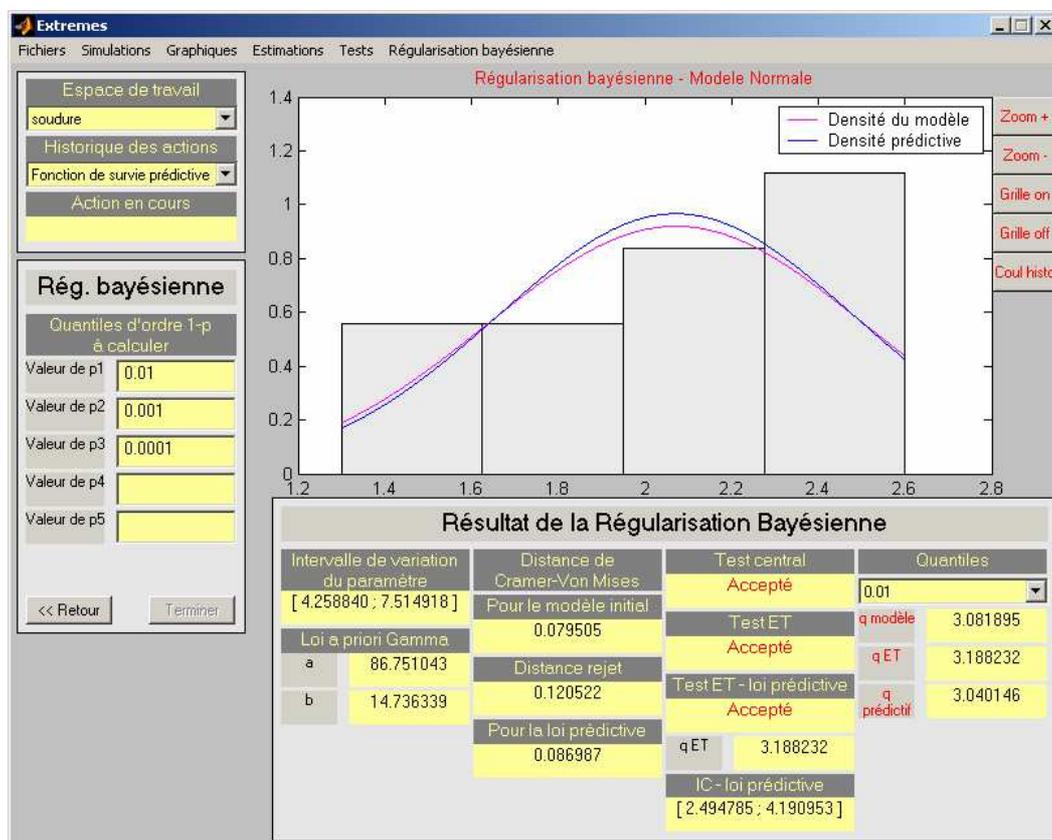
- Quantiles d'ordre 1-p à calculer:** A header for the following fields.
- Valeur de p1:** An empty text input field.
- Valeur de p2:** An empty text input field.
- Valeur de p3:** An empty text input field.
- Valeur de p4:** An empty text input field.
- Valeur de p5:** An empty text input field.
- At the bottom, there are two buttons: "<< Retour" and "Terminer".

Fenêtre 127 - Ordres des quantiles à calculer

Un clic sur *Terminer* et la procédure de calcul est lancée. A chaque étape de la saisie, il est possible de revenir sur les écrans précédents en appuyant sur *Retour*.

Exemple : Ouvrons un jeu de données réelles de taille 11 (hauteur de défauts de soudure) (*defsoud.txt* est présent dans le sous-répertoire *Donnees* du répertoire d'installation). Les tests

centraux acceptent par exemple la loi Normale mais aussi beaucoup d'autres lois vu la faible taille de l'échantillon. Choisissons le modèle Normale, les données n'étant pas simulées, et prenons comme avis d'expert 3.2mm que l'on encadre en terme de risque par $p1 = 0.01$ et $p2 = 0.001$. Prenons le test de Cramer-Von Mises avec risque 5% et 4 excès. Calculons les quantiles d'ordre 1-0.01, 1-0.001 et 1-0.0001. Le programme calcul et affiche une boîte résultat dans laquelle sont affichés l'intervalle de variation du paramètre, les paramètres de la loi a priori (toujours une loi gamma sauf quand le modèle est Weibull paramètre de forme, auquel cas il s'agit d'une loi beta). Diverses distances de Cramer-Von Mises seront également éditées, ce qui donnera une indication à l'utilisateur sur la validité du modèle initial et prédictif. Enfin, la décision des tests et la valeur des quantiles apparaîtront. Un nombre de classes sera demandé à l'utilisateur pour construire l'histogramme des données sur lequel seront superposées la densité du modèle et la densité prédictive (cf. Fenêtre 128). Cet exemple fait partie de la thèse de Myriam Garrido¹ (chapitre 2.1.4, p.76).



Fenêtre 128 - Résultat de la régularisation bayésienne

L'intervalle de variation du paramètre sur lequel on a mis une loi a priori est déterminé grâce à l'avis d'expert. On peut ensuite calculer les hyperparamètres (paramètres de la loi a priori) à partir de cet intervalle de variation et de la confiance en l'expert. La valeur **Distance rejet** est tabulée pour le modèle initial. Or on aimerait tester l'adéquation générale de la loi prédictive mais ces lois ne sont pas usuelles et par conséquent, il n'existe pas de valeur de rejet tabulée pour elles. Pour avoir une indication sur la bonne adéquation de la loi prédictive,

¹ Garrido M., Modélisation des événements rares et estimation des quantiles extrêmes, Méthodes de sélection de modèles pour les queues de distribution, Thèse de doctorat, Université Grenoble 1, 2002.

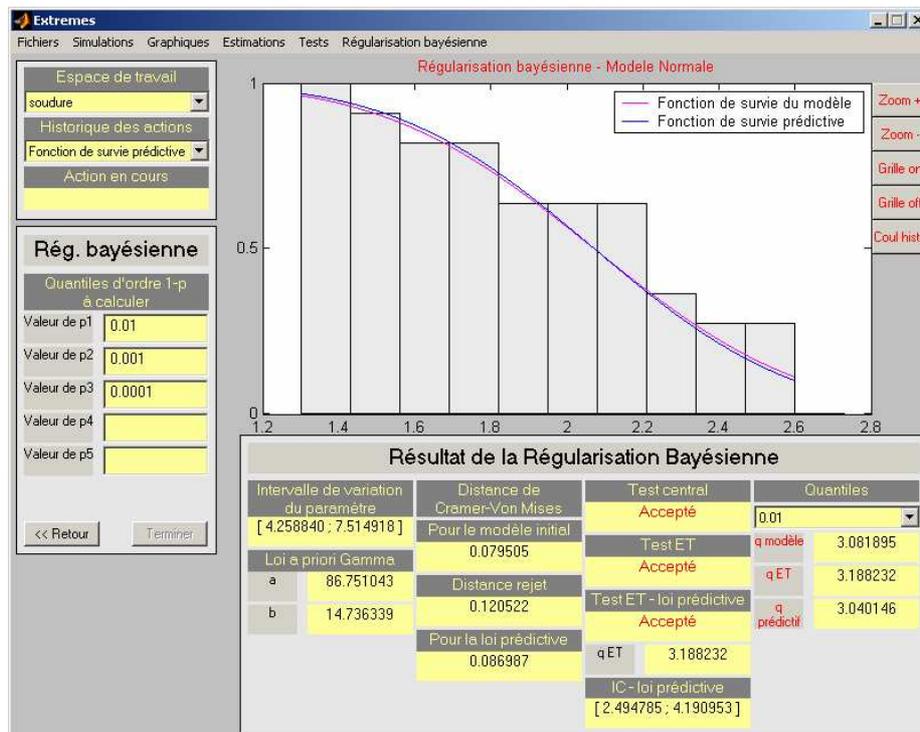
on peut comparer sa distance de Cramer-Von Mises (distance CVM) avec celle du modèle de départ. On peut un peu plus « quantifier » ceci en comparant la distance CVM du modèle à sa valeur de rejet : plus la distance CVM est petite par rapport à la valeur de rejet, meilleur était le modèle de départ, et plus on va s'autoriser des valeurs de la distance CVM relativement lointaines pour la loi prédictive. Par contre, si la distance CVM du modèle est proche de sa valeur de rejet, le modèle est peu adéquat et on va se méfier dès que la distance CVM de la prédictive augmente un peu.

Un clic droit sur le graphique offre le menu contextuel suivant :



Fenêtre 129 - Menu contextuel du graphique

La mention *Fonction de survie – RB* permet de visualiser la fonction de survie empirique sur laquelle les fonctions de survie du modèle et prédictive seront dessinées.



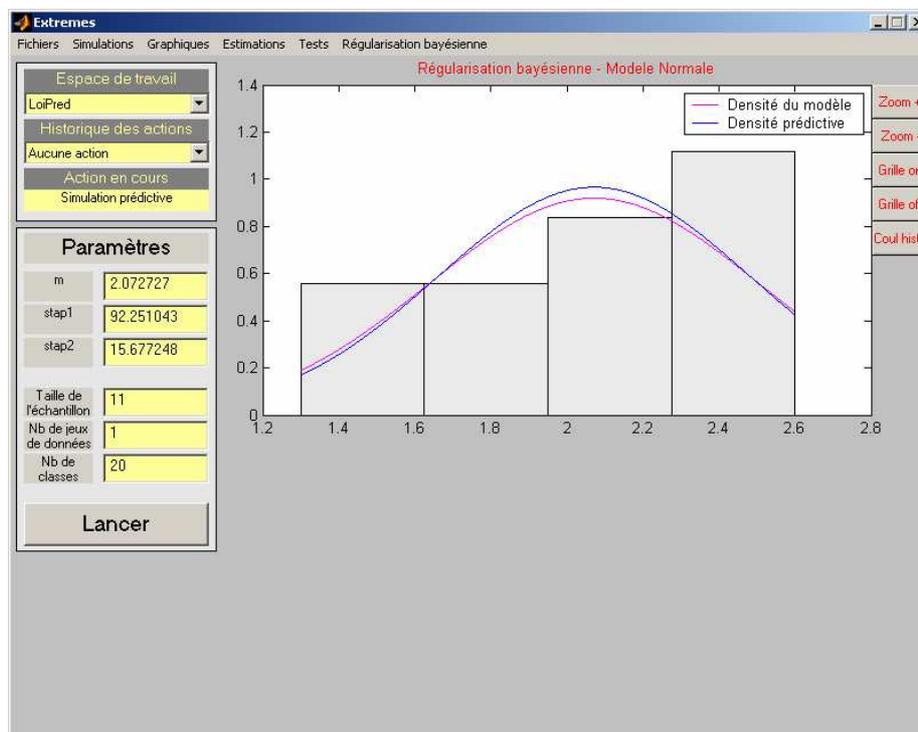
Fenêtre 130 - Résultats de la régularisation bayésienne (2)

Pour voir les valeurs des quantiles calculés, il faut dérouler la liste **Quantiles** dans la boîte **Résultat de la Régularisation bayésienne**.

Remarque 9 : Le temps de calcul de cette procédure peut prendre de quelques secondes à plusieurs dizaines de minutes.

