Contrôle des connaissances du cours :
« Méthodes de Traitement des Données »
du DEA d'Astrophysique de Paris 6
Corrigé

Paris le lundi 24 février 1997 de 9h à 10h30 amphithéâtre de l'IAP.

Un choix d'exercices vous est proposé pour un total de 40 points. Naturellement le contrôle est noté sur 20, vous pouvez donc choisir les exercices que vous voulez. Le barème est indiqué en fin d'exercice. Les questions à 1 point sont relativement faciles, celles à 2 points demandent un peu plus de réflexion, ne passez pas trop de temps sur les questions à 3 points.

Enoncés des exercices :

1. **Fond du ciel.**

   On a observé un objet peu lumineux par rapport au fond du ciel. Le nombre de photons du ciel intégré dans la zone où l'on observe l'objet est de 10^4, celui de l'objet est de 100.

   ① En supposant que l'on retire le niveau exact du fond de ciel, quel est l'écart type du signal obtenu par cette soustraction ?

   ② Au lieu du niveau exact du ciel (que l'on ignore) on retire une observation du ciel telle qu'elle a été observée dans un zone très proche de l'objet. Quel est alors l'écart type du signal obtenu par cette procédure.

   (1 + 1 = 2 points)

   Corrigé.

   ❶ $\sigma \approx 100$.

   ❷ $\sigma \approx 100\sqrt{2}$.

   Sauf erreur de ma part, une seconde d'arc carrée du ciel possède une magnitude de 22.5 dans le visible ($M_V = 22.5$) et correspond à un type spectral $\mathrm{S}_p = \mathrm{G2V}$ ($T = 5000\,\mathrm{K}$). Un télescope de $1\,\mathrm{m}$ de diamètre reçoit environ $6\,\mathrm{photons/sec}$ dans cette bande spectrale.

2. **Carré d'une variable aléatoire.**

   ① Une variable aléatoire X possède une moyenne μ et une variance σ^2. Calculer la moyenne du carré de X.

   (2 points)

   Corrigé.

   ❶ On a la formule $\mathrm{E}\{(X-\mu)^2\} = \mathrm{E}\{X^2\} - \mu^2 = \sigma^2$ d'où $\mathrm{E}\{X^2\} = \sigma^2 + \mu^2$.

3. **Spectroscopie.**
On observe un objet ponctuel à l'aide d'un spectroscope à fente. La turbulence de l'atmosphère fait que l'objet apparaît sous la forme d'une tache dont le profil peut être assimilé à une loi normale à deux dimensions. On suppose que l'image est à symétrie circulaire.
   ① Combien de paramètres sont-ils nécessaire afin de décrire le profil 2D de l'image ?
   ② On suppose que l'image est centrée sur l'origine d'un système d'axes xOy. Donner l'expression du profil $f(x,y)$ de l'image.
   ③ La fente du spectroscope est centrée sur l'image et sa largeur est 2δ on peut la considérer comme étant infinie. Qu'elle est alors en fonction de δ et des paramètres définis plus haut la fraction p de l'intensité pénétrant dans le spectroscope ?
   ④ Que doit valoir δ par rapport à la largeur a mi-hauteur de la tache afin que la fente laisse pénétrer 90% de la lumière de l'image ?

(1 + 2 + 2 + 2)

Corrigé.
   ❶ Quatre paramètres sont nécessaires : un pour l'intensité de l'objet, deux pour sa position et un seul pour sa dimension.
   ❷ Si A désigne l'intensité intégrée de l'objet et σ le paramètre d'échelle on a :
$$f(x,y) = \frac{A}{2\pi\sigma^2}\exp\Big\{-\frac{1}{2\sigma^2}[x^2+y^2]\Big\}\,.$$
Voir formule (4.11) page 64 du cours.
   ❸ Les lois marginales d'une distribution normale sont normales, la proportion de lumière pénétrant le spectroscope est l'intégrale de $-\delta$ à δ de cette loi marginale, c'est-à-dire que p se calcule à partir de la fonction de répartition de la loi marginale. On a :
$$p = \Phi\Big(\frac{\delta}{\sigma}\Big) - \Phi\Big(-\frac{\delta}{\sigma}\Big) = 2\Phi\Big(\frac{\delta}{\sigma}\Big) - 1\,,$$
où Φ est la fonction de répartition de la loi normale réduite. Voir §4.2.7 page 71 du cours.
   ❹ Soit Δ la largeur à mi-hauteur de la tache, on a : $\Delta \approx 2.355\sigma$ (formule (4.7) page 62). La table 4.1 page 64 nous apprend que δ/σ doit être égal à 1.65 environ. Soit en fonction de Δ :
$$\frac{\delta}{\Delta} = \frac{1.65}{2.35} \approx 0.7\,.$$
La largeur de la fente doit être égale à 1.4 fois la largeur à mi-hauteur de la tache correspondant à l'image.

