Contrôle des connaissances du cours
"Bruits et Signaux"

$\boxed{\text{Corrigé}}$

Meudon le 16 décembre 1994 de 9h à 11h salle 204.

Le contrôle est noté sur 40, la note finale sur 20 sera : note$_{20} = \min(20,\text{note}_{40})$. Vous pouvez donc choisir les exercices que vous voulez. Le barême est indiqué en fin d'exercice.

1. Une variable aléatoire ξ possède une moyenne μ et une variance σ^2. Calculer la moyenne du carré de ξ.
   (1 point)
   On a la formule $\mathrm{E}\{(\xi-\mu)^2\} = \mathrm{E}\{\xi^2\} - \mu^2 = \sigma^2$ d'où $\mathrm{E}\{\xi^2\} = \sigma^2 + \mu^2$.
2. 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.
   i) Comment transformer U de façon à obtenir une variable aléatoire T suivant la loi exponentielle de paramètre λ ?
   ii) Comment obtenir une variable aléatoire N suivant la loi de Poisson de moyenne μ ?
   ($1+2$ points)
   i) 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). On a $F(t) = 1 - \exp(-\lambda t)$. D'où la variable aléatoire :
   $$T = -\frac{1}{\lambda}\ln(1-U)$$
   suit la loi exponentielle.

   ii) Posons $\mu = \lambda t$. D'après les considérations du chapitre 5 il suffit d'ajouter les T_i obtenus avec la loi exponentielle jusqu'à juste dépasser le seuil t.
   $$\sum_{i=1}^{N+1} T_i > t \text{ avec } \sum_{i=1}^{N} T_i \le t \quad t = \frac{\mu}{\lambda}.$$
   On simule ainsi l'arrivée de N photons dans le temps t lorsque le taux d'arrivée des photons est λ.
3. On fait l'hypothèse qu'un amas de galaxies est réparti uniformément sur la surface d'une plaque photographique de format carré. Cette hypothèse signifie que le nombre de galaxies susceptibles d'être présentes dans une surface quelconque S découpée sur la plaque est égale au rapport de cette surface S par la surface totale de la plaque. Soit $A = a^2$ cette surface. On repère la position des galaxies par leurs coordonnées cartésiennes X, Y par rapport au centre de la plaque et d'axes parallèles aux bords.
   i) Donnez l'expression des densités de probabilité suivies par X et Y.
   ii) Montrez que ces variables aléatoires sont indépendantes.

iii) Calculer la loi suivie par le carré de la distance d'une galaxie par rapport au centre de la plaque (évaluer la densité de probabilité sous forme d'intégrale).

$(2 + 1 + 2$ points)

F désigne une fonction de répartition, f une densité de probabilité. On a :
i) $\Pr X \leq x = a(x + a/2)/a^2 = (x + a/2)/a \equiv F_X(x)$, et $f_X(x) = 1/a$. $\Pr Y \leq y = a(y + a/2)/a^2 = (y + a/2)/a \equiv F_Y(y)$, et $f_Y(y) = 1/a$.
ii) $F(x, y) = \Pr X \leq x, Y \leq y = (x+a/2)(y+a/2)/a^2 = F_X(x)F_Y(y)$ d'où l'indépendance. Par dérivation on a $f(x, y) = 1/a^2$.
iii) On a $R^2 = X^2 + Y^2$. La loi suivie par X^2 est donnée par la formule (2.12), il vient : $f_{X^2}(u) = 1/(a\sqrt{u})$ et $f_{Y^2}(v) = 1/(a\sqrt{v})$ et nulle pour $u \notin [0, a^2/4]$ ou $v \notin [0, a^2/4]$. Pour la loi suivie par R^2 les variables aléatoires X^2 et Y^2 étant indépendantes, on obtient sa densité de probabilité par la convolution de f_{X^2} et f_{Y^2} (voir table 2.2). Il faut faire attention aux bornes d'intégration, soit en introduisant une distribution indicatrice du support des densités, soit en explicitant les bornes. C'est cette dernière option qui a été préférée dans la formule suivante qui donne la densité de probabilité de R^2

$$f_{R^2}(r^2) = \frac{1}{a^2} \int_{\max(0, r^2 - a^2/4)}^{\min(r^2, a^2/4)} \frac{du}{\sqrt{u(r^2 - u)}}$$

4. Une variable aléatoire Z est la somme de N variables aléatoires X_i indépendantes : $Z = \sum_{i=1}^{N} X_i$. Les variables X_i possèdent la même moyenne μ_X.

   i) Calculer la moyenne μ_Y de Y dans le cas où N est une constante.
   ii) Calculer cette moyenne dans le cas où N est une variable aléatoire discrète de moyenne μ_N. On posera $\Pr N = n = p_n$.