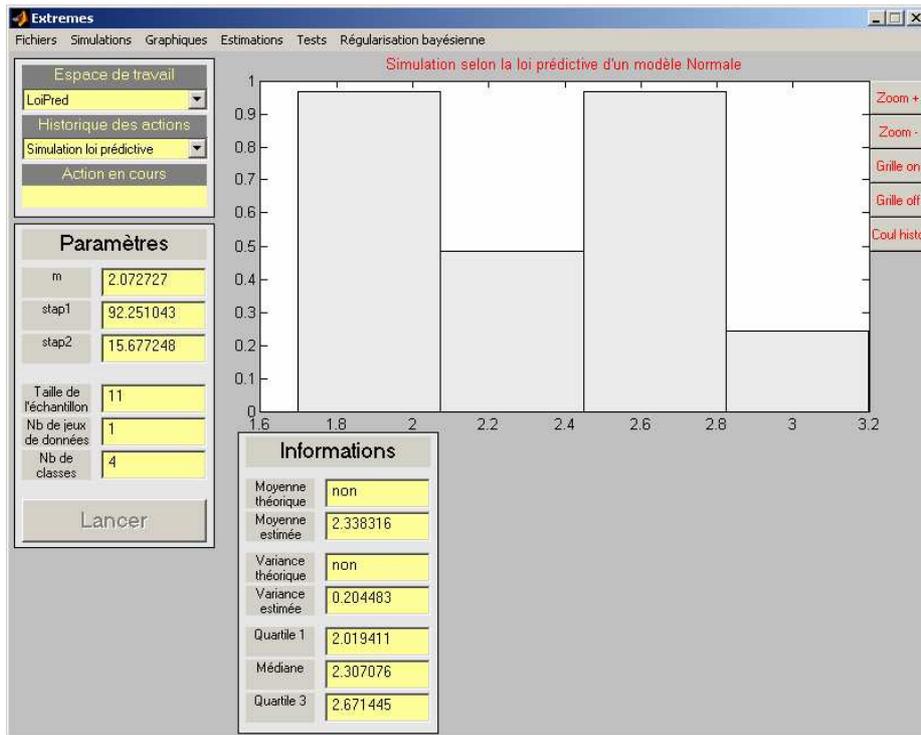
- Simuler selon la loi prédictive

Il est maintenant possible de simuler des variables aléatoires selon la loi prédictive. En effet, la mention **Loi prédictive** est dégrisée dans le menu **Simulations**. Un ou plusieurs jeux de données vont pouvoir être créés, c'est pourquoi le programme demande la saisie d'un nouvel espace de travail (par exemple **LoiPred**) afin d'exploiter ces nouvelles données. L'écran suivant est alors présenté :



Fenêtre 131 - Simulation selon la loi prédictive

Les paramètres de la loi prédictive sont affichés, ainsi que la taille de l'échantillon de départ qu'il est possible de modifier, tout comme les autres champs. En cliquant sur Lancer, on voit :



Fenêtre 132 - Résultat de la simulation

3.2.7. Le menu Aide



Fenêtre 133 - Organisation du menu Aide

L'aide en ligne est visible grâce à *Internet Explorer*. Si l'utilisateur ne dispose pas de ce logiciel, il peut ouvrir manuellement le fichier ManU.htm (dans le sous-répertoire **Doc** du répertoire d'installation) avec son propre navigateur.

3.3. Boutons de contrôle du graphique

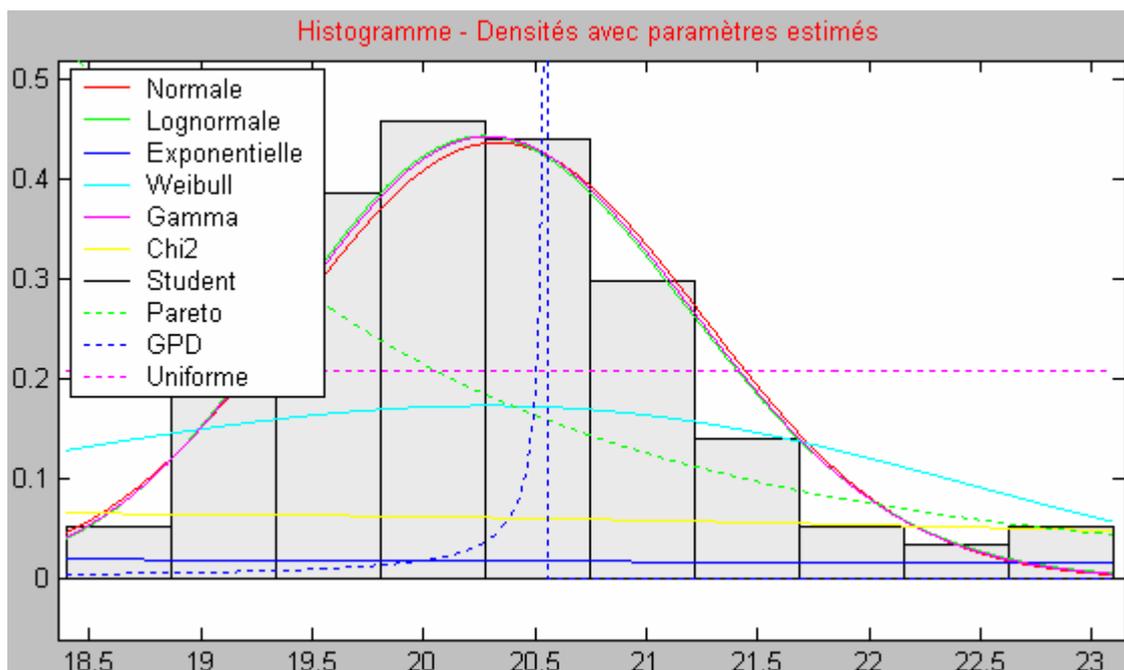
5 boutons sont associés au graphique. Ils sont utilisables à tout moment dans le logiciel et permettent une meilleure visualisation des sorties graphiques.



Fenêtre 134 - Contrôle du graphique

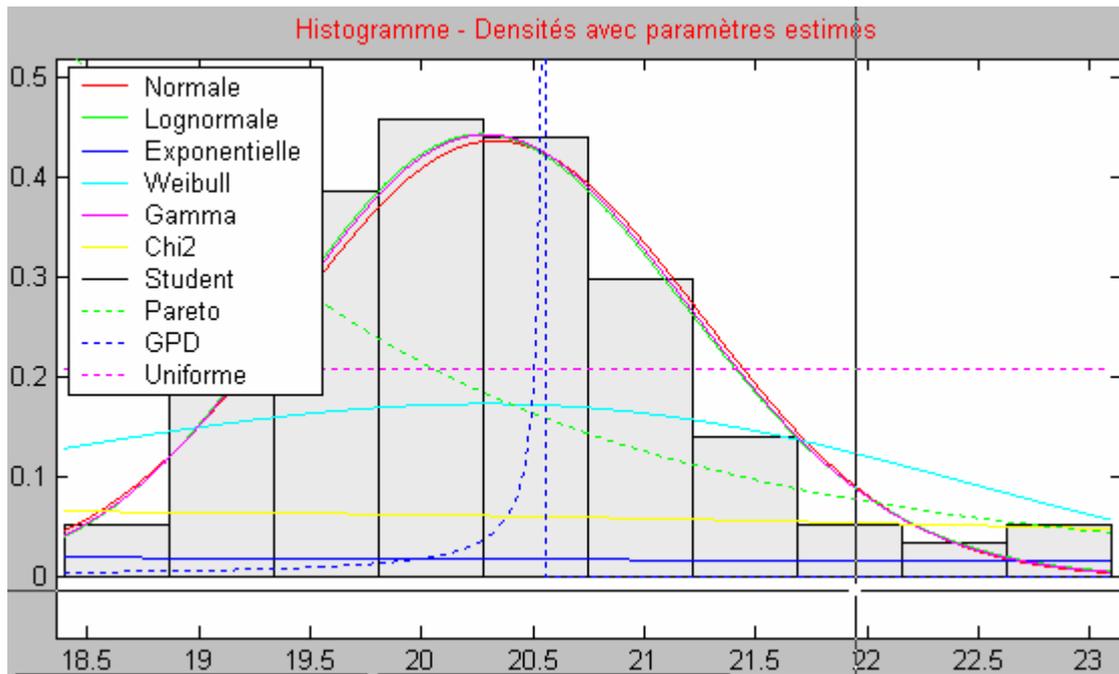
- Zoom + et Zoom -

Ces 2 accessoires permettent d'agrandir une zone du graphique (**Zoom +**) ou de revenir à l'écran précédent (**Zoom -**). L'utilisation du bouton **Zoom +** est simple. Pour grossir une partie du graphique, il suffit d'imaginer une petite boîte encadrant la zone à agrandir, puis de cliquer une fois au coin inférieur gauche et une fois au coin supérieur droit. Par exemple, considérons l'écran suivant :

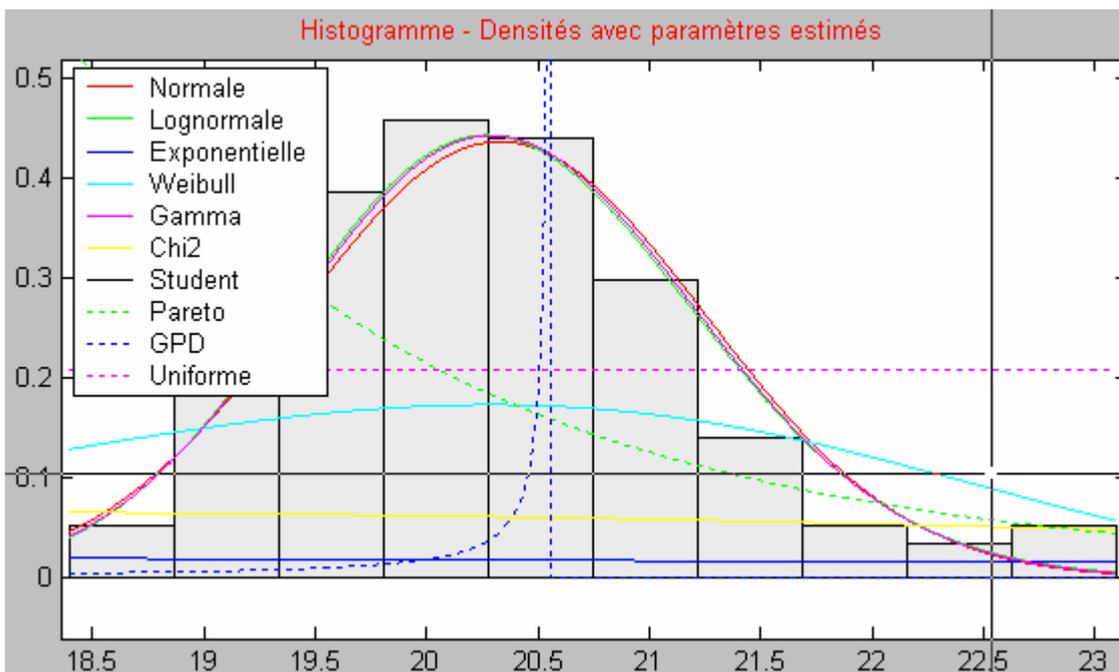


Fenêtre 135 - Graphique de densités

On veut zoomer sur l'intervalle $[22, 22.5]$ afin de dissocier les courbes. En cliquant sur **Zoom +**, le curseur de la souris devient une croix que l'on peut déplacer partout sur le graphique. On peut imaginer une petite boîte autour de l'intervalle voulu et on clique au coin inférieur gauche puis au coin supérieur droit :