4. **Une coquille absorbante.**
Une étoile est entourée d'une coquille de nuages absorbants. Cette coquille se présente sous la forme d'une sphère creuse de rayon r et d'épaisseur Δr négligeable devant r. Tous les nuages sont identiques et ils sont en nombre N. Le nombre moyen de nuages sur la ligne de visée est n et ils sont placés les uns par rapport aux autres dans la coquille de façon indépendante.

① Quel modèle probabiliste permet de calculer la probabilité de rencontrer k nuages sur la ligne de visée ?

② Donner, suivant ce modèle, l'expression de la probabilité : p_k, de rencontrer effectivement k nuages sur la ligne de visée.

On considére que le rayonnement de l'étoile est absorbé s'il rencontre au moins un nuage sur la ligne de visée.

③ Quelle est la probabilité : p_{abs}, pour que la source soit absorbée ?

④ Calculer cette probabilité pour $N = 10^4$ et $n = 0.1$.

⑤ Quel modèle probabiliste approximatif pourrait-on admettre dans le cas où $n \ll N$?

⑥ Calculer p_{abs} suivant ce modèle.

$(1 + 1 + 1 + 1 + 2 + 1 = 7$ points)

Corrigé.

❶ Le modèle probabiliste est le modèle binomial.

En effet on peut considérer que l'on « lance » les nuages un par un au hasard comme à « pile ou face » et qu'un nuage a la probabilité p d'intercepter la ligne de visée. On définit les variables aléatoires X_i qui valent 1 si le nuage numéro i intercepte le ligne de visée et 0 dans le cas contraire. Ce sont des variables de Bernoulli dont la moyenne vaut p, elles sont indépendantes. Le nombre ν de nuages interceptant la ligne de visée est la somme des N variables aléatoires identiques et indépendantes de Bernoulli X_i, ce nombre ν suit une loi binomiale de moyenne $Np = n$.

❷ La probabilité p_k de rencontrer k nuages est donnée par l'expression :

$$p_k = C_N^k p^k (1-p)^{N-k} \quad \text{avec} \quad p = \frac{n}{N}\,.$$

❸ On a $p_{\text{abs}} = 1 - p_0$, c'est-à-dire :

$$p_{\text{abs}} = 1 - C_N^0 p^0 (1-p)^N = 1 - (1-p)^N\,.$$

❹ Dans ce cas on a :

$$p_{\text{abs}} = 1 - (1-p)^N \approx 0.0951630\,.$$

On aurait pu aussi employer l'approximation :

$$1 - (1-p)^N = 1 - (1-p)^{\frac{n}{p}} \approx 1 - e^{-n} \approx n = 0.1\,.$$

On peut aussi raisonner ainsi : lorsque la moyenne n est petite on a pratiquement que deux événements possibles : soit on ne rencontre pas de nuages ; soit on en rencontre qu'un. La probabilité de rencontrer un nuage est p_1 par conséquent la probabilité d'absorption est identique à p_1. On a : $n = 0 \times (1 - p_1) + 1 \times p_1 = p_1 \equiv p_{\text{abs}}$ et on retrouve bien $p_{\text{abs}} = n$.

❺ Le modèle de Poisson de paramètre μ, où μ est la moyenne. On a alors $\mu = n$ et $p_k = \dfrac{n^k}{k!} e^{-n}$.

❻ On a suivant ce modèle :

$$p_{\text{abs}} = 1 - \frac{n^0}{0!} e^{-n} = 1 - e^{-n} \, .$$

Avec les nombres ci-dessus on trouve :

$$p_{\text{abs}} = 1 - e^{-0.1} \approx 0.0951626 \, .$$

5. **Simulation de la loi de Poisson.**
On dispose d'un programme permettant de « tirer » des nombres au hasard suivant la loi uniforme entre 0 et 1. Soit U une variable aléatoire simulée par ce programme.

① Comment transformer U de façon à obtenir une variable aléatoire T suivant la loi exponentielle de paramètre λ ($\lambda > 0$) ?

② Comment obtenir une variable aléatoire N suivant la loi de Poisson de moyenne μ ?

($2 + 3 = 5$ points)
Corrigé.

