Séance 4 : Régression Non Linéaire avec R

Exercice 1 *Définition d'une fonction sous R*

1. *Définir la fonction $f1$ de trois variables (t, a, b) par*

```
f1 <- function(t,a,b){# paramètres d'entrée: t,a,b
out=a*exp(b*t)
return(out)}#renvoie la valeur de out
```

et la tester en calculant $f1(0, 1, 2)$ puis tracer le graphe de $f1(t, 1, 2)$ pour t allant de 0 à 5

```
f1( 0,1,2)# calcul de f(0,1,2)
curve(f1(x,1,2),from=0,to=5)# curve: attention la variable est toujours x
```

2. *Si (a, b) sont des paramètres, on peut aussi les regrouper dans un vecteur indexé par des entiers et définir f comme fonction de (t, par). Définir f comme suit*

```
f <- function(t,par){# paramètres d'entrée: t (un réel) et par (vecteur)
out=par[1]*exp(par[2]*t) # par contient deux valeurs
return(out)}
par=c(1,2)# initialisation du vecteur par
f(0,par)
```

Remarque: si (a, b) sont des paramètres, on peut aussi les regrouper dans un vecteur indexé par des noms et définir f comme fonction de (t, par) (ne pas le faire).

```
f3 <- function(t,par){
out=par["a"]*exp(par["b"]*t)
return(out)}
par=c(a=1,b=2)# initialisation du vecteur par
```

Exercice 2 *Rendement d'un procédé batch dans un fermenteur et modélisation de l'évolution de la biomasse et du substrat*
Après différentes calibrations, on a obtenu les concentrations contenues dans le fichier chemostat.csv (à télécharger).

On veut modéliser l'évolution des concentrations x et s sous les conditions "batch" par les équations différentielles suivantes :

$$\begin{cases} \dfrac{dx}{dt} &= \mu x & x(0) = x_0 \\ \dfrac{ds}{dt} &= -\dfrac{\mu}{Y_{x/s}} x & s(0) = s_0 \end{cases}$$

où μ le taux de croissance de la biomasse, et $Y_{x/s}$ le rendement instantané sont supposés constants au cours du temps et $(x_0,\ s_0)$ sont les concentrations initiales (des constantes strictement positives).

1. *Détermination du rendement $Y_{x/s}$*

   (a) *Lire le fichier chemostat.csv dans la data frame donnees.Tracer le nuage de points expérimentaux (s_i, x_i).*

   (b) *Montrer (à la main!) que pour tout t, $x(t) + Y_{x/s}s(t) = x_m$ où x_m est la constante égale à $x_m = x_0 + Y_{x/s}s_0$. On a donc $x = -Y_{x/s}s + x_m$ (une droite).*

   (c) *Déterminer, à partir des données expérimentales, les estimations de $Y_{x/s}$ et x_m et leur écart-type. Tracer la droite de régression sur le nuage des points expérimentaux. Afficher son équation sur le graphe.*

   (d) *Vérifier les résidus de la régression pour valider ou non la régression.*

2. **Modèle exponentiel : sélection des données correspondant au modèle**

   (a) *Tracer le nuage de points (t_i, x_i).*

   (b) *On voit que la croissance s'infléchit en fin d'expérience, on peut donc supposer que la phase exponentielle se termine avant. Tracer le nuage de points $(t_i, \ln(x_i))$ pour sélectionner les points (t_i, x_i) qui "suivent" le modèle exponentiel.*

(c) *On décide de sélectionner les 8 premiers points pour déterminer les paramètres de croissance de la phase exponentielle (ce choix est subjectif). Créer une nouvelle data frame nommée donnees2 où les deux dernières lignes de donnees ont été supprimées. Déterminer une initialisation pour les paramètres* (μ, x_0).