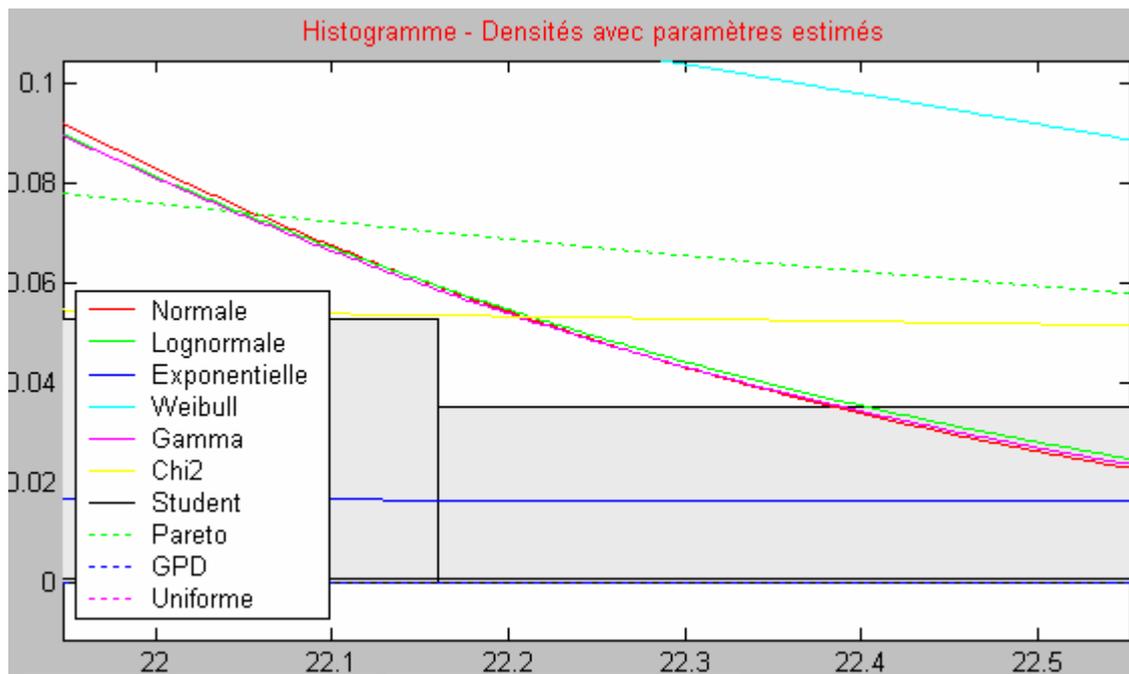


Fenêtre 136 - Zoom avant



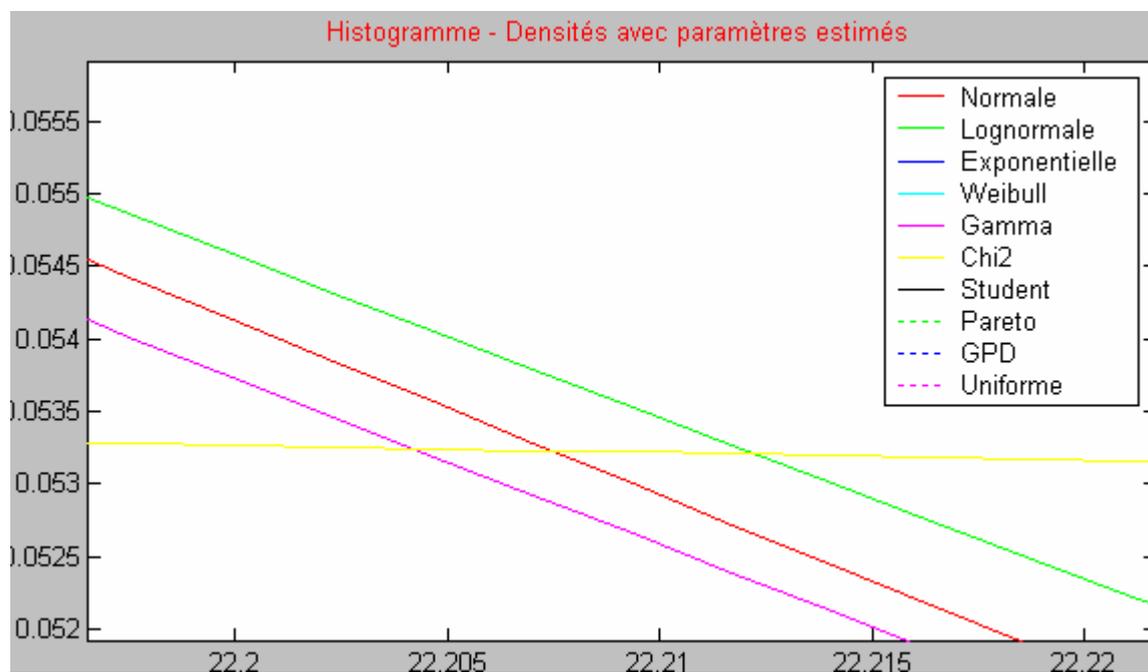
Fenêtre 137 - Zoom avant (2)

Après le 2^{ème} clic, on obtient le graphique voulu :



Fenêtre 138 - Zoom sur une partie du graphique

On peut répéter l'opération pour mieux voir l'intersection des courbes vertes, rouges et jaunes :



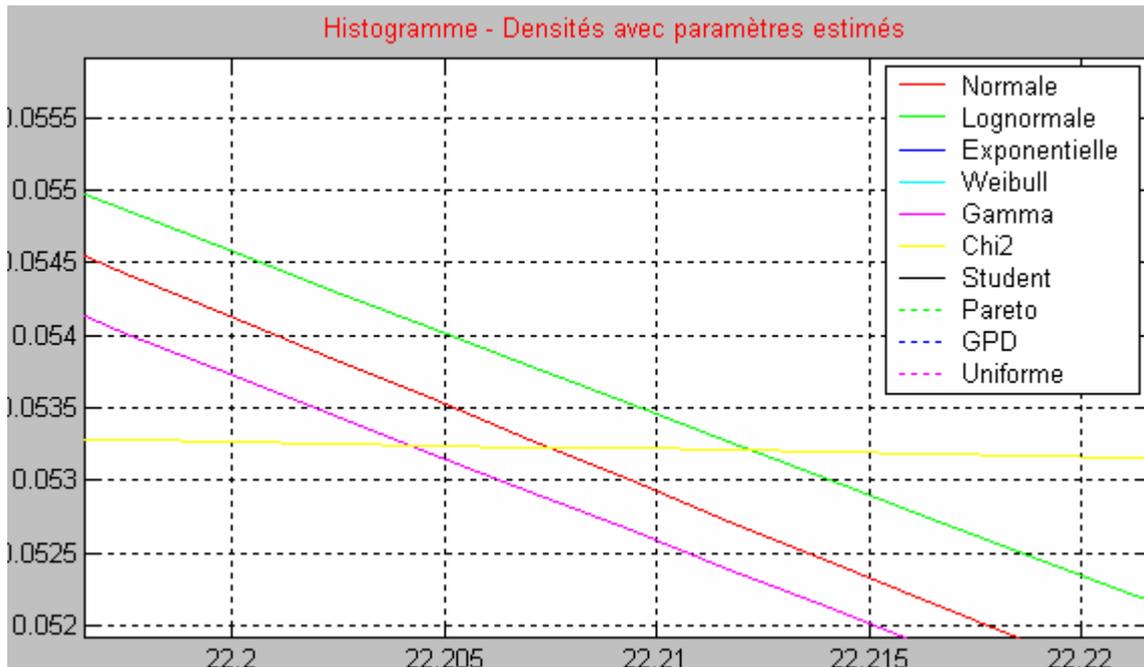
Fenêtre 139 - Zoom avant (3)

Un 1^{er} clic sur **Zoom** – fait réapparaître la Fenêtre 138 et un 2^{ème} la Fenêtre 137. On peut cliquer sur **Zoom** – autant de fois que l'on a fait de Zoom avant.

Remarque 10 : Il est aisé de déplacer la légende. Pour cela, il suffit de laisser appuyer le bouton gauche de la souris dessus, puis de la faire glisser à l'endroit voulu.

- Grille on et Grille off

Ces 2 boutons font respectivement apparaître et disparaître la grille sur le graphique. En revenant sur la Fenêtre 139 et en appuyant sur **Grille on**, on obtient :

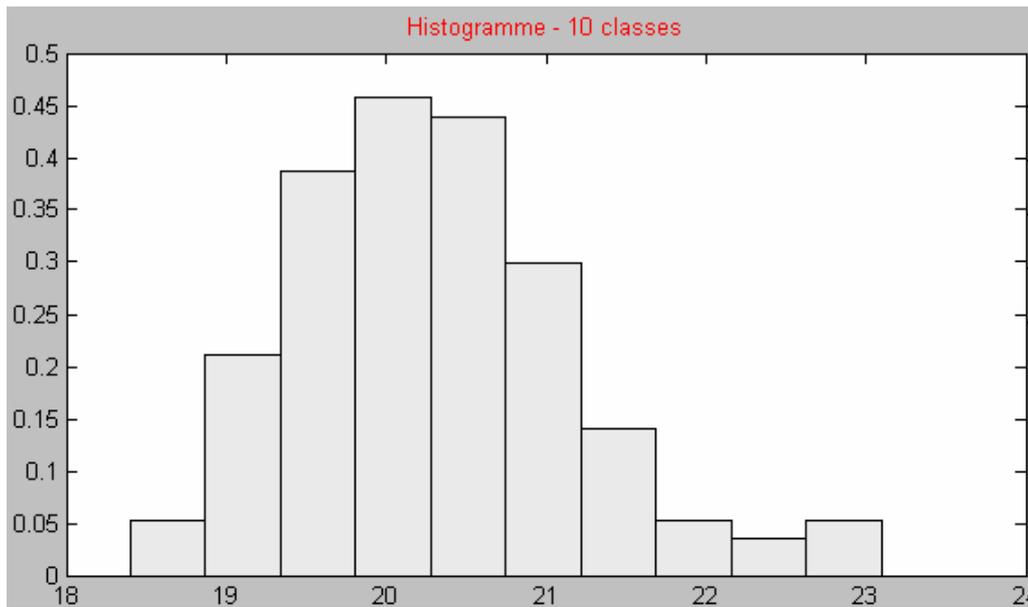


Fenêtre 140 - Affichage de la grille

Un clic sur **Grille off** fera disparaître la grille.