❶ Soit F la fonction de répartition de la loi exponentielle, le changement de variable aléatoire $T = F^{-1}(U)$ répond à la question (voir page 30 du cours). On a $F(t) = 1 - \exp(-\lambda t)$. D'où la variable aléatoire :

$$T = -\frac{1}{\lambda} \ln(1 - U)$$

suit la loi exponentielle. Mais $1 - U$ suit aussi la loi uniforme d'où :

$$T = -\frac{1}{\lambda} \ln(U)$$

suit aussi la loi exponentielle.

❷ Posons $\mu = \lambda t$. D'après les considérations du chapitre 7 sur les flux de Poisson, il suffit d'ajouter les T_i obtenus avec la loi exponentielle jusqu'à juste dépasser le seuil t, c'est-à-dire :

$$\sum_{i=1}^{N+1} T_i > t \text{ avec } \sum_{i=1}^{N} T_i \leq t \quad t = \frac{\mu}{\lambda} \, .$$

On simule ainsi l'arrivée de N photons dans le temps t lorsque le taux d'arrivée des photons est λ. Notez qu'il existe des algorithmes plus rapides que celui-là.

6. **Simulation d'un flux de Poisson stationnaire.**
On prétend simuler un flux de Poisson de paramètre λ pendant le temps t en mettant en œuvre l'algorithme suivant :

(a) On tire $n = \lambda t$ nombres aléatoires suivant la loi uniforme entre $[0, t]$. Soient U_i ces variables.

(b) On trie par ordre croissant ces variables de façon à obtenir un échantillon ordonné. Soient $U_{(i)}$, les nouvelles variables ainsi triées.

(c) On identifie les dates d'arrivées des événements avec les variables triées, soit $T_i = U_{(i)}$.

① Quelle est la densité de probabilité des variables U_i ?

② Quelle est la densité de probabilité f_1 de la variable $T_1 = U_{(1)}$?

③ Peut-on dire que le flux ainsi construit est un flux de Poisson ?

④ Dire, sans justification, comment modifier l'algorithme précédent de façon à obtenir effectivement un flux de Poisson.

($1 + 2 + 1 + 3 = 7$ points)

Corrigé.

❶ En tant que variables aléatoires uniformes entre $[0, t]$ leur densité de probabilité est $f(u) = \dfrac{1}{t}$ entre $[0, t]$ et 0 ailleurs.

❷ Voir cours §8.4.1 page 153. La fonction de répartition du minimum des U_i est donnée entre $[0, t]$ par :

$$F_1(\tau) = 1 - \left(1 - \frac{\tau}{t}\right)^n .$$

On trouve la densité de probabilité en dérivant F_1 :

$$f_1(\tau) = \frac{d}{d\tau} F_1(\tau) = \frac{n}{t}\left(1 - \frac{\tau}{t}\right)^{n-1} .$$

❸ Il est clair que l'expression précédente n'est pas la densité de probabilité d'une loi exponentielle. Elle en est cependant assez proche.

Plus généralement les dates d'arrivées T_i suivent la loi bêta et non la loi gamma comme ils devraient le faire si le flux était réellement de Poisson.

Le flux ainsi simulé n'est donc pas un flux de Poisson.

❹ Il faudrait modifier le point (a) et ne pas tirer un nombre fixe n de points mais un nombre variable N suivant la loi de Poisson de moyenne $\mu = \lambda t$.

7. **Information de Fisher.**

Soit U une variable aléatoire suivant la loi uniforme sur $[0, \theta]$, $\theta > 0$. La borne supérieure : θ de l'intervalle de définition de U est inconnue et on va chercher à l'estimer à l'aide d'un échantillon de taille n.

① Donner l'expression de la fonction de vraisemblance L de l'échantillon pour les valeurs de θ.

② Donner l'expression de l'information de Fisher $I_n(\theta)$ fournie par l'échantillon pour l'estimation de θ.

③ Calculer cette information de Fisher.

($2 + 1 + 2 = 5$ points)

Corrigé.

❶ Pour une réalisation u_i de U_i, on a la fonction de vraisemblance $F_i(u_i)$:

$$F_i(u_i) = \begin{cases} \dfrac{1}{\theta} & \text{si } u_i \le \theta \ ; \\ 0 & \text{si } u_i > \theta \, . \end{cases}$$

Ce qui s'écrit encore : $F_i(u_i) = \dfrac{1}{\theta}\mathbf{1}_{u_i \le \theta}$, où $\mathbf{1}_{u_i \le \theta}$ est la fonction indicatrice qui vaut 1 si $u_i \le \theta$ et 0 dans le cas contraire. Les variables aléatoires étant indépendantes, il vient : $L(u_1, \ldots, u_n|\theta) = \prod_{i=1}^n \dfrac{1}{\theta}\mathbf{1}_{u_i \le \theta}$. Soit :

$$L(u_1, \ldots, u_n|\theta) = \frac{1}{\theta^n}\mathbf{1}_{\max_i(u_i) \le \theta} \, .$$