   $(1 + 2$ points)

   i) La moyenne d'une somme est égale à la somme des moyennes, d'où : $\mu_y = N\mu_X$.
   ii) Il faut employer la formule (1.48) des espérances conditionnelles totales d'où $\mu_y = \sum_{n=1}^{\infty} n\mu_X p_n = \mu_X \sum_{n=1}^{\infty} np_n = \mu_X \mu_N$.

5. On a observé un spectre formé de 30 pixels : $X_1, \ldots, X_{30}$. Sous hypothèse nulle (H_0), ce spectre ne représente qu'un bruit normal de moyenne μ et d'écart-type σ. On suppose que les paramètres μ et σ sont connus et que les pixels sont indépendants les uns des autres.

   i) On s'attend à trouver une raie en émission ou en absorption au pixel numéro 15 (par exemple). A quelle valeur de la fausse-alarme α correspond un critère de détection basé sur la "règle des 3-sigmas" ?

   ii) Même question, mais on s'attend à trouver une raie en émission.

   iii) On s'attend de nouveau à une raie en émission ou en absorption, mais on a aucune idée du pixel sur lequel elle est susceptible d'apparaître. Que devient la fausse-alarme α si l'on applique toujours la règle des 3-sigmas dès qu'au moins 1 pixel dépasse le seuil critique ?

   $(1 + 1 + 2$ points)

   i) On détecte une raie si le pixel numéro 30 dépasse le niveau 3σ en dessus ou au dessous de la moyenne μ. Soit Φ la fonction de répartition de la loi normale réduite on a $\alpha = 1 - (\Phi(3) - \Phi(-3))$ d'après la table 3.1 on a $\alpha \approx 0.003$
   ii) Dans ce cas on a $\alpha = 1 - \Phi(3) \approx 0.0015$.
   iii) La probabilité pour qu'au moins 1 pixel dépasse le seuil critique est "$1-$ la

probabilité pour qu'ils soient tous en dessous du seuil". Dans le cas d'une raie en émission ou en absorption cela donne :

$$\alpha = 1 - (\Phi(3) - \Phi(-3))^{30} \approx 0.0862 \approx 30 \times 0.003$$

6. Dans les conditions de l'exercice précédent, on suppose maintenant que le bruit est corrélé de matrice des variances-covariances $\boldsymbol{V}$ connue.

   i) Précisez un test qui permettrait de détecter une raie d'émission ou d'absorption avec la fausse alarme α.

   (2 points)

   i) C'est le test du χ^2 du chapitre 3.3.11 il faut former la quantité scalaire $\chi^2 = (\boldsymbol{x} - \boldsymbol{\mu})^t \boldsymbol{V^{-1}} (\boldsymbol{x} - \boldsymbol{\mu})$ où $\boldsymbol{x}$ désigne le spectre. Cette quantité suit une loi du χ^2 à n degrés de liberté et il faut la comparer à un seuil donné par la formule (3.101) ou la table 3.4. Par exemple, on détecte une raie avec $\alpha = 0.001$ si $\chi^2 > (7.727)^2$.

7. Le spectre précédent est supposé normal et IID, mais on ne connait pas les paramètres μ et σ. On envisage alors de les estimer à l'aide des estimateurs suivants :

$$M = \frac{1}{n} \sum_{i=1}^{n} X_i, \quad S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - M)^2, \quad S = \sqrt{S^2}\,.$$

   On réduit les variables aléatoires X_i à l'aide de ces estimateurs, suivant la formule :

$$Y_i = \frac{X_i - M}{S}\,.$$

   i) Montrer que la variable

$$\sqrt{\frac{n}{n-1}} Y_i,$$

   suit une loi de Student à $n-1$ degrés de liberté.