- Coul histo

Ce bouton permet de changer la couleur de remplissage des histogrammes. La couleur par défaut est le rouge. Par exemple, ceci est utile lors de la superposition de plusieurs histogrammes. Prenons l'exemple du graphique suivant :



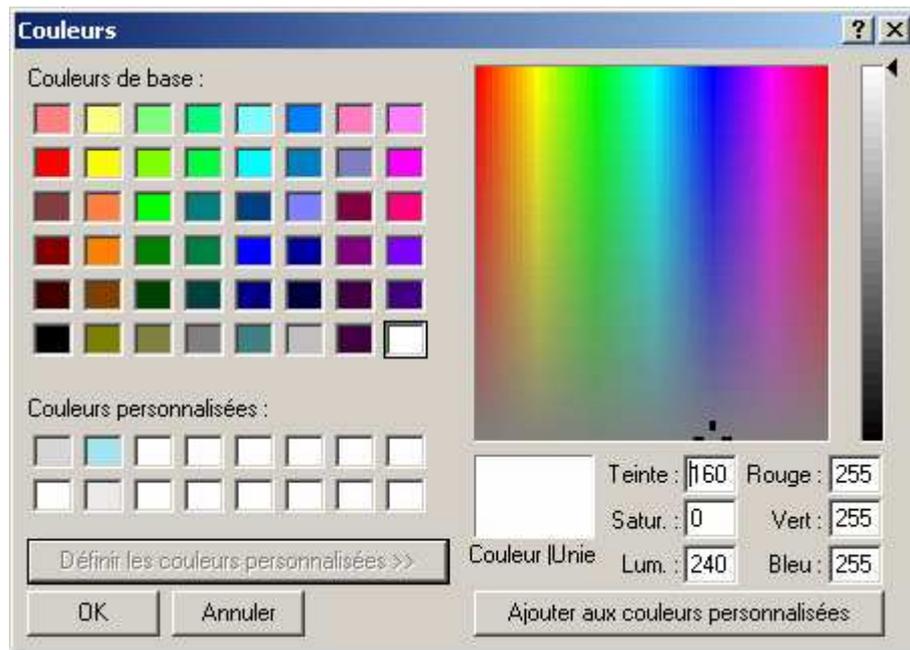
Fenêtre 141 – Histogramme

En cliquant sur *Coul histo*, une petite fenêtre s'affiche :



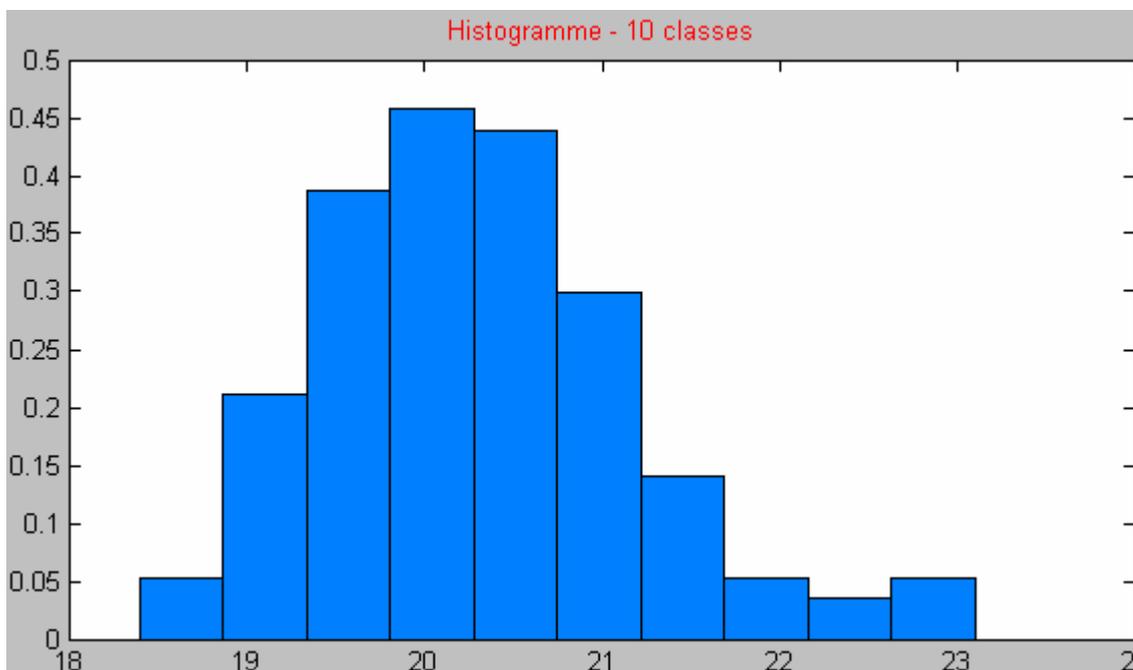
Fenêtre 142 - Palette des couleurs

Il suffit ensuite de sélectionner une couleur proposée ou alors de fabriquer sa propre couleur en sélectionnant *Définir les couleurs personnalisées* :



Fenêtre 143 - Palette des couleurs (2)

Choisissons par exemple un bleu quelconque et appuyons sur **OK**. La nouvelle couleur de l'histogramme ne sera effective que lors d'une prochaine création d'un histogramme ou alors en réaffichant celui-ci par l'intermédiaire de la liste déroulante *Historique des actions* :



Fenêtre 144 - Nouvelle couleur de l'histogramme

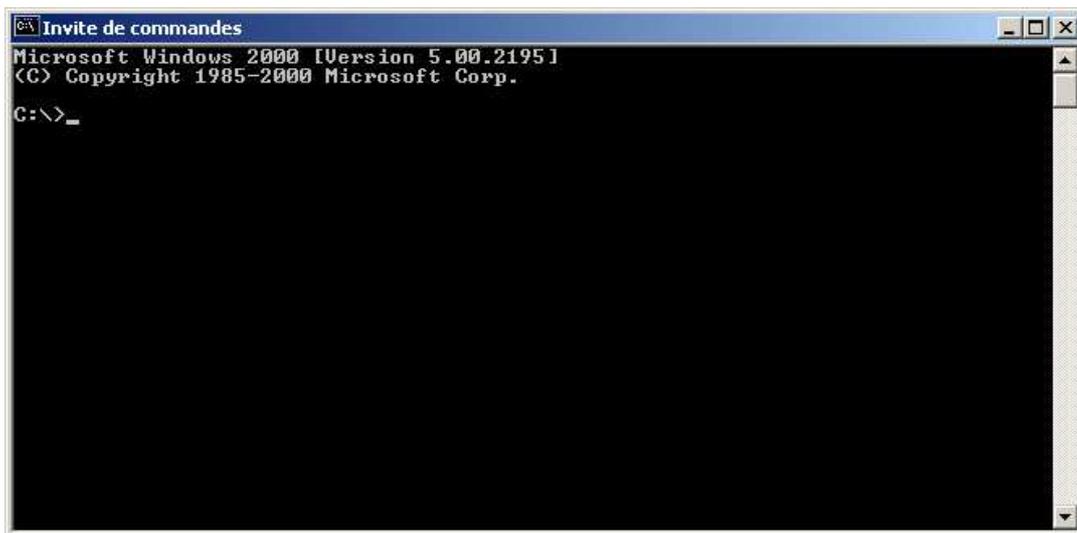
4. EXECUTION MANUELLE

Le logiciel *EXTREMES* peut être utilisé sans interface graphique. Pour cela, il est nécessaire de construire des fichiers d'entrée de type texte qui seront des paramètres de l'exécutable *liste.exe*. Ce programme génère des fichiers de sorties de type texte. Il y a 5 grands modèles de fichiers d'entrée du programme (ce qui sous entend 5 types de chargements), et 46 fonctionnalités différentes désignés par des « numéros de tâches » (de 1 à 46). Des exemples de fichiers pour chaque fonctionnalité sont décrits en annexe C.

Les fichiers en argument de l'exécutable *liste.exe* doivent impérativement être appelés *Tmpxx.txt* où *xx* est un numéro (par exemple *Tmp1.txt*).

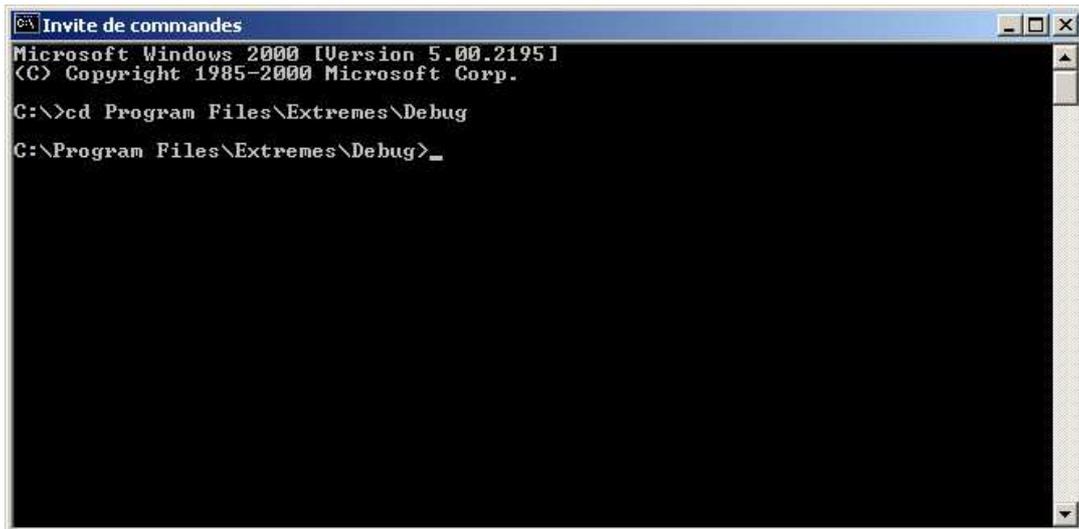
4.1. Ligne de commandes

Pour ouvrir une fenêtre MS-DOS sous Windows, il faut cliquer sur *Démarrer*, puis *Programmes*. En général, *Invite de commandes* se trouve dans le groupe *Accessoires*. On découvre la fenêtre suivante :



Fenêtre 145 - Invite de commandes

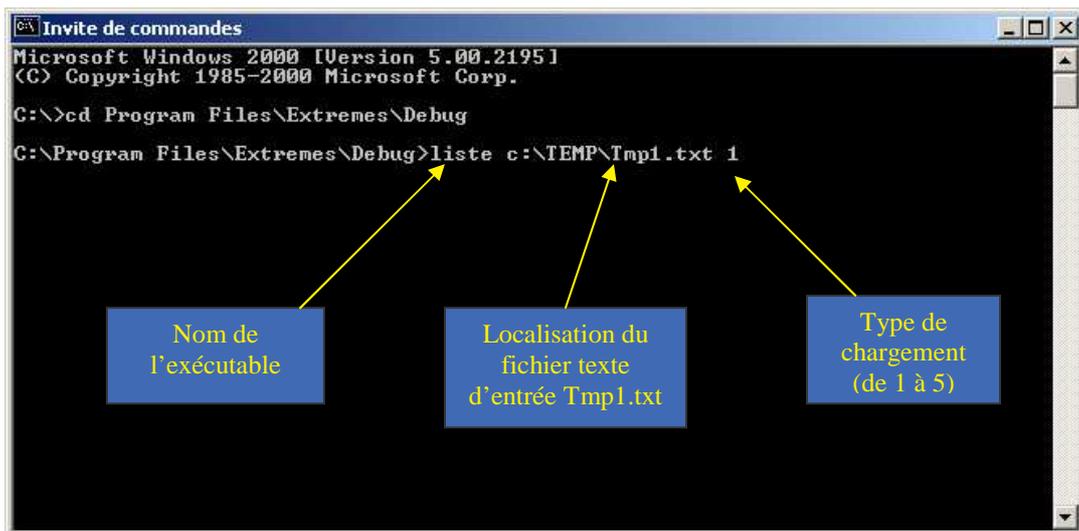
Supposons qu'*EXTREMES* soit installé sous *c:\Program Files\Extremes* et que le fichier d'entrée *Tmp1.txt* se situe dans le répertoire *c:\TEMP*. Il faut se placer dans le sous-répertoire *Debug* d'*Extremes* :



```
Microsoft Windows 2000 [Version 5.00.2195]
(C) Copyright 1985-2000 Microsoft Corp.
C:\>cd Program Files\Extremes\Debug
C:\Program Files\Extremes\Debug>_
```

