%newcommand10

Contrôle des connaissances du cours "Méthodes de traitement des données" du DEA de Strasbourg.

Corrigé

Le vendredi 11 février 2000 de 16h à 18h à l'Observatoire de Strasbourg.

Le contrôle est noté sur 20, le barème est indiqué à la fin de chaque question. Le polycopié du cours et les notes de cours sont les seuls documents autorisés.

Probabilités

On joue à pile ou face n fois, et on considère la variable aléatoire K obtenue en sommant le nombre de pile.

– Quelle est la moyenne de K
  [0.5 point]
  Réponse : $n/2$
– Quel est son écart type ?
  [0.5 point]
  Réponse : $n^{1/2}/2$
– Si n est grand, donner une approximation de la fonction de répartition de K
  [1 point]
  Réponse : c'est la fonction de répartition de la loi normale de moyenne $n/2$ et d'écart type $n^{1/2}/2$

Détecteurs de photons

Le système de d'etection d'un instrument est un détecteur à comptage de photon ; il tombe souvent en panne et on est obligé de remplacer par un détecteur de rechange, moins performant. Alors que le détecteur normal D1 a un rendement de 65 %, celui de rechange D2 a un rendement de 30 %. Une moyenne faite sur l'année indique un rendement moyen de 50 %

– Quelle est la probabilité de tomber sur le bon détecteur un jour donné ?
  [1 point]
  Réponse : La probabilité p recherchée satisfait :

$$50 = 65p + 30(1 - p)$$

  Ce qui donne $p = 0.57$

Lors de l'observation d'une source de calibration, pour laquelle on s'attendrait à détecter 20 photons pendant la durée de l'observation si le rendement était de 100 %, on n'en trouve que 14.

– a. Donner la loi de probabilité suivie par le nombre de photons qui auraient été détectés par D1 ou par D2
[1 point]
Réponse : Le nombre de photons reçus suit une loi de Poisson de moyenne 13 pour D1 et 7 pour D2
– b. Quelle est la probabilité que ce soit le détecteur D1 qui soit monté ce jour là ?
[2 points]
Réponse : En utilisant la formule de Bayes,

$$P(D1|k=14) = \frac{P(k=14|D1)p(D1)}{P(k=14|D1)p(D1) + P(k=14|D2)p(D2)}$$

Comme $P(14|D1) = 13^{14}/14!\ e^{-13} = 0.09176$ et que $P(14|D2) = 6^{14}/14!\ e^{-6} = 0.00248$, que $p(D1) = 0.57$ et $p(D2) = 0.43$, la probabilité d'avoir utilisé le détecteur D1 est de 0.98
– Si le détecteur comptait non pas les photons, mais les interactions entre le photons et le détecteur, pourquoi la réponse à la question (a) ci-dessus aurait-elle dû être différente si on se place dans un domaine d'énergie où un photon est susceptible de subir plusieurs interactions ?
[1 point]
Réponse : alors que les photons sont détectés de façon indépendante, il n'en va pas de même pour les interactions qui sont corrélées.

Ajustement d'une courbe de lumière.

Cet exercice est inspiré du paragraphe 5 et de la figure 2 de l'article de Monnier, Tuthill & Danchi joint en annexe (*Ap. J. Letters* **525**, L99). Les auteurs ajustent une courbe de lumière à un ensemble de données expérimentales et en déduisent la valeur de certains paramètres.
– Les auteurs mentionnent que "[a] sinusoid was fit to this data", pouvez vous préciser le nombre de paramètres libres qui entrent dans leur modèle ?
[1 point]
Réponse : quatre paramètres libres.
– Donnez la forme $\mu(t;\theta)$ de la fonction représentant la courbe de lumière ainsi que la nature des paramètres θ.
[2 points]
Réponse :

$$\mu(t;\theta) = A\cos\Big(2\pi\frac{t-t_{\max}}{T}\Big) + K_0\,,$$

où $\theta_1 = A$ est l'amplitude, $\theta_2 = t_{\max}$ la date du maximum, $\theta_3 = T$ la période et $\theta_4 = K_0$ la magnitude K moyenne.
– Le modèle est-il linéaire ?
[1 point]
Non, il ne s'écrit pas $\mu(t) = \sum_{i=1}^{n}\theta_i f_i(t)$. Seuls θ_1 et θ_4 interviennent linéairement.
– Quelles sont les valeurs des θ trouvées par les auteurs ?
[1 point]

$$A = 0.65/2\,,\quad t_{\max} = 1992.59\,\mathrm{yr}\,,\quad T = 1.61\,\mathrm{yr}\,,\quad K_0 = 4.52\,.$$