   (3 points)

   On sait d'après le chapitre 6.7.5 que le quotient d'une loi normale réduite par la racine d'une loi du χ^2 à n degrés de liberté le tout multiplié par $\sqrt{n}$, suit une loi de Student à n degrés de liberté . Il faut faire apparaître ces termes. Le calcul ne dépend pas de l'indice i, on posera $i = 1$. La variable aléatoire $X_1 - M$ est une somme de variables aléatoires normales indépendantes elle a pour moyenne 0 et pour écart-type $\sigma^2(1 - 1/n)$. La variable $(n-1)S^2/\sigma^2$ suit une loi du χ^2 à $n-1$ degrés de liberté (voir chapitre 6.7.3). D'où :

$$Y_i = \frac{X_1 - M}{\sigma(1 - 1/n)^{1/2}} \frac{\sigma}{S(n-1)^{1/2}} \frac{n-1}{n^{1/2}}$$

$$\sqrt{\frac{n}{n-1}} Y_i = \sqrt{n-1} \frac{X_1 - M}{\sigma(1 - 1/n)^{1/2}} \frac{\sigma}{S(n-1)^{1/2}} = T$$

   Ce qui montre que T est le quotient d'une loi normale réduite et de la racine carrée d'une loi du χ^2 à $n-1$ degrés de liberté, le tout multiplié par la racine du nombre de degrés de liberté. T suit donc une loi de Student à $n-1$ degrés de liberté.

8. On a observé un échantillon IID de taille n : $(x_1, \ldots, x_n)$, issu d'une loi de Poisson de moyenne μ.

   i) Donner l'expression du logarithme naturel de la fonction de vraisemblance de l'échantillon.

   ii) Trouver l'estimateur du maximum de vraisemblance de μ.

   $(1 + 1$ points)

   i) La loi de Poisson de moyenne μ est telle que $\Pr X = x = e^{-\mu}\mu^x/x!$ d'où le $\ln$ la fonction de vraisemblance :

   $$\ln L(x_1, \ldots, x_n|\mu) = -n\mu + \ln\mu \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \ln x_i!$$

   ii) $\partial \ln L/\partial\mu = -n + \sum_{i=1}^{n} x_i/\mu$ d'où $\widehat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$. C'est la moyenne arithmétique de l'échantillon, on vérifie que c'est bien un maximum : $\partial^2 \ln L/\partial\mu^2 = -\sum_{i=1}^{n} x_i/\mu^2 < 0$.

9. Le résultat X d'une expérience aléatoire dépend d'une autre variable aléatoire Θ qui représente *l'état de la nature*. On a de bonnes raisons de penser que le couple (Θ, X) suit une loi normale à deux dimensions dont tous les paramètres sont connus. Soient $\mu_\Theta, \mu, \sigma_\Theta, \sigma$ et ρ ces paramètres ; μ_Θ, σ_Θ font référence à Θ ; μ, σ font référence à X et ρ est le coefficient de corrélation du couple (Θ, X). On note θ une valeur prise par la variable aléatoire Θ.

   i) Donner l'expression de la loi *a priori* suivie par X, c'est-à-dire la loi qui préside au choix de X, θ n'étant pas connu. Quels sont la moyenne et la variance de cette loi.

   ii) Donner la loi *a posteriori* suivie par X, c'est-à-dire la loi qui préside au choix de X, θ étant connu. Quels sont la moyenne et la variance de cette loi.

   On connait le résultat donné par l'expérience, soit x ce résultat, en revanche on ne connait pas la valeur de θ.

   iii) Donner l'estimation du maximun de vraisemblance de θ.

   iv) Donner l'estimation du maximum de densité de probabilité *a posteriori* de θ.