Fenêtre 146 - Localisation de l'exécutable liste.exe

Il suffit ensuite de taper la ligne de commandes suivante :



```
Microsoft Windows 2000 [Version 5.00.2195]
(C) Copyright 1985-2000 Microsoft Corp.
C:\>cd Program Files\Extremes\Debug
C:\Program Files\Extremes\Debug>liste c:\TEMP\Tmp1.txt 1
```

Nom de l'exécutable

Localisation du fichier texte d'entrée Tmp1.txt

Type de chargement (de 1 à 5)

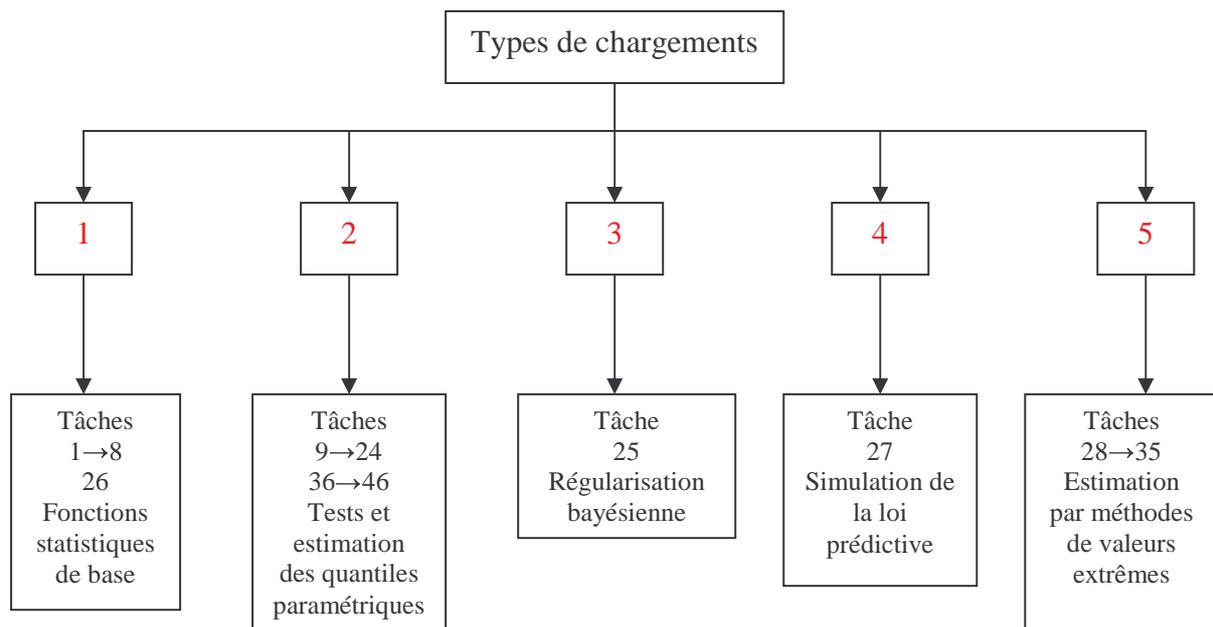
Fenêtre 147 - Exécution du programme

Selon la fonctionnalité voulue, le programme créera un ou plusieurs fichiers résultats (nommés comme indiqué dans le fichier d'entrée. Il est recommandé d'utiliser des noms de type *Resultxx.txt* où *xx* est le même numéro que le fichier d'entrée *Tmpxx.txt*). Voyons maintenant quel type de chargement utiliser en fonction de l'opération demandée par l'utilisateur ainsi que les modèles de fichiers d'entrées.

4.2. Fichier d'entrée du programme

4.2.1. Type de chargement et numéros de tâches

La concordance entre les numéros de tâches et leurs véritables significations est explicitée en annexe B. Les schémas ci-dessous désignent le type de chargement à utiliser en fonction de la fonctionnalité demandée, représentée ici par son numéro de tâche.



4.2.2. Modèles des fichiers d'entrée selon le type de chargement

- Type de chargement 1

Pour les tâches 1 à 8 :

```

Chemin_de_travail (ex : c:\extremes\Temp\loinormale\
Nom_de_fichier_de_données ou * si inexistant (ex : c:\extremes\Temp\loinormale\Donnees.txt) (ex : *)
Numéro_de_tâche (ex : 3)
Nom_de_fichier_résultat (ex : c:\extremes\Temp\loinormale\Result2.txt)
Nom_de_loi ou * si inexistant (ex : Weibull)
1er_paramètre ou Choix_de_densité (ex : 2.63) (ex : 1 ou 2)
2ème_paramètre ou * si inexistant (ex : 0.25698) (ex : *)
3ème_paramètre ou * si inexistant ou parametre_de_lissage (ex : 125.2369) (ex : *) (ex : 0.2)
Taille_de_l'échantillon ou Borne_inférieure_d'un_intervalle (ex : 10000) (ex : -2)
Nombre_de_simulations ou Borne_supérieure_d'un_intervalle (ex : 10) (ex : 3.5)
Nombre_de_classe_de_l'histogramme ou Nombre_de_points (ex : 50) (ex : 2000)

```

Pour la tâche 26 :

```

Chemin_de_travail (ex : c:\extremes\Temp\loinormale\
Nom_des_fichiers_de_données = Tous
Numéro_de_tâche (ex : 3)
Nom_de_fichier_résultat (ex : c:\extremes\Temp\loinormale\Result2.txt)
Nom_de_modèle (ex : Weibull)
*
*
*
0
0
0

```

- Type de chargement 2

Pour les tâches 9 à 24 et 36 à 46 :

```

Chemin_de_travail (ex : c:\extremes\Temp\loinormale\
Nom_de_fichier_de_données ou tous (ex : c:\extremes\Temp\loinormale\Donnees.txt) (ex : Tous)
Numéro_de_tâche (ex : 10)
Nom_de_fichier_résultat (ex : c:\extremes\Temp\loinormale\Result8.txt)
Nom_des_modèles
Nombre_d'excès ou 0 si inexistant (ex : 200) (ex : 0)
Niveau_de_signification_du_test ou 0 si inexistant (ex : 0.05) (ex : 0)
Borne_inférieure_des_excès ou 0 si inexistant (ex : 100) (ex : 0)
Borne_supérieure_des_excès ou 0 si inexistant (ex : 110) (ex : 0)
Version_test_ET ou 0 si inexistant (ex : 1 = asymptotique) (ex : 2 = bootstrap complet) (ex : 3 = simplifié)
Nombre_d'échantillons_de_bootstrap ou 0 si inexistant (ex : 500) (ex : 0)
Valeur_de_p_du_quantile_d'ordre_1-p ou 0 si inexistant (ex : 0.001) (ex : 0)
Nom_de_méthode_d'estimation ou rien (ex : HillG) (ex : )

```

- Type de chargement 3 (tâche 25)

Chemin_de_travail (ex : c:\extremes\Temp\loinormale\
Nom_de_fichier_de_données (ex : c:\extremes\Temp\loinormale\Donnees.txt)
Numéro_de_tâche (ex : 25)
Nom_de_fichier_résultat (ex : c:\extremes\Temp\loinormale\Result15.txt)
Nom_de_modèle (ex : Gamma)
Borne_inférieure_pour_la_loi_beta_a_priori ou * si inexistant (Weibull paramètre de forme) (ex : 0.32)
Borne_supérieure_pour_la_loi_beta_a_priori ou * si inexistant (Weibull paramètre de forme) (ex : 0.88)
Simulées (ex : 1 si données simulées, 0 sinon)
Nom_de_la_vraie_loi ou * si inexistant (ex : Exponentielle) (ex : *)
Paramètre_1_de_vraie_loi ou * si inexistant (ex : 0.5) (ex : *)
Paramètre_2_de_vraie_loi ou * si inexistant (ex : 2) (ex : *)
Paramètre_3_de_vraie_loi ou * si inexistant (ex : 15.26) (ex : *)
Avis_d'expert (ex : 1 si oui, 0 sinon)
Qmax ou * si inexistant (ex : 18.5698) (ex : *)
p1 ou * si inexistant (ex : 0.01) (encadrement en terme d'ordre (1 - p1) de la valeur du quantile de l'expert)
p2 ou * si inexistant (ex : 0.0001) (ex : *)
Epsilon (confiance en l'expert) (ex : 0.01)
Confiance ou * si inexistant (concerne Weibull paramètre de forme) (ex : Forte ou Moyenne ou Faible)
Type_de_test_central (ex : AD ou CVM)
Niveau_de_signification_test_central (ex : 0.05)
Niveau_de_signification_test_ET (ex : 0.05)
Nombre_d'excès (ex : 200)
pET (1 - pET = 1 - ordre du quantile) (ex : 0.001)
q1 ou 0 si inexistant (1 - q1 = 1 - ordre du 1^{er} quantile à calculer) (ex : 0.01)
q2 ou 0 si inexistant (ex : 0.001)
q3 ou 0 si inexistant (ex : 0.0001)
q4 ou 0 si inexistant (ex : 0.00001)
q5 ou 0 si inexistant (ex : 0.000001)

- Type de chargement 4 (tâche 27)

Chemin_de_travail (ex : c:\extremes\Temp\pred\
Nom_de_fichier_de_données (ex : c:\extremes\Temp\RB\Donnees.txt)
Numéro_de_tâche (ex : 27)
Nom_de_fichier_résultat (ex : c:\extremes\Temp\pred\Result1.txt)
Nom_du_modèle_de_régularisation_bayésienne (ex : Normale)
1^{er} paramètre_de_la_loi_a_priori (ex : 86.751)
2^{ème} paramètre_de_la_loi_a_priori (ex : 14.859)
3^{ème} paramètre_de_la_loi_a_priori ou * si inexistant (ex : 2.156) (ex : *)
Taille_de_l'échantillon (ex : 10000)
Nombre_de_simulations (ex : 10)
Nombre_de_classe_de_l'histogramme (ex : 50)
Borne_inférieure_pour_la_loi_beta_a_priori ou * si inexistant (Weibull paramètre de forme) (ex : 0.32)
Borne_supérieure_pour_la_loi_beta_a_priori ou * si inexistant (Weibull paramètre de forme) (ex : 0.88)
1^{er} paramètre_de_la_loi_prédictive (ex : 86.751)
2^{ème} paramètre_de_la_loi_prédictive (ex : 14.859)
3^{ème} paramètre_de_la_loi_prédictive ou * si inexistant (ex : 2.156) (ex : *)

- Type de chargement 5

Pour les tâches 28 à 35 :

<p><i>Chemin_de_travail</i> (ex : c:\extremes\Temp\loinormale\) <i>Nom_de_fichier_de_données</i> ou <i>tous</i> (ex : c:\extremes\Temp\loinormale\Donnees.txt) (ex : Tous) <i>Numéro_de_tâche</i> (ex : 10) <i>Nom_de_fichier_résultat</i> (ex : c:\extremes\Temp\loinormale\Result8.txt) <i>Nom_des_méthodes_d'estimation</i> <i>Nombre_d'excès</i> ou <i>borne_inférieure_des_excès</i> (ex : 20) <i>Borne_supérieure_des_excès</i> ou 0 si inexistant (ex : 50) (ex : 0) <i>Ordre_du_quantile_extrême</i> ou 0 si inexistant (ex : 0.01) (ex : 0)</p>

Les valeurs prises par certaines variables (Nom de modèles, nom de méthode d'estimation, ...) sont décrites en annexe C.