– En supposant que les mesures de magnitudes sont indépendantes, que vaut la matrice des variances-covariances des observations ? (On supposera que les “uncertainties” dont parlent les auteurs sont les écarts types.)
[2 points]
Si les données sont non-corrélées et comme les erreurs sont visiblement les mêmes, on a :

$$\mathbf{V} = \sigma^2 \mathbf{I} \,,$$

où **I** est la matrice identité $(10, 10)$ et avec $\sigma = 0.05$ comme on peut le voir sur la figure.

– En supposant que la méthode d'ajustement du modèle aux données est la méthode des moindres carrés, écrire la formule qui a permis d'obtenir la valeur des paramètres.
[1 point]

$$\chi^2_{\min} = \min_{\theta} \sum_{j=1}^{10} \Big(\frac{K_j - \mu_j}{\sigma} \Big)^2 \,,$$

où les K_j sont les valeurs expérimentales et les μ_j les valeurs prédites par le modèle aux temps t_j où les K_j sont observés. On a :

$$K_j = \mu(t_j; \theta) = \theta_1 \cos\Big(2\pi \frac{t_j - \theta_2}{\theta_3} \Big) + \theta_4 \,.$$

– Que vaut la quantité $\chi^2_{\min}$ pour les paramètres trouvés par les auteurs ?
[1 point]
On peut assez facilement estimer visuellement cette valeur :

$$\begin{aligned} \chi^2_{\min} &= 1^2 + 0^2 + 1.5^2 + 2^2 + 5^2 + 3^2 + 0^2 + 0.5^2 + 0.5^2 + 0^2 \,, \\ &= 41.75 \approx 42 \,. \end{aligned}$$

– La valeur trouvée pour $\chi^2_{\min}$ permet-elle d'accepter l'ajustement ?
[2 points]
La quantité χ^2 suit approximativement une loi du χ^2 à : "nombre de données" − "nombre de paramètres libres" $= 10 - 4 = 6$ degrés de liberté. Cela serait rigoureusement exact dans le cas linéaire et normal mais on fait souvent cette approximation. Une variable aléatoire à f degrés de liberté possède une moyenne égale à f et une variance égale à $2f$. On s'attend donc ici à une moyenne de 6 et un écart type de $\sqrt{2f} \approx 3.46$. Le $\chi^2_{\min}$ observé s'écarte donc à : $\frac{42-6}{3.46} \approx 10.4$ écarts types de la moyenne attendu. C'est une valeur énorme (on dépasse tout seuil de fausse alarme raisonnable) on ne peut donc pas accepter cet ajustement. La raison de cette si mauvaise valeur est l'observation numéro 5 qui s'écarte beaucoup trop de la courbe théorique.

– Les auteurs donnent une valeur qui semble très précise à certains paramètres, voyez par exemple $t_{\max}$ qui est donné avec une précision de 3 jours environ. Comment est-il possible de savoir si cette précision est illusoire ou non ? On demande le principe pas un calcul détaillé.
[2 points]
Il faut calculer la matrice des variances-covariances des θ et prendre la racine carrée de ces éléments diagonaux. Ces valeurs donnent les barres d'erreurs dues aux erreurs statistiques, c'est à dire sans tenir compte d'éventuelles erreurs systématiques. Cette matrice s'obtient facilement dans le cas linéaire, il faut donc

au préalable linéariser le modèle au voisinage de la solution. On obtient ainsi une matrice modèle : $\mathbf{X}$ formée de 10 lignes et 4 colonnes.
La matrice $\mathbf{\Sigma}$ des variances-covariances des θ s'obtient alors par (voir cours) :