❷ Par définition l'information de Fisher est donnée par :

$$I_n(\theta) = \mathrm{E}\left\{ \left(\frac{\partial \ln L}{\partial \theta} \right)^2 \right\} \, .$$

Il ne fallait *surtout pas* dire que :

$$I_n(\theta) = \mathrm{E}\left\{ -\frac{\partial^2 \ln L}{\partial \theta^2} \right\} \, ,$$

car nous ne sommes pas dans les conditions où l'inégalité de Rao-Cramér s'applique (la borne supérieure de l'intégration dans le calcul de l'espérance dépend du paramètre à estimer : θ).

❸ Dans le domaine $[0, \theta]$ où U est défini on a :

$$I_n(\theta) = \int \cdots \int_0^\theta \left(\frac{\partial \ln L}{\partial \theta} \right)^2 L \, du_1 \ldots du_n \, .$$

On a : $L = \dfrac{1}{\theta^n}$, $\ln L = -n \ln \theta$, $\dfrac{\partial}{\partial \theta} \ln L = -\dfrac{n}{\theta}$, d'où :

$$\begin{aligned} I_n(\theta) &= \int \cdots \int_0^\theta \frac{n^2}{\theta^2} \frac{1}{\theta^n} \, du_1 \ldots du_n \, , \\ &= \frac{n^2}{\theta^2} \, . \end{aligned}$$

On trouve $I_n(\theta) = n^2 I_1(\theta)$ et non pas $I_n(\theta) = n I_1(\theta)$. On aurait trouvé cette valeur si l'inégalité de Rao-Cramér avait été applicable.

8. **Courbe de lumière.**
On décide d'interpréter la courbe de lumière d'un astre variable par la méthode des moindres carrés. On suppose que la période de la courbe de lumière est T et que les observations y_i ; $i = 1, \ldots, n$ correspondent à un échantillonnage (non nécessairement régulier) de cette courbe à des temps t_i, $0 \le t_i \le T$. Le modèle $\boldsymbol{\mu}$ choisit est une série de Fourier tronquée :

$$\mu_i = \theta_0 + \sum_{p=1}^{k} \theta_{2p-1} \cos\left(\frac{2\pi p}{T} t_i \right) + \theta_{2p} \sin\left(\frac{2\pi p}{T} t_i \right) \, .$$

Les paramètres inconnus sont les θ_i.

① Ce modèle est-il linéaire ?

② Préciser la matrice $\boldsymbol{X}$ du modèle.

③ Le nombre de points est $n = 30$ et le nombre d'harmoniques $k = 5$. On estime que le bruit autour de la vraie courbe de lumière est de variance unité ($\sigma^2 = 1$). Le moindre carré est atteint pour la valeur $S = 20$. Pouvez-vous dire, en termes très généraux si le modèle semble bien adapté aux données ?

($1 + 2 + 2$ points)

❶ Oui bien-sûr, il est linéaire en θ_i.

❷ $\boldsymbol{X}$ est une matrice au format $(n, 2k+1)$ telle que $\boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\theta}$. Soit en posant $\omega = 2\pi/T$

$$\boldsymbol{X} = \begin{pmatrix} 1 & \cos\omega t_1 & \sin\omega t_1 & \cdots & \cos k\omega t_1 & \sin k\omega t_1 \\ 1 & \cos\omega t_2 & \sin\omega t_2 & \cdots & \cos k\omega t_2 & \sin k\omega t_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \cos\omega t_n & \sin\omega t_n & \cdots & \cos k\omega t_n & \sin k\omega t_n \end{pmatrix}$$

❸ Une estimation de la variance des données est $S/(30-(2k+1)) = S/19 \approx 1.0526$ (voir chapitre 13.3.10). Cette valeur est proche de 1 et le modéle semble raisonnable. Pour conclure de façon précise, il faudrait connaître la loi suivie par S. Dans le cas normal, c'est une loi du χ^2 à 19 degrés de liberté de moyenne 19 et d'écart-type $\sqrt{2 \times 19} \approx 6.16$. La valeur de S trouvée (S=20) s'écarte de la moyenne à moins d'un sixième de l'écart-type, l'accord est donc excellent. Dans le cas général, on prend aussi la loi du χ^2 comme approximation de la loi suivie par S.