4.3. Fichiers de sortie

Il existe pratiquement un fichier de sortie différent pour chaque fonctionnalité, ce qui donc en représente à peine moins de 46. Des exemples détaillés pour chaque tâche sont exposés en annexe D.

ANNEXE A – Paramétrage des lois

- **Loi Normale**

$$N(m, \sigma) : f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) \quad x \in \mathbb{R}$$

- **Loi Lognormale**

$$\text{Logn}(m, \sigma) : f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln(x)-m)^2}{2\sigma^2}\right) \quad x \in \mathbb{R}^{+*}$$

- **Loi Exponentielle**

$$\text{Exp}(\lambda) : F(x) = 1 - \exp\left(-\frac{x}{\lambda}\right) \quad x \in \mathbb{R}^+$$

Régularisation bayésienne :

$$\text{Exp}(\lambda) : F(x) = 1 - \exp(-\lambda x) \quad x \in \mathbb{R}^+$$

- **Loi Gamma**

$$G(a, b) : f(x) = \frac{1}{b^a \Gamma(a)} e^{-\frac{x}{b}} x^{a-1} \quad x \in \mathbb{R}^{+*} \text{ si } a < 1, x \in \mathbb{R}^+ \text{ si } a \geq 1$$

- **Loi de Weibull**

$$W(\eta, \beta) : F(x) = 1 - \exp\left(-\frac{x^\beta}{\eta}\right) \quad x \in \mathbb{R}^+$$

Régularisation bayésienne :

$$W(\eta, \beta) : F(x) = 1 - \exp\left(-\left(\frac{x}{\eta}\right)^\beta\right) \quad x \in \mathbb{R}^+$$

- **Loi Beta**

$$\text{Beta}(a, b) : f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1} (1-x)^{b-1} \quad x \in [0,1]$$

- **Loi de Student**

$$T(n): f(x) = \frac{1}{\sqrt{n\pi}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\left(\frac{n+1}{2}\right)} \quad x \in \mathbb{R}^+$$

- **Loi du Chi2**

$$X^2(n): f(x) = \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} x^{\frac{n-2}{2}} e^{-\frac{x}{2}} \quad x \in \mathbb{R}^{+*}$$

- **Loi de Pareto**

$$Pa(x_0, \alpha): F(x) = 1 - \left(\frac{x_0}{x}\right)^\alpha \quad x \geq x_0$$

- **Loi GPD**

$GPD(\gamma, \sigma)$:

$$\text{Si } \gamma \neq 0, f(x) = \frac{1}{\sigma} \left(1 + \frac{x}{\sigma}\right)^{\frac{1}{\gamma}-1} \begin{cases} x \in \mathbb{R}^+ \text{ si } \gamma > 0 \\ x \in \left[0, -\frac{\sigma}{\gamma}\right] \text{ si } \gamma < 0 \end{cases}$$

$$\text{Si } \gamma = 0, f(x) = \frac{1}{\sigma} \exp\left(-\frac{x}{\sigma}\right) \quad x \in \mathbb{R}^+$$

Régularisation bayésienne :

$$GPD(\alpha, \beta): f(x) = \frac{\alpha}{\beta} \left(1 + \frac{x}{\beta}\right)^{-\alpha-1} \quad \text{pour } \begin{cases} x \in \mathbb{R}^+ \text{ si } \alpha > 0 \\ x \in [0, -\beta] \text{ si } \alpha < 0 \end{cases} \quad \alpha = \frac{1}{\gamma} \text{ et } \beta = \frac{\sigma}{\gamma}$$

ANNEXE B – Nom réel de la fonctionnalité et numéro de tâche

Numéro de tâche → Nom réel :

- 1 = Simulation
- 2 = Densité
- 3 = Fonction de répartition
- 4 = Fonction quantile
- 5 = Histogramme
- 6 = Estimation de la densité
- 7 = Estimation des paramètres
- 8 = Fonction de survie
- 9 = Test d'Anderson-darling
- 10 = Test de Cramer-Von Mises
- 11 = Test d'Anderson-Darling – multi données
- 12 = Test de Cramer-Von Mises – multi données
- 13 = Test d'exponentialité des excès – QQ-plot
- 14 = Test d'exponentialité des excès – mode multi données
- 15 = Test d'exponentialité des excès – excès multiples
- 16 = Test ET – multi données – version asymptotique
- 17 = Test ET – multi données – version bootstrap paramétrique complet
- 18 = Test ET – multi données – version bootstrap paramétrique simplifié
- 19 = Test ET – 1 seul nombre d'excès – version asymptotique
- 20 = Test ET – 1 seul nombre d'excès – version bootstrap paramétrique complet
- 21 = Test ET – 1 seul nombre d'excès – version bootstrap paramétrique simplifié
- 22 = Test ET – Excès multiples – version asymptotique
- 23 = Test ET – Excès multiples – version bootstrap paramétrique complet
- 24 = Test ET – Excès multiples – version bootstrap paramétrique simplifié
- 25 = Régularisation bayésienne
- 26 = Estimation des paramètres – multi données
- 27 = Simulation selon la loi prédictive
- 28 = Indice des valeurs extrêmes – 1 seule méthode d'estimation – 1 seul nombre d'excès – QQ-plot
- 29 = Indice des valeurs extrêmes – plusieurs méthodes d'estimation – 1 seul nombre d'excès
- 30 = Indice des valeurs extrêmes – plusieurs méthodes d'estimation – Excès multiples
- 31 = Indice des valeurs extrêmes – multi données
- 32 = Estimation d'un quantile extrême – 1 seule méthode d'estimation – 1 seul nombre d'excès
- 33 = Estimation d'un quantile extrême – plusieurs méthodes d'estimation – 1 seul nombre d'excès
- 34 = Estimation d'un quantile extrême – plusieurs méthodes d'estimation – Excès multiples
- 35 = Estimation d'un quantile extrême – multi données
- 36 = Estimation d'un quantile paramétrique
- 37 = Estimation d'un quantile paramétrique – multi données
- 38 = Test GPD – multi données – version asymptotique (non disponible)
- 39 = Test GPD – multi données – version bootstrap paramétrique complet
- 40 = Test GPD – multi données – version bootstrap paramétrique simplifié
- 41 = Test GPD – 1 seul nombre d'excès – version asymptotique (non disponible)
- 42 = Test GPD – 1 seul nombre d'excès – version bootstrap paramétrique complet
- 43 = Test GPD – 1 seul nombre d'excès – version bootstrap paramétrique simplifié
- 44 = Test GPD – Excès multiples – version asymptotique (non disponible)
- 45 = Test GPD – Excès multiples – version bootstrap paramétrique complet
- 46 = Test GPD – Excès multiples – version bootstrap paramétrique simplifié

Nom réel → Numéro de tâche :

Densité = 2
 Fonction de répartition = 3
 Estimation d'un quantile extrême – 1 seule méthode d'estimation – 1 seul nombre d'excès = 32
 Estimation d'un quantile extrême – multi données = 35
 Estimation d'un quantile extrême – plusieurs méthodes d'estimation – 1 seul nombre d'excès = 33
 Estimation d'un quantile extrême – plusieurs méthodes d'estimation – Excès multiples = 34
 Estimation d'un quantile paramétrique = 36
 Estimation d'un quantile paramétrique – multi données = 37
 Estimation de la densité = 6
 Estimation des paramètres = 7
 Estimation des paramètres – multi données = 26
 Fonction de survie = 8
 Fonction quantile = 4
 Histogramme = 5
 Indice des valeurs extrêmes – 1 seule méthode d'estimation – 1 seul nombre d'excès – QQ-plot = 28
 Indice des valeurs extrêmes – multi données = 31
 Indice des valeurs extrêmes – plusieurs méthodes d'estimation – 1 seul nombre d'excès = 29
 Indice des valeurs extrêmes – plusieurs méthodes d'estimation – Excès multiples = 30
 Régularisation bayésienne = 25
 Simulation = 1
 Simulation selon la loi prédictive = 27
 Test d'Anderson-Darling = 9
 Test d'Anderson-Darling – multi données = 11
 Test d'exponentialité des excès – excès multiples = 15
 Test d'exponentialité des excès – mode multi données = 14
 Test d'exponentialité des excès – QQ-plot = 13
 Test de Cramer-Von Mises = 10
 Test de Cramer-Von Mises – multi données = 12
 Test ET – 1 seul nombre d'excès – version asymptotique = 19
 Test ET – 1 seul nombre d'excès – version bootstrap paramétrique complet = 20
 Test ET – 1 seul nombre d'excès – version bootstrap paramétrique simplifié = 21
 Test ET – Excès multiples – version asymptotique = 22
 Test ET – Excès multiples – version bootstrap paramétrique complet = 23
 Test ET – Excès multiples – version bootstrap paramétrique simplifié = 24
 Test ET – multi données – version asymptotique = 16
 Test ET – multi données – version bootstrap paramétrique complet = 17
 Test ET – multi données – version bootstrap paramétrique simplifié = 18
 Test GPD – 1 seul nombre d'excès – version asymptotique (non disponible) = 41
 Test GPD – 1 seul nombre d'excès – version bootstrap paramétrique complet = 42
 Test GPD – 1 seul nombre d'excès – version bootstrap paramétrique simplifié = 43
 Test GPD – Excès multiples – version asymptotique (non disponible) = 44
 Test GPD – Excès multiples – version bootstrap paramétrique complet = 45
 Test GPD – Excès multiples – version bootstrap paramétrique simplifié = 46
 Test GPD – multi données – version asymptotique (non disponible) = 38
 Test GPD – multi données – version bootstrap paramétrique complet = 39
 Test GPD – multi données – version bootstrap paramétrique simplifié = 40

ANNEXE C – Exemples de fichiers d’entrée

Tâche 1 – Simulation

```
z:\extremes\Temp\gamma\  
*  
1  
z:\extremes\Temp\gamma\Result1.txt  
Gamma  
2  
0.5  
*  
100  
1  
10
```

Tâche 2 – Densité

```
z:\extremes\Temp\gamma\  
*  
2  
z:\extremes\Temp\gamma\Result2.txt  
Gamma  
2  
0.5  
*  
0  
5  
1000
```

Tâche 3 – Fonction de répartition

```
z:\extremes\Temp\gamma\  
*  
3  
z:\extremes\Temp\gamma\Result3.txt  
Gamma  
2  
0.5  
*  
0  
5  
1000
```

Tâche 4 – Fonction quantile

```
z:\extremes\Temp\gamma\  
*  
4  
z:\extremes\Temp\gamma\Result4.txt  
Gamma  
2  
0.5  
*  
0.01  
0.99  
1000
```

Tâche 5 – Histogramme

```
z:\extremes\Temp\gamma\  
z:\extremes\Temp\gamma\Donnees.txt  
5  
z:\extremes\Temp\gamma\Result6.txt  
*  
*  
*  
*  
*  
10
```

Tâche 6 – Estimation de la densité

```
z:\extremes\Temp\gamma\  
*  
6  
z:\extremes\Temp\gamma\Result7.txt  
*  
1  
*  
0.2  
0  
3  
200
```