$$\mathbf{\Sigma} = \sigma^2(\mathbf{X}^t\mathbf{X})^{-1} .$$

Le programme qui suit calcule numériquement ces quantités"

```
! Calculs pour l'examen de Strasbourg
! f95 S9900.f90 -ldxml -lpgplot -lX11
Program S9900
  ! Donnees.
  Real, Dimension(10) :: t_i ! Temps.
  Real, Dimension(10) :: y_i ! Observations.
  Real :: sigma = 0.05 ! Ecart type des observations.

  ! Design matrix du model linearise':
  Double Precision, Dimension(10,4) :: X
  ! Matrice des va-co des parametres.
  Double Precision, Dimension(4,4)  :: V

  ! Parametres trouves par les auteurs
  Real, Parameter    :: A=0.65/2.0, t_max= 1992.59, T=1.61, K0=4.52
  Real, Dimension(4) :: theta=(/A,t_max,T,K0/)

  ! Variables locales.
  Real :: pi
  Real, Dimension(10) :: arg, Dy
  Integer :: i, info

  ! Pour le dessin :
  Integer, Parameter :: nplot = 1000
  Real, Dimension(nplot) :: tt, yy
  Real :: tmin, tmax, Dt
  Integer, Dimension(nplot) :: ii

  pi = 4.0*Atan(1.0)

  ! Ici on rentre les temps approximatifs.
  t_i = (/1990.2,1990.28,1991.18,1991.5,1991.55, &
       1991.63,1991.7,1992.35,1992.53,1992.61/)

  ! Et ici on entre les magnitudes approximatives (lues sur le graphe).
  y_i = (/4.8,4.83,4.2,4.55,5.0,4.66,4.83,4.35,4.24,4.21/)

  ! Calcul de la matrice du modele linearise'.
  arg(:) = 2.0*pi*(t_i(:)-t_max)/T
  X(:,1) = -Cos(arg(:))
  X(:,2) = -A*Sin(arg(:))*(2.0*pi)/T
  X(:,3) = A*Sin(arg(:))*(2.0*pi*(t_i(:)-t_max))/T**2
  X(:,4) = 1.0
  V = Matmul(Transpose(X),X)
```

```
  ! Inversion par la routine appropriee de `lapack'
  Call DPOTRI('U',4,V,4,info)
  Do i=1,3
   V(:,i) = V(i,:)
  End Do
  V = sigma**2*V
  Do i=1,4
     Write(*,'(f8.3,x,"+/-",x,f6.4)') theta(i),Sqrt(V(i,i))
  End Do

  ! A titre d'exemple on calcule les corrections sur les parametres
  ! (au sens de moindres-carres) dues au fait que nous avons pris
  ! des valeurs t_i et y_i approximatives.
  Dy(:) = y_i(:) + A*Cos(2.0*pi*(t_i(:)-t_max)/T) - K0
  theta(:) = Matmul(V,Matmul(Transpose(X),Dy))
  Write(*,*) "Corrections :"
  Write(*,*) theta

  ! Dessin
  tmin = 1990.0 ; tmax = 1993.0
  Dt = (tmax-tmin)/Real(nplot-1)
  ii(:) = (/((i-1),i=1,nplot)/)
  tt(:) = tmin + ii(:)*Dt
  yy(:) = -A*Cos(2.0*pi*(tt(:)-t_max)/T) + K0
  Call PgBegin(0,"?",1,1)
  Call PgWindow(tmin,tmax,5.2,4.0)
  Call PgSLW(2)
  Call PgBox("BCNST",1.0,10,"BCNSTV",0.2,4)
  Call PgLabel("Date (year)","K-band Magnitude"," ")
  Call PgPoint(10,t_i,y_i,17)
  Do i=1,10
     Call PgErry(1,t_i(i),y_i(i)+sigma,y_i(i)-sigma,1.0)
  End Do
  Call PgSLS(2)
  Call PgLine(nplot,tt,yy)
  Call PgEnd
End Program S9900
```

Et donne les résultats suivants :