# Économétrie

**5e édition**

## Annexes : exercices et corrigés

William Greene
New York University

Édition française dirigée par Didier Schlacther,
IEP Paris, université Paris II

Traduction :
Stéphanie Monjon, université Paris I Panthéon-Sorbonne

PEARSON
Education

Pearson Education France

47 bis, rue des Vinaigriers

75010 PARIS

Tél. : 01 72 74 90 00

Structuration des documents et mise en pages : edito.biz

# Annexe A

## Exercice 1

Pour les matrices A = $\begin{bmatrix} 1 & 3 & 3 \\ 2 & 4 & 1 \end{bmatrix}$ et B = $\begin{bmatrix} 2 & 4 \\ 1 & 5 \\ 6 & 2 \end{bmatrix}$, calculer AB, A′B′, et BA.

$$\mathbf{AB} = \begin{bmatrix} 23 & 25 \\ 14 & 30 \end{bmatrix}, \ \mathbf{BA} = \begin{bmatrix} 10 & 22 & 10 \\ 11 & 23 & 8 \\ 10 & 26 & 20 \end{bmatrix}$$

$$\mathbf{A'B'} = (\mathbf{BA})' = \begin{bmatrix} 10 & 11 & 10 \\ 22 & 23 & 26 \\ 10 & 8 & 20 \end{bmatrix}.$$

## Exercice 2

Prouver que $tr(\mathbf{AB}) = tr(\mathbf{BA})$ avec **A** et **B** deux matrices quelconques, non nécessairement carrées, pouvant être multipliées.

Le $i$-ième élément de la diagonale de **AB** est $\sum_j a_{ij} b_{ji}$. En sommant sur $i$, on obtient $tr(\mathbf{AB}) = \sum_i \sum_i a_{ij} b_{ji}$. Le $j$-ième élément de la diagonale de **BA** est $\sum_j b_{ji} a_{ij}$. En sommant sur $i$, on obtient $tr(\mathbf{BA}) = \sum_i \sum_j b_{ji} a_{ij}$.

## Exercice 3

Prouver que $tr(\mathbf{A'A}) = \sum_i \sum_j a_{ij}^2$.

Le $j$-ième élément de la diagonale de **A′A** est le produit de la $j$-ième colonne de **A**, $\sum_i a_{ij}^2$. En sommant sur $j$, on obtient $tr(\mathbf{A'A}) = \sum_j \sum_i a_{ij}^2 = \sum_i \sum_j a_{ij}^2$.

## Exercice 4

Développer le produit de la matrice $\mathbf{X} = \{[\mathbf{AB} + (\mathbf{CD})'][(\mathbf{EF})^{-1} + \mathbf{GH}]\}'$. On suppose que toutes les matrices sont carrées et que $\mathbf{E}$ et $\mathbf{F}$ sont non singulières.

On développe d'abord $(\mathbf{CD})' = \mathbf{D}'\mathbf{C}'$ et $(\mathbf{EF})^{-1} = \mathbf{F}^{-1}\mathbf{E}^{-1}$. Le produit est alors :

$$\{[\mathbf{AB} + (\mathbf{CD})'][(\mathbf{EF})^{-1} + \mathbf{GH}]\}' = (\mathbf{ABF}^{-1}\mathbf{E}^{-1} + \mathbf{ABGH} + \mathbf{D}'\mathbf{C}'\mathbf{F}^{-1}\mathbf{E}^{-1} + \mathbf{D}'\mathbf{C}'\mathbf{GH})'$$

$$= (\mathbf{E}^{-1})'(\mathbf{F}^{-1})'\mathbf{B}'\mathbf{A}' + \mathbf{H}'\mathbf{G}'\mathbf{B}'\mathbf{A}' + (\mathbf{E}^{-1})'(\mathbf{F}^{-1})'\mathbf{CD} + \mathbf{H}'\mathbf{G}'\mathbf{CD}$$

## Exercice 5

Prouver que, pour des vecteurs colonnes $K \times 1$, $\mathbf{x}_i$ $i = 1, ..., n$, et un vecteur non nul, $\mathbf{a}$,

$$\textstyle\sum_{i=1}^{n}(\mathbf{x}_i - \mathbf{a})(\mathbf{x}_i - \mathbf{a})' = \mathbf{X}'\mathbf{M}^0\mathbf{X} + n(\bar{\mathbf{x}} - \mathbf{a})(\bar{\mathbf{x}} - \mathbf{a})'$$

On écrit $\mathbf{x}_i - \mathbf{a}$ comme $[(\mathbf{x}_i - \bar{\mathbf{x}}) + (\bar{\mathbf{x}} - \mathbf{a})]$. La somme est alors :

$$\textstyle\sum_{i=1}^{n} [(\mathbf{x}_i - \bar{\mathbf{x}}) + (\bar{\mathbf{x}} - \mathbf{a})]\,[(\mathbf{x}_i - \bar{\mathbf{x}}) + (\bar{\mathbf{x}} - \mathbf{a})]' = \sum_{i=1}^{n} (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})' + \sum_{i=1}^{n}$$
$$\textstyle(\bar{\mathbf{x}} - \mathbf{a})(\bar{\mathbf{x}} - \mathbf{a})' + \sum_{i=1}^{n} (\mathbf{x}_i - \bar{\mathbf{x}})(\bar{\mathbf{x}} - \mathbf{a})' + \sum_{i=1}^{n} (\bar{\mathbf{x}} - \mathbf{a})(\mathbf{x}_i - \bar{\mathbf{x}})'$$

