

# TP1 : Ajustement par maximum de vraisemblance

marie-anne.poursat@universite-paris-saclay.fr

23 janvier 2024

## Objectifs du TP :

- tracer le graphe de la log-vraisemblance d'un échantillon;
- estimer une densité de probabilité par maximum de vraisemblance;
- interpréter un qqplot pour valider (graphiquement) un modèle.

## 1 Vraisemblance d'un échantillon de Bernoulli

1. Ecrire une fonction R qui calcule la log-vraisemblance d'un échantillon de Bernoulli en une valeur du paramètre  $p$ , en utilisant l'expression développée en cours puis en utilisant la fonction `dbinom`. Testez votre fonction en quelques valeurs du paramètre, puis en un vecteur de quelques points. Quel est le problème?

`Vectorize` transforme une fonction  $g$  dont le paramètre d'entrée est un scalaire en une *fonction vectorielle*, c'est-à-dire une fonction qui évalue  $g$  en chaque point d'un vecteur d'entrée:

```
f=function(x,y) x+y
x=1; y=1:10
f(x,y) # ajoute le scalaire 1 au vecteur 1:10
x=rep(1,2)
f(x,y) # ne s'applique pas à un vecteur x
vf=Vectorize(f,"x") # vectorise la fonction en x
vf(x,y) # tableau croisé des (x[i],y[j])
```

2. Générer un échantillon de Bernoulli de taille  $n = 20$  et représenter sa "densité" (fonction de probabilité).
3. Tracer le graphe de la log-vraisemblance associée à l'échantillon; renseigner les axes et ajouter un titre. Quel est la valeur de l'EMV?
4. Générer un nouvel échantillon et superposer au graphe précédent le tracé de la log-vraisemblance du nouvel échantillon. Que mettez-vous en évidence?

## 2 Ajustement de données

Le fichier `duree_de_vie.txt` contient les durées de vie de  $n = 20$  composants en milliers d'heures. L'import et l'export de fichiers de données sous forme de fichier texte peuvent se faire en appelant les fonctions `read.table()` (import dans un data.frame d'un fichier texte) et `write.table()` (export dans un data.frame d'un fichier texte).

1. Consultez le help de la fonction `read.table()` et attribuer la bonne valeur aux arguments `skip=` et `header=` pour importer le fichier texte `duree_de_vie.txt` dans un `data.frame`. Extraire la variable `times`. Vérifiez que votre échantillon est de taille  $n = 20$ . Calculer la moyenne et la variance empiriques de l'échantillon.
2. Représentez un histogramme normalisé en densité de l'échantillon `times`. Quel modèle de loi générative pouvez-vous proposer?

En fiabilité, la durée de vie d'un composant électronique est souvent modélisée par une loi de Weibull (suffixe `weibull` dans R). Une variable aléatoire distribuée selon une loi de Weibull de paramètre  $a > 0$  a pour densité:

$$f_a(x) := a x^{a-1} \exp(-x^a), x > 0.$$

3. Ecrire une fonction `logL` qui calcule la log-vraisemblance de l'échantillon au point  $a$ . Testez votre fonction.
4. Représentez graphiquement la fonction de vraisemblance sur l'intervalle  $[0.17, 6]$ .
5. Le calcul de l'estimateur du maximum de vraisemblance (EMV) se fait en trouvant les racines de la fonction suivante,

$$(\log L)'(a) = \frac{n}{a} + \sum_{i=1}^n X_i - \sum_{i=1}^n \log(X_i) X_i^a$$

L'équation de vraisemblance  $(\log L)'(a) = 0$  admet une racine mais elle n'a pas d'expression analytique. On propose donc de la calculer en utilisant un algorithme d'optimisation de R (par exemple `optimize`). Consultez le help de la fonction `optimize`.

Calculez l'EMV de votre échantillon. Vérifiez qu'il réalise bien le maximum du graphe de la log-vraisemblance.

6. Superposez le graphe de la densité estimée à l'histogramme de l'échantillon.

### 3 Q-Q plot

Lorsque l'on compare deux distributions (celles de deux échantillons, ou bien la distribution d'un échantillon avec une distribution théorique ou estimée), les graphiques des histogrammes ne sont pas (visuellement) informatifs.

Le graphe quantile/quantile ou Q-Q plot est un outil pour vérifier visuellement l'adéquation d'un jeu de données à une loi. Il représente les **quantiles théoriques** du modèle, associés à la loi estimée, en fonction des **quantiles empiriques**, observés sur l'échantillon. Si la loi théorique est pertinente, les points s'alignent 'à peu près, pas exactement!) suivant la première diagonale.

En pratique, il y a plusieurs façons de calculer les quantiles empiriques de l'échantillon observé. L'implémentation peut différer légèrement selon les logiciels, qui prévoient d'ailleurs plusieurs types de calcul (argument `type` de la fonction `quantile` de R).

7. Comprendre les fonctions R `ppoints` et `quantile`.
8. Retrouver le graphique suivant avec la fonction `qqplot`  
Que pensez-vous de l'ajustement obtenu par de la loi de Weibull sur ces données?

```
n=length(times)
plot(qweibull(ppoints(n),shape=EMV),quantile(times,ppoints(n)),col="blue",
     xlab="quantile théorique", ylab="quantile empirique")
abline(0,1)
```