

## Ajustement de lois

Loi	Expression	Moyenne	Variance
Poisson	$P(\lambda)(k) = \frac{\lambda^k}{k!} e^{-\lambda}$	$\lambda$	$\lambda$
binômiale négative	$Neg(r, p)(k) = \frac{\Gamma(k+r)}{k! \Gamma(r)} p^r (1-p)^k$	$k(1-p)/p$	$k(1-p)/p^2$
log-normale	$LogN(\mu, \sigma)(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-[\ln(x) - \mu]^2 / (2\sigma^2)\right)$	$e^{\mu + \sigma^2/2}$	$(e^{\sigma^2} - 1)e^{2\mu + \sigma^2}$
Gamma	$G(\lambda, \kappa)(x) = \frac{\lambda^\kappa}{\Gamma(\kappa)} x^{\kappa-1} e^{-\lambda x}$	$k/\lambda$	$k/\lambda^2$
normale	$N(\mu, \sigma)(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	$\mu$	$\sigma^2$
exponentielle	$E(\lambda)(x) = \lambda e^{-\lambda x}$	$1/\lambda$	$1/\lambda^2$
valeurs extrêmes	$P(\mu, \sigma, \xi)(x) = \exp\left[-\left(1 + \xi \cdot \frac{x-\mu}{\sigma}\right)^{1/\xi}\right]$	-	-

Tableau 1 – Quelques lois de probabilités.

### 1 Méthode des moments (*rappel séance 1*)

L'ajustement le plus simple est de chercher la loi ayant les mêmes moments que les données. Les moments des données sont :

moyenne (1 <sup>er</sup> moment)	$E = \int X dP = \frac{1}{N} \sum x_i$
variance (2 <sup>e</sup> moment)	$\sigma^2 = \int (X - \mu)^2 dP = \frac{1}{N} \sum (x_i - \mu)^2$
⋮	⋮ ⋮

Le nombre de paramètres de la loi donne le nombre de moments à égaliser. Trois séries de données (*serie01.txt*, *serie02.txt* et *serie03.txt*) ont été obtenues lors d'expériences et on cherche à leur affecter l'une des lois de probabilité suivante : loi log-normale, loi binômiale négative et loi de Poisson.

- Donner les principales caractéristiques de ces trois lois (on pourra s'aider d'internet), à savoir le nombre de paramètres, la moyenne et la variance. Quelles sont leurs particularités respectives ?
- Calculer les densités de probabilité (pdf) de ces trois séries et tracer les avec leurs pdf théoriques ajustées à l'aide de la méthode des moments. [**hist**, **lognpdf**, **poisspdf**, **nbinpdf**]
- Pour chacune de ces séries, tracer un diagramme quantile expérimental/quantile théorique ainsi que la droite  $y = x$ . [**ecdf**, **logninv**, **poissinv**, **nbininv**]
- Que dire de ce diagramme pour la série 3 ?

### 2 Méthode du maximum de vraisemblance

Nous allons dans cette section appliquer la méthode du maximum de vraisemblance aux données *serie-gamma.txt*. Nous voulons ici caler une loi Gamma à deux paramètres.

1. Pour commencer nous allons créer une fonction matlab qui calcule l'opposée de la log-vraisemblance  $l(\kappa, \lambda)$  lorsqu'on lui donne les paramètres  $(\kappa, \lambda)$  et les données  $x_i$ . La vraisemblance est définie de la manière suivante :

$$L(\kappa, \lambda) = \prod_{i=1}^N G_a[\lambda, \kappa](x_i).$$

- Calculer **analytiquement** la fonction  $l(\kappa, \lambda, x_i) = -\ln(L)$  et simplifier au maximum son expression.
- Pour créer une fonction dans matlab, il faut créer un fichier *mafonction.m*. (cf premier TD) La syntaxe est ensuite du type :

```
function lvrai = mafonction(l,k,x)
    lvrai = - (l * k * x *... ; % à modifier...
end
```

Une fois cette fonction créée, la commande matlab `mafonction(l,k,x)` vous donne l'opposé de la valeur de la log-vraisemblance correspondant au triplet  $(l, k, x)$ .

— Modifier l'exemple ci-dessus avec l'expression analytique trouvée.

2. Nous allons ensuite chercher le minimum de la fonction créée avec la fonction :

```
k=fminbnd(@k mafonction(l,k,x),inf,sup)
```

Ici, on cherche la valeur de  $k$  qui minimise *mafonction*. *inf* est la borne inférieure de recherche et *sup* la borne supérieure. Comme notre fonction a deux paramètres, nous allons effectuer une recherche par itérations. L'algorithme est le suivant :

— fixer un paramètre  $l$  arbitrairement ;

— pour  $i=1\dots 10$  :

— calculer  $k$  avec la fonction `fminbnd` ,

— calculer  $l$  avec la fonction `fminbnd` en prenant  $k$  trouvé à l'étape précédente.

— fin de la boucle.

3. Coder cet algorithme. L'algorithme converge t'il ? Tracer la densité de probabilité de la loi ajustée, et la comparer à celle obtenue grâce à la méthode des moments.

### 3 Ajustement de lois de valeurs extrêmes sur des données

Nous allons maintenant faire un ajustement de loi de valeurs extrêmes sur les données de températures et de précipitations à Davos. Il s'agit de déterminer le type de loi de valeurs extrêmes (Weibull, Gumbel ou Fréchet) qui décrit le mieux le comportement des maxima de chaque série. Dans cette exercice, nous utiliserons directement la fonction `gevfit` qui estime les paramètres de la loi généralisée des valeurs extrêmes par la méthode du maximum de vraisemblance vue dans l'exercice précédent. Nous procédons de la manière suivante :

— définir des blocs (classiquement des années) ;

— calculer le maximum de chaque bloc (donc par exemple le maximum annuel) ;

— utiliser ces maxima pour caler la loi des valeurs extrêmes ;

— tracez cette loi et comparer avec la courbe empirique.

1. Précipitations

De quel type (Gumbel, Fréchet ou Weibull) est la distribution ?

2. Températures

Le fichier possède plusieurs colonnes de données, les trois dernières colonnes représentent la température à 7 h, 13 h et 21 h. Choisissez-en une et trouver la loi de valeurs extrêmes qui convient le mieux pour décrire la distribution des maxima.