# TP 6 - Méthode du maximum de vraisemblance - Agrégation externe de mathématiques - 2005/2006.

## 1 Maximum de vraisemblance et lois de Bernouilli

On lance une pièce mais des petites asymétries influencent la probabilité d'obtenir une face, qui n'est plus forcément 1/2.
On désigne $Y_1, \dots Y_n$ les résultats d'essais indépendants et l'on note

$$P(Y_j = 1) = \theta \qquad P(Y_j = 0) = 1 - \theta \qquad 0 \leq \theta \leq 1$$

1. Programmer une fonction Matlab qui simule une suite de lancés de pièce biaisée de paramètre $\theta$.
2. Calculer la vraisemblance ainsi que la log-vraisemblance du modèle.
3. Calculer l'estimateur du maximum de vraisemblance.
4. Utiliser cet estimateur pour calculer dans ce cas $\hat{\theta}_n$.
5. Dessiner les vraisemblances obtenues en fonction de $\theta$, pour différentes simulations de $n$ pièces.
6. On a en réalité le théorème sur l'efficacité du maximum de vraisemblance :
   « Soient $Y_1, \dots Y_n$ un échantillon aléatoire issu d'une densité paramétrique $f(y, \theta)$, et soit $\hat{\theta}$ l'EMV de $\theta$. Si $f$ satisfait des conditions de régularité, alors

   $$J(\hat{\theta})^{1/2}(\hat{\theta} - \theta) \longmapsto \mathcal{N}(0, 1)$$

   Donc pour $n$ grand, $\hat{\theta} \sim \mathcal{N}(\theta, J(\hat{\theta}^{-1})$. »Pour des valeurs de $n$ petites, certains estimateurs ont une variance inférieure. En revanche, lorsque $n$ tend vers $\infty$, l'EMV est à variance minimale (borne de Cramer-Rao), il est donc optimal en ce sens.

## 2 Maximum de Vraisemblance et loi de Pareto

On rappelle qu'une variable aléatoire $X$ suit une loi de Pareto de paramètres $\alpha$ et $\theta$ si la densité s'écrit de la forme

$$f_{X,\alpha,\theta}(t) = ke^{-\alpha t}\chi_{t \geq \theta}$$

### 2.1 Généralités

1. Montrer que la constante $k$ de normalisation est donnée par

   $$k = (\alpha - 1)\theta^{\alpha - 1}$$

2. Montrer que si $X$ suit une loi de Pareto $(\theta, \alpha)$, alors

   $$\mathbb{E}(X) = \frac{\alpha - 1}{\alpha - 2}\theta$$

   et

   $$V(X) = \frac{(\alpha - 1)\theta^2}{(\alpha - 3)(\alpha - 2)^2}$$

3. Montrer que la fonction de répartition $F_{X,\theta,\alpha}$ est donnée par

$$F_{X,\,\theta,\,\alpha}(t) = 1 - (\theta/t)^{\alpha-1}$$

4. Démontrer que si l'on effectue le changement de variable

$$U = (\alpha - 1)\ln\left(\frac{X}{\theta}\right)$$

alors $U$ suit une loi exponentielle.

5. Écrire une fonction matlab pareto.m qui simule la loi de Pareto avec comme argument $(\theta, \alpha)$. Écrire une fonction echantillon-pareto.m qui simule un $n$-échantillon de loi de Pareto. Tracer l'histogramme empirique obtenu.
6. On note $Z = \min(X_1, \ldots, X_n)$ où $(X_1, \ldots, X_n)$ est un $n$-échantillon de loi $P(\theta, \alpha)$. Déterminer la loi de $Z$.
7. Programmer la fonction min-pareto.m qui simule la variable $Z$ en fonction de l'entier $n$ taille de l'échantillon, de $\theta$ et $\alpha$. Comparer l'histogramme empirique obtenu ici avec l'histogramme empirique construit à partir de $p$ simulations de $n$-échantillons $(X_1, \ldots, X_n)$ de loi de Pareto $\theta, \alpha$.

## 2.2 Étude à $\theta$ connu

On suppose le paramètre $\theta$ connu, on cherche à estimer $\alpha$.

1. Donner une statistique exhaustive pour $\alpha$.
2. Estimer $\alpha$ par la méthode du maximum de vraisemblance. On notera $\hat{\alpha}$ l'EMV de $\alpha$.
3. Écrire une fonction emv-alpha.m qui prend un argument un $n$-échantillon et qui donne l'EMV de $\alpha$.
4. $\hat{\alpha}$ est-il sans biais ? Si non, construire un estimateur sans biais de $\alpha$ noté $\hat{\alpha}_1$. Modifier alors la fonction matlab précédente pour en déduire un ESB de $\alpha$.
5. L'estimateur $\hat{\alpha}_1$ est-il efficace, asymptotiquement efficace ?
6. Construire l'histogramme empirique obtenu à partir de $p$ simulations de $n$-échantillons. Comparer l'histogramme avec la borne de Cramer-Rao.

## 2.3 Étude à $\alpha$ connu

On suppose le paramètre $\alpha$ connu, on cherche à estimer $\theta$.

1. Donner une statistique exhaustive de $\theta$.
2. Estimer $\theta$ par la méthode du maximum de vraisemblance. On notera $\hat{\theta}$ l'EMV de $\theta$.
3. Écrire une fonction emv-theta.m qui prend un argument un $n$-échantillon et qui donne l'EMV de $\theta$.
4. Donner la loi exacte de $\hat{\theta}$.
5. Calculer l'information de Fisher $I_X(\theta)$ apportée par $X$ sur $\theta$. En déduire $I_n(\theta)$.
6. Calculer $\mathbb{E}(\hat{\theta})$. En déduire un estimateur sans biais $\tilde{\theta}$ de $\theta$.
7. $\hat{\theta}$ est-il un estimateur efficace, asymptotiquement efficace de $\theta$ ?
8. Construire l'histogramme empirique obtenu à partir de $p$ simulations de $n$-échantillons. Comparer l'histogramme avec la borne de Cramer-Rao.