Tâche 7 – Estimation des paramètres

```
z:\extremes\Temp\gamma\  
z:\extremes\Temp\gamma\Donnees.txt  
7  
z:\extremes\Temp\gamma\Result8.txt  
Gamma  
*  
*  
*  
0  
0  
0
```

Tâche 8 – Fonction de survie

```
z:\extremes\Temp\gamma\  
*  
8  
z:\extremes\Temp\gamma\Result5.txt  
Gamma  
2  
0.5  
*  
0  
5  
1000
```

Tâche 9 – Test d'Anderson-Darling

```
z:\extremes\Temp\gamma\  
z:\extremes\Temp\gamma\Donnees.txt  
9  
z:\extremes\Temp\gamma\Result11.txt  
Normale  
Lognormale  
Exponentielle  
Weibull  
Gamma  
Chi2  
Student  
Pareto  
GPD  
Uniforme  
Beta  
0  
0.05  
0  
0  
0  
0  
0  
0
```

=====
=====
Tâche 10 – Test de Cramer-Von Mises

z:\extremes\Temp\gamma\
z:\extremes\Temp\gamma\Donnees.txt
10
z:\extremes\Temp\gamma\Result22.txt
Normale
Lognormale
Exponentielle
Weibull
Gamma
Chi2
Student
Pareto
GPD
Uniforme
Beta
0
0.05
0
0
0
0
0
0

=====

Tâche 11 – Test d'Anderson-darling – multi données

```
z:\extremes\Temp\gamma1000\  
Tous  
11  
z:\extremes\Temp\gamma1000\Result3.txt  
Gamma  
0  
0.05  
0  
0  
0  
0  
0
```

=====
=====
Tâche 12 – Test de Cramer-Von Mises – multi données

```
z:\extremes\Temp\gamma1000\  
Tous  
12  
z:\extremes\Temp\gamma1000\Result4.txt  
Weibull  
0  
0.05  
0  
0  
0  
0  
0  
0
```

Tâche 13 – Test d’exponentialité des excès – QQ-Plot

```
z:\extremes\Temp\gamma\  
z:\extremes\Temp\gamma\Donnees.txt  
13  
z:\extremes\Temp\gamma\Result33.txt  
Exponentielle  
20  
0.05  
0  
0  
0  
0  
0
```

Tâche 14 – Test d'exponentialité des excès – multi données

```
z:\extremes\Temp\gamma1000\  
z:\extremes\Temp\gamma1000\Donnees.txt  
14  
z:\extremes\Temp\gamma1000\Result5.txt  
Exponentielle  
20  
0.05  
0  
0  
0  
0  
0
```

Tâche 15 – Test d'exponentialité des excès – Excès multiples

```
z:\extremes\Temp\gamma\  
z:\extremes\Temp\gamma\Donnees.txt  
15  
z:\extremes\Temp\gamma\Result34.txt  
Exponentielle  
0  
0.05  
0  
0  
0  
0  
0  
0
```

=====
=====
Tâche 16 – Test ET – multi données – version asymptotique

```
z:\extremes\Temp\gamma1000\  
Tous  
16  
z:\extremes\Temp\gamma1000\Result6.txt  
Weibull  
20  
0.05  
0  
0  
1  
0  
0.01
```

Tâche 17 – Test ET – multi données – version bootstrap paramétrique

```
z:\extremes\Temp\gamma1000\  
Tous  
17  
z:\extremes\Temp\gamma1000\Result7.txt  
Weibull  
20  
0.05  
0  
0  
2  
500  
0.01
```

Tâche 18 – Test ET – multi données – version bootstrap paramétrique simplifié

```
z:\extremes\Temp\gamma1000\  
Tous  
18  
z:\extremes\Temp\gamma1000\Result8.txt  
Weibull  
20  
0.05  
0  
0  
3  
200  
0.01
```

=====

Tâche 19 – Test ET – 1 seul nombre d’excès – version asymptotique

```
z:\extremes\Temp\gamma\  
z:\extremes\Temp\gamma\Donnees.txt  
19  
z:\extremes\Temp\gamma\Result35.txt  
Weibull  
Gamma  
20  
0.05  
0  
0  
1  
0  
0.01
```

**Tâche 20 – Test ET – 1 seul nombre d'excès – version bootstrap
paramétrique**

```
z:\extremes\Temp\gamma\  
z:\extremes\Temp\gamma\Donnees.txt  
20  
z:\extremes\Temp\gamma\Result36.txt  
Weibull  
Gamma  
20  
0.05  
0  
0  
2  
500  
0.01
```

**Tâche 21 – Test ET – 1 seul nombre d'excès – version bootstrap
paramétrique simplifié**

```
z:\extremes\Temp\gamma\  
z:\extremes\Temp\gamma\Donnees.txt  
21  
z:\extremes\Temp\gamma\Result37.txt  
Weibull  
Gamma  
20  
0.05  
0  
0  
3  
500  
0.01
```

Tâche 22 – Test ET – Excès multiples – version asymptotique

```
z:\extremes\Temp\gamma\  
z:\extremes\Temp\gamma\Donnees.txt  
22  
z:\extremes\Temp\gamma\Result38.txt  
Weibull  
Gamma  
0  
0.05  
10  
40  
1  
0  
0.01
```

Tâche 23 – Test ET – Excès multiples – version bootstrap paramétrique

```
z:\extremes\Temp\gamma\  
z:\extremes\Temp\gamma\Donnees.txt  
23  
z:\extremes\Temp\gamma\Result39.txt  
Weibull  
Gamma  
0  
0.05  
10  
40  
2  
500  
0.01
```

Tâche 24 – Test ET – Excès multiples – version bootstrap paramétrique simplifié

```
z:\extremes\Temp\gamma\  
z:\extremes\Temp\gamma\Donnees.txt  
24  
z:\extremes\Temp\gamma\Result40.txt  
Weibull  
Gamma  
0  
0.05  
10  
40  
3  
500  
0.01
```

=====
=====
Tâche 25 – Régularisation bayésienne

```
z:\extremes\Temp\RB\  
z:\extremes\Temp\RB\Donnees.txt  
25  
z:\extremes\Temp\RB\Result1.txt  
Normale  
*  
*  
0  
*  
*  
*  
*  
*  
1  
3.2  
0.01  
0.001  
0.01  
*  
CVM  
0.05  
0.05  
4  
0.01  
0.01  
0.001  
0.0001  
0  
0
```

Tâche 26 – Estimation des paramètres – multi données

```
z:\extremes\Temp\gamma1000\  
Tous  
26  
z:\extremes\Temp\gamma1000\Result2.txt  
Gamma  
*  
*  
*  
0  
0  
0
```

Tâche 27 – Simulation selon la loi prédictive

```
z:\extremes\Temp\predictive\  
z:\extremes\Temp\RB\Donnees.txt  
27  
z:\extremes\Temp\predictive\Result1.txt  
Normale  
86.751  
14.7363  
*  
11  
1  
20  
*  
*  
2.0727  
92.251  
15.6772
```

**Tâche 28 – Estimation de (γ, σ) - 1 seule méthode – 1 seul nombre d'excès –
QQ-Plot**

```
z:\extremes\Temp\gamma\  
z:\extremes\Temp\gamma\Donnees.txt  
29  
z:\extremes\Temp\gamma\Result45.txt  
HW  
20  
0  
0
```

Tâche 29 - Estimation de (γ, σ) - Plusieurs méthodes – 1 seul nombre d'excès

```
z:\extremes\Temp\gamma\  
z:\extremes\Temp\gamma\Donnees.txt  
29  
z:\extremes\Temp\gamma\Result45.txt  
HW  
EMV  
Hill  
HillG  
Zipf  
20  
0  
0
```

Tâche 30 - Estimation de (γ, σ) - Plusieurs méthodes – Excès multiples

```
z:\extremes\Temp\gamma\  
z:\extremes\Temp\gamma\Donnees.txt  
30  
z:\extremes\Temp\gamma\Result46.txt  
HW  
EMV  
Hill  
HillG  
Zipf  
20  
50  
0
```

=====

Tâche 31 - Estimation de (γ, σ) - multi données

```
z:\extremes\Temp\gamma1000\  
Tous  
31  
z:\extremes\Temp\gamma1000\Result11.txt  
Zipf  
20  
0  
0
```

**Tâche 32 – Estimation d'un quantile extrême – 1 seule méthode – 1 seul
nombre d'excès**

```
z:\extremes\Temp\gamma\  
z:\extremes\Temp\gamma\Donnees.txt  
32  
z:\extremes\Temp\gamma\Result47.txt  
HW  
25  
0  
0.01
```

Tâche 33 – Estimation d'un quantile extrême – Plusieurs méthodes – 1 seul nombre d'excès

```
z:\extremes\Temp\gamma\  
z:\extremes\Temp\gamma\Donnees.txt  
33  
z:\extremes\Temp\gamma\Result47.txt  
HW  
EMV  
Hill  
HillG  
Zipf  
25  
0  
0.01
```

Tâche 34 – Estimation d’un quantile extrême – Plusieurs méthodes – Excès multiples

z:\extremes\Temp\gamma\
z:\extremes\Temp\gamma\Donnees.txt
34
z:\extremes\Temp\gamma\Result48.txt
HW
EMV
Hill
HillG
Zipf
20
50
0.01

Tâche 35 – Estimation d'un quantile extrême – multi données

```
z:\extremes\Temp\gamma1000\  
Tous  
35  
z:\extremes\Temp\gamma1000\Result12.txt  
Zipf  
20  
0  
0.01
```

=====

Tâche 36 – Estimation d’un quantile paramétrique

```
z:\extremes\Temp\gamma\  
z:\extremes\Temp\gamma\Donnees.txt  
36  
z:\extremes\Temp\gamma\Result49.txt  
Normale  
Lognormale  
Exponentielle  
Weibull  
Gamma  
Chi2  
Student  
Pareto  
GPD  
Uniforme  
Beta  
0  
0  
0  
0  
0  
0  
0  
0.01
```

Tâche 37 – estimation d'un quantile paramétrique – multi données

```
z:\extremes\Temp\gamma1000\  
Tous  
37  
z:\extremes\Temp\gamma1000\Result13.txt  
Weibull  
0  
0  
0  
0  
0  
0  
0  
0.01
```

Tâche 39 – Test GPD – multi données – version bootstrap paramétrique

```
z:\extremes\Temp\gamma1000\  
Tous  
39  
z:\extremes\Temp\gamma1000\Result9.txt  
Weibull  
20  
0.05  
0  
0  
2  
200  
0.01  
HillG
```

Tâche 40 – Test GPD – multi données – version bootstrap paramétrique simplifié

```
z:\extremes\Temp\gamma1000\  
Tous  
40  
z:\extremes\Temp\gamma1000\Result10.txt  
Lognormale  
20  
0.05  
0  
0  
3  
200  
0.01  
HillG
```

**Tâche 42 – Test GPD – 1 seul nombre d'excès – version bootstrap
paramétrique**

```
z:\extremes\Temp\gamma\  
z:\extremes\Temp\gamma\Donnees.txt  
42  
z:\extremes\Temp\gamma\Result41.txt  
Weibull  
Gamma  
20  
0.05  
0  
0  
2  
500  
0.01  
Hill
```

**Tâche 43 – Test GPD – 1 seul nombre d’excès – version bootstrap
paramétrique**

```
z:\extremes\Temp\gamma\  
z:\extremes\Temp\gamma\Donnees.txt  
43  
z:\extremes\Temp\gamma\Result42.txt  
Weibull  
Gamma  
20  
0.05  
0  
0  
3  
500  
0.01  
Hill
```