   $(1 + 1 + 2 + 2$ points)

   i) La loi *a priori* est la loi marginale pour X du couple (Θ, X) c'est-à-dire ici une loi normale de moyenne μ et de variance σ^2 (voir chapitre 3.2.7).

   ii) La loi *a posteriori* est la loi conditionnelle pour la condition $\Theta = \theta$ c'est aussi une loi normale de moyenne $\mu + \rho\frac{\sigma}{\sigma_\Theta}(\theta - \mu_\Theta)$ et de variance $\sigma^2(1 - \rho^2)$ (voir chapitre 3.2.2)

   iii) La fonction de vraisemblance est la loi conditionnelle pour la condition $\Theta = \theta$. C'est la même loi qu'en ii) son maximum est atteint pour sa moyenne. Dans l'expression de cette moyenne, θ est inconnu et x est donné. Suivant le principe du maximum de vraisemblance pour un échantillon de taille 1 (voir exemple 6.2) on doit choisir θ tel que le maximum de la fonction de vraisemblance corresponde à l'observation x. Ces considérations conduisent à l'estimation :

   $$\widehat{\theta} = \mu_\Theta + \frac{1}{\rho}\frac{\sigma_\Theta}{\sigma}(x - \mu)$$

   iv) L'estimation du maximum de densité de probabilité *a posteriori* est donné par le maximum de la loi conditionnelle pour $X = x$ c'est-à-dire pour sa moyenne. On a alors :

$$\widehat{\theta} = \mu_\Theta + \rho \frac{\sigma_\Theta}{\sigma}(x - \mu)$$

10. Dans les conditions du problème précédent on procède à une série de n tirages, indépendants et suivant la même loi, de X. On dispose donc d'un échantillon IID : $(x_1, \ldots, x_n)$, issu de la loi suivie par X et dont les paramètres dépendent de θ. L'échantillon étant IID, θ ne change naturellement pas d'un tirage à l'autre.

    i) Donner l'estimation du maximun de vraisemblance de θ.

    (3 points)

    i) L'estimation du maximum de vraisemblance de la moyenne de la loi conditionnelle (θ étant connu) est la moyenne arithmétique $\bar{x}$ de l'échantillon. La moyenne de la loi est égale à $\mu + \rho\sigma(\theta - \mu_\Theta)/\sigma_\Theta$. L'estimation du maximum de vraisemblance d'un fonction d'un paramètre étant égale à la fonction de l'estimateur (voir chapitre 9.2.2 et équation 9.16) il vient :

$$\widehat{\theta} = \mu_\Theta + \frac{1}{\rho}\frac{\sigma_\Theta}{\sigma}(\bar{x} - \mu)$$

11. Un expérimentateur a interprété ses données $\boldsymbol{x}$ à l'aide d'un modèle théorique $\boldsymbol{\mu}$. La différence entre ses données et le modèle lui donne un bruit résiduel $\boldsymbol{\epsilon}$ que l'on peut considérer comme issu d'une loi normale à 30 dimensions de moyenne nulle et de matrice des variances-covariances $\boldsymbol{V}$ connue. Pour juger de l'adéquation de son modèle vis-à-vis des données, il décide de calculer la valeur du χ^2 attachée à ce bruit résiduel.

    i) Donner l'expression qui lui permet de calculer cette quantité χ^2.
    ii) Il a trouvé $\chi^2 = 49$, Cette valeur est-elle acceptable au niveau de "fausse-alarme" $\alpha = 0.01$?
    iii) Pouvez-vous dire, en des termes précis, quelle devra être sa conclusion ?
    ($1 + 1 + 1$ points)
    i) $\chi^2 = \boldsymbol{\epsilon}^t \boldsymbol{V}^{-1} \boldsymbol{\epsilon}$
    ii) Pour la loi normale cette quantité suit une loi du χ^2_{30} à 30 degrés de liberté. Des quantiles de cette loi sont donnés par la table 3.4, on trouve pour $\alpha = 0.01$ ($\gamma = 0.99$), $k_{0.99} = 7.134$. On a $\chi^2 = 49 < 7.134^2 \approx 51$, et la valeur est acceptable.
    iii) En conséquence, on doit admettre que les données ne contredisent pas l'hypothèse nulle. Le modèle $\boldsymbol{\mu}$ est donc compatible avec les données $\boldsymbol{x}$ au niveau de signification de 99%.

12. 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 :

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

    i) Ce modèle est-il linéaire ?
    ii) Préciser la matrice $\boldsymbol{X}$ du modèle.

iii) 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)

i) Oui bien-sûr, il est linéaire en θ_i.

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

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

iii) Une estimation de la variance des données est $S/(30 - (2k+1)) = S/19 \approx 1.0526$ (voir chapitre 10.3.9). 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.