(d) *Déterminer le taux de croissance* μ *à partir des données2* (t_i, x_i). **On ne travaillera qu'avec les points de donnees2** *(on élimine les points qui sortent du domaine de validité du modèle exponentiel). Vérifier graphiquement le résultat.*

```
# appel de nls:  construire la suite (a_n,b_n)
# start contient valeurs initiales des parametres (a_0,b_0)=(x0,mu) déterminés avant
# formule xi~f(ti,c(a,b)) à chaque iteration a et b seront remplacés par (a_n,b_n)
initialisation=list(a=x0,b=mu)
modele <- nls(xi~ f(ti,c(a,b)), data = donnees2, start = initialisation)
summary(modele)
x0=coef(modele)[1];mu=coef(modele)[2]
eq = paste0("Equation: x = (", round(x0,4)," )*exp(", round(mu,4),"*t)" )# sous titre equation
titre=paste0("Détermination taux de croissance de la phase expo. mu=",round(mu,4))#titre
plot(ti,xi,main=titre,sub=eq,xlab="temps (h)",ylab="biomasse (g/L)")#tracé du nuage de points
curve(f(x,c(x0,mu)),from=0,to=8,col="red" ,lwd = 2,add=TRUE)# ajout courbe prédite
```

Il faudrait contrôler les résidus (cf séance précédente).

3. **Extension du modèle: croissance logistique (prise en compte de toutes les données)** *On étudie toujours la croissance dans un chemostat en batch mais la modélisation qui suit vise à élargir le domaine de validité du modèle.*

On conserve le système d'équations différentielles mais on suppose maintenant que μ *le taux de croissance de la biomasse suit une fonction linéaire* $\mu(s) = \lambda s(t)$ *où* λ *est une constante positive. On suppose aussi à l'instant* $t^* > 0$ *le substrat est épuisé c'est-à-dire* $s(t^*) = 0$ *et on pose* $x(t^*) = x_m$. *On a toujours* $x_m - x(t) = Y_{x/s}s(t)$. *On conservera donc les valeurs des paramètres* $Y_{x/s}$ *et* x_m *obtenues dans la modélisation précédente.*

(a) *Montrer que* x *vérifie l'équation différentielle*

$$\frac{dx}{dt} = k(1 - \frac{x}{x_m})x$$

où $k = \frac{\lambda x_m}{Y_{x/s}}$.

(b) *Montrer (à la main) que la solution de cette équation différentielle (en posant* $z = \frac{1}{x}$*) est*

$$x(t) = \frac{x_m x_0}{x_0 + e^{-kt}(x_m - x_0)}$$

(c) *Définir la fonction* $g(t, par) = \frac{x_m par[1]}{par[1]+e^{-par[2]t}(x_m-par[1])}$. *Déterminer avec nls les paramètres* k, x_0 *(*x_m *fixé à la valeur déterminée précédemment 1.8285).*

(d) *Tracer le nuage de points et la courbe théorique obtenue.*
(Remarque: Il faudrait contrôler les résidus).

(e) *Améliorations possibles: ajouter* x_m *comme paramètre ? Peut-être faut-il ajouter des bornes sur les paramètres ? (cf Help: nls)*

Exercice 3 *On introduit maintenant des bactéries dans le fermenteur. La concentration en oxygène dissous dans un fermenteur en régime transitoire est donnée par l'équation*

$$\frac{dO_{2L}}{dt} = Kla(O_{2L}^* - O_{2L}) - Q_{O_2}X \quad O_{2L}(0) = O_{2L_0}$$

où O_{2L}^* *est la concentration saturante d'oygène liquide à 30øC,* X *la concentration bactérienne du milieu, et* Q_{O_2} *la vitesse spécifique de respiration des bactéries. On suppose que* $O_{2L}^* = 100$*. A l'aide de données expérimentales ci-dessous, on veut déterminer* Kla*,* $Q_{O_2}X$ *et* O_{2L_0}*.*

t_i *temps (s)*	*10*	*20*	*30*	*40*	*50*	*60*	*70*	*100*	*130*
O_{2L_i} *(mgl^{-1})*	*43,5*	*53,5*	*60*	*67,5*	*70,5*	*72*	*73*	*73,5*	*73,5*

1. *Tracer le nuage de points expérimentaux* (t_i, O_{2L_i})*.*
2. *Résoudre l'équation différentielle vérifiée par* O_{2L}*. On montrera que* $O_{2L}(t) = (O_{2L}(0) - \overline{O_{2L}})e^{-Klat} + \overline{O_{2L}}$ *où* $\overline{O_{2L}} = O_{2L}^* - \dfrac{Q_{O_2}X}{Kla}$*.*
3. *Déterminer les paramètres.*
4. *Tracer la courbe théorique obtenue et tracer les résidus dans un autre graphique.*