Tâche 45 – Test GPD – Excès multiples – version bootstrap paramétrique

```
z:\extremes\Temp\gamma\  
z:\extremes\Temp\gamma\Donnees.txt  
45  
z:\extremes\Temp\gamma\Result43.txt  
Weibull  
Gamma  
0  
0.05  
10  
30  
2  
500  
0.01  
Hill
```

Tâche 46 – Test GPD – Excès multiples – version bootstrap paramétrique simplifié

```
z:\extremes\Temp\gamma\  
z:\extremes\Temp\gamma\Donnees.txt  
46  
z:\extremes\Temp\gamma\Result44.txt  
Weibull  
Gamma  
0  
0.05  
10  
20  
3  
500  
0.001  
Hill
```

ANNEXE D – Fichiers de sortie

- Tâche 1 – Simulation

Si l'utilisateur ne veut qu'un jeu de données, un fichier *Donnees.txt* sera créé et comportera le résultat de la simulation. Si plusieurs jeux de données sont souhaités, des fichiers *Donneesxx.txt* seront générés (où *xx* est un nombre allant de 2 au nombre de jeux de données voulus) en plus du fichier *Donnees.txt*. Parallèlement, dans le fichier résultat on trouve :

- § La moyenne théorique
- § La moyenne estimée
- § La variance théorique
- § La variance estimée
- § Le 1^{er} quartile
- § Le 3^{ème} quartile
- § La médiane

- Tâche 2, 3, 4, 8 – Densité, fonction de répartition, fonction quantile et fonction de survie

Le fichier résultat comporte 2 colonnes : dans la 1^{ère} on trouve les abscisses et dans la 2^{ème} la valeur de la fonction pour ces points.

- Tâche 5 – Histogramme

Le fichier résultat comporte les mêmes champs que lors de l'exécution de la tâche 1.

- Tâche 6 – Estimation de la densité

Le fichier résultat comporte les mêmes champs que lors de l'exécution de la tâche 2.

- Tâche 7 – Estimation des paramètres

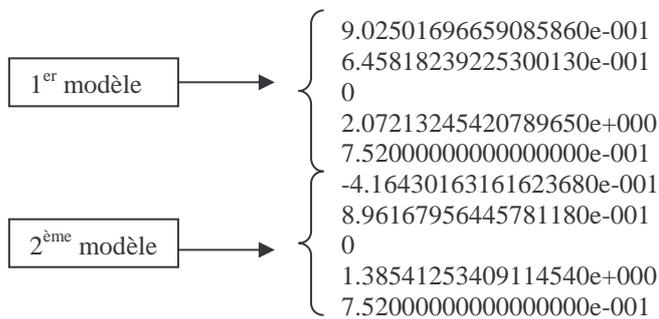
Le fichier résultat contient l'estimation des paramètres du modèle choisi suivant le paramétrage de l'annexe A.

- Tâche 9, 10 – Test d’Anderson-Darling et Cramer-Von Mises

Pour chaque modèle testé (suivant l’ordre du fichier d’entrée), il est écrit dans le fichier :

- § Les estimations des paramètres du modèle (0 si pas d’estimation)
- § La décision (0 ou 1)
- § La statistique de test (0 si statistique non calculée pour rejet flagrant)
- § La valeur de rejet (0 si non calculée pour rejet flagrant)

Exemple :



A la fin du fichier, les indices des modèles dont les densités ont pu être calculées sont listés. Le dernier champ est un nombre entier indiquant pour combien de modèle les densités n’ont pu être calculées.

- Tâche 11, 12, 14, 16, 17, 18, 38, 39, 40 – Tests d’Anderson-Darling, de Cramer-Von Mises, d’exponentialité des excès, ET et GPD – multi données

Le fichier résultat contient respectivement :

- § Le pourcentage de rejet
- § Le pourcentage d’acceptation
- § Le pourcentage d’écart type

- Tâche 13 – Test d’exponentialité des excès – QQ-Plot

Le fichier résultat est un fichier à 2 colonnes :

- § La 1^{ère} contient le calcul du quantile de la loi $Exp(1)$ pour le QQ-Plot
- § La 2^{ème} contient la valeur des excès correspondant

- § L'avant-dernier champ de la 1^{ère} colonne recèle la valeur de l'estimation du paramètre de la loi exponentielle que suivent les excès
- § Le dernier champ de la 1^{ère} colonne est 1 si le test d'exponentialité est accepté, 0 sinon
- § Les 2 derniers champs de la 2^{ème} colonne sont factices

- Tâche 15 – Test d'exponentialité des excès – Excès multiples

Le fichier contient les 2 colonnes suivantes :

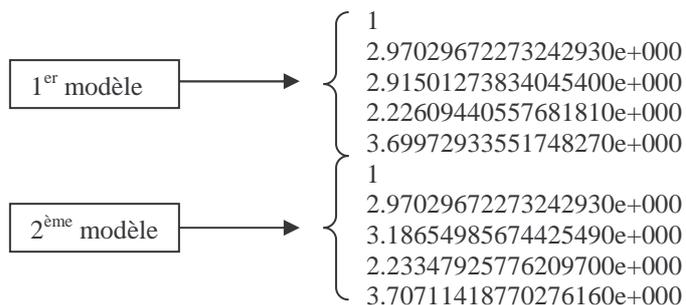
- § Les nombres d'excès
- § Le résultat du test correspondant (1 si accepté, 0 sinon)

- Tache 19, 20, 21, 41, 42, 43 – Test ET ou Test GPD– 1 seul nombre d'excès

Pour chaque modèle testé, on a :

- § Le résultat du test (1 si accepté, 0 sinon)
- § L'estimation du quantile ET (ou GPD) q_{ET} (q_{GPD})
- § L'estimation du quantile paramétrique q_{param}
- § La borne inférieure de l'intervalle de confiance
- § La borne supérieure de l'intervalle de confiance

Exemple :



- Tâche 22, 23, 24, 44, 45, 46 – Test ET ou Test GPD– Excès multiples

Les 2 colonnes de ce fichier contiennent :

-
-
- § 1^{ère} colonne : les nombres d'excès répétés pour chaque modèle testé
 - § 2^{ème} colonne : la valeur du test (1 si accepté, 0 sinon)

- Tâche 25 – Régularisation bayésienne

Le fichier résultat contient les champs ci-dessous :

- § Borne inférieure de l'intervalle de variation du paramètre
- § Borne supérieure de l'intervalle de variation du paramètre
- § Borne inférieure de l'intervalle de variation du paramètre pour le modèle Weibull paramètre de forme (0 sinon)
- § Borne supérieure de l'intervalle de variation du paramètre pour le modèle Weibull paramètre de forme (0 sinon)
- § Les paramètres de la loi a priori
- § Les paramètres de la loi prédictive
- § La valeur du test central (0 ou 1)
- § La valeur de la distance de Cramer-Von Mises pour le modèle
- § La valeur de rejet de la distance de Cramer-Von Mises
- § La valeur de la distance de Cramer-Von Mises pour le modèle prédictif
- § La valeur du test ET (0 ou 1)
- § La valeur du test ET pour le modèle prédictif (0 ou 1)
- § L'estimation du quantile ET prédictif
- § La borne inférieure de l'intervalle de confiance
- § La borne supérieure de l'intervalle de confiance
- § L'estimation des quantiles du modèle
- § L'estimation des quantiles ET
- § L'estimation des quantiles prédictifs
- § La valeur des quantiles de la vraie loi de simulation (sinon rien)
- § L'estimation des quantiles de la vraie loi dont on a estimé les paramètres (sinon rien)

- Tâche 26 – Estimation des paramètres – multi données

Les résultats sont sur une seule colonne dans le fichier. On y trouve :

- § La liste des estimateurs du 1^{er} paramètre estimé du modèle
- § La liste des estimateurs du 2^{ème} paramètre estimé du modèle (s'il y a lieu)
- § La liste des estimateurs du 3^{ème} paramètre estimé du modèle (s'il y a lieu)
- § La moyenne et la variance estimées de chacune des listes des estimateurs
- § La taille des listes des estimateurs

- Tâche 27 – Simulation selon la loi prédictive

Les résultats sont semblables à la tâche 1. La moyenne et la variance théoriques ne sont pas définies et sont égales à -1 dans le fichier résultat.

- Tâche 28 – Estimation de (γ, σ) paramètres GPD – 1 seule méthode d'estimation – 1 seul nombre d'excès – QQ-Plot

Le fichier résultat contient 2 colonnes. Sur la 1^{ère} ligne, on trouve respectivement les estimation de γ et σ . Puis, les excès sont listés en ordonnées et on trouve la valeur correspondante du QQ-plot en abscisse. Une estimation de γ est ensuite proposée par la pente d'une droite dont les points sont ajoutés aux 2 colonnes.

- Tâche 29 - Estimation de (γ, σ) paramètres GPD – plusieurs méthodes d'estimation – 1 seul nombre d'excès

2 colonnes sont présentes dans ce fichier résultat :

- § 1^{ère} colonne : les estimations de γ pour chacune des méthodes (-10 quand estimation impossible)
- § 2^{ème} colonne : les estimations de σ correspondantes (-10 quand estimation impossible)

- Tâche 30 – Estimation de (γ, σ) paramètres GPD – plusieurs méthodes d'estimation – Excès multiples

Pour chacune des méthodes on trouve : les nombres d'excès dans la colonne de gauche et les estimation de γ dans la colonne de droite. Ensuite, de nouveau pour chacune des méthodes on a : les nombres d'excès à gauche et les estimations de σ à droite.

- Tâche 31 – Estimation de (γ, σ) paramètres GPD – multi données

Résultats identiques à la tâche 26.

- Tâche 32 – Estimation d'un quantile extrême – 1 seule méthode d'estimation – 1 seul nombre d'excès

Dans les 2 colonnes du fichier résultats on trouve :

- § L'estimation des paramètres de la loi GPD en ligne
- § Dessous, 0 est présent dans la colonne de gauche et à droite on a l'estimation du quantile extrême

- Tâche 33 – Estimation d'un quantile extrême – Plusieurs méthodes d'estimation – 1 seul nombre d'excès

Dans les 2 colonnes du fichier résultats on trouve :

- § Les estimations des paramètres de la loi GPD en ligne et pour chacune des méthodes
- § Dessous, des 0 sont présents dans la colonne de gauche et à droite on a les estimations des quantiles extrêmes pour chaque méthode.

- Tâche 34 – Estimation d'un quantile extrême – Plusieurs méthodes d'estimation – Excès multiples

Pour chacune des méthodes d'estimation, on a :

- § 1^{ère} colonne : les nombres d'excès
- § 2^{ème} colonne : l'estimation du quantile

- Tâche 35 – Estimation d'un quantile extrême – multi données

- § La liste des quantiles estimés
- § La moyenne des estimations des quantiles
- § La variance des estimations des quantiles
- § La taille de la liste des estimations

- Tâche 36 – Estimation d'un quantile paramétrique

Les estimations du quantiles sont écrites en colonne dans l'ordre de départ des modèles, c'est-à-dire l'ordre dans lequel ils sont notés dans le fichier d'entrée. On trouve la valeur -10 si le quantile n'a pu être estimé.

- Tâche 37 – Estimation d'un quantile paramétrique – multi données

Résultats identiques à la tâche 35.