Puisque $(\bar{\mathbf{x}} - \mathbf{a})$ est un vecteur de constantes, il peut être extrait des sommes. Ainsi, le quatrième terme est $(\bar{\mathbf{x}} - \mathbf{a}) \sum_{i=1}^{n} (\mathbf{x}_i - \bar{\mathbf{x}})' = \mathbf{0}$. Le troisième terme est similaire. Le premier terme est $\mathbf{X}'\mathbf{M}^0\mathbf{X}$ par définition alors que le deuxième est $n(\bar{\mathbf{x}} - \mathbf{a})(\bar{\mathbf{x}} - \mathbf{a})'$.

## Exercice 6

On note $\mathbf{A}$ une matrice carrée dont les colonnes sont $[\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_M]$ et $\mathbf{B}$ tout réarrangement des colonnes de la matrice identité $M \times M$. Quelle opération est exécutée par la multiplication $\mathbf{AB}$ ? Que dire de $\mathbf{BA}$ ?

$\mathbf{B}$ est appelée une matrice de permutation. Chaque colonne de $\mathbf{B}$, notée $\mathbf{b}_i$, est une colonne d'une matrice identité. La $j$-ième colonne du produit des matrices $\mathbf{AB}$ est $\mathbf{A}\,\mathbf{b}_i$ qui est la $j$-ième colonne de $\mathbf{A}$. Par conséquent, la multiplication de $\mathbf{A}$ par $\mathbf{B}$ réarrange simplement (permute) les colonnes de $\mathbf{A}$ (d'où le nom). Chaque ligne du produit $\mathbf{BA}$ est l'une des lignes de $\mathbf{A}$, de sorte que $\mathbf{BA}$ est un réarrangement des lignes de $\mathbf{A}$. Bien sûr, $\mathbf{A}$ n'a pas besoin d'être carrée pour permuter ses lignes ou ses colonnes. Sinon, la matrice de permutation applicable est d'ordre différent pour les lignes et les colonnes.

## Exercice 7

On considère le cas $3 \times 3$ de la matrice **B** de l'exercice 6. Par exemple, $\mathbf{B} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$

On calcule $\mathbf{B}^2$ et $\mathbf{B}^3$. On répète cela pour une matrice $4 \times 4$. Peut-on généraliser le résultat trouvé ?

$$\mathbf{B}^2 = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \quad \mathbf{B}^3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Comme chaque puissance de **B** est un réarrangement de **I**, certaines puissances de **B** sont égales à **I**. Si $n$ est cette puissance, on trouve donc $\mathbf{B}^{n-1} = \mathbf{B}^{-1}$. Ce résultat est valable de façon générale.

## Exercice 8

Calculer $|\mathbf{A}|$, $tr(\mathbf{A})$ et $\mathbf{A}^{-1}$ pour $\mathbf{A} = \begin{bmatrix} 1 & 4 & 7 \\ 3 & 2 & 5 \\ 5 & 2 & 8 \end{bmatrix}$

$$|\mathbf{A}| = 1(2)(8) + 4(5)(5) + 3(2)(7) - 5(2)(7) - 1(5)(2) - 3(4)(8) = -18$$

$$tr(\mathbf{A}) = 1 + 2 + 8 = 11$$

$$\mathbf{A}^{-1} = \frac{-1}{18}\begin{bmatrix} \det\begin{pmatrix} 2 & 5 \\ 2 & 8 \end{pmatrix} & -\det\begin{pmatrix} 4 & 7 \\ 2 & 8 \end{pmatrix} & \det\begin{pmatrix} 4 & 7 \\ 2 & 5 \end{pmatrix} \\ -\det\begin{pmatrix} 3 & 5 \\ 5 & 8 \end{pmatrix} & \det\begin{pmatrix} 1 & 7 \\ 5 & 8 \end{pmatrix} & -\det\begin{pmatrix} 1 & 7 \\ 3 & 5 \end{pmatrix} \\ \det\begin{pmatrix} 3 & 2 \\ 5 & 2 \end{pmatrix} & -\det\begin{pmatrix} 1 & 4 \\ 5 & 2 \end{pmatrix} & \det\begin{pmatrix} 1 & 4 \\ 3 & 2 \end{pmatrix} \end{bmatrix} = \begin{bmatrix} -6/18 & 18/18 & -6/18 \\ -1/18 & 27/18 & -16/18 \\ 4/18 & -18/18 & 10/18 \end{bmatrix}$$

## Exercice 9

Déterminer la décomposition de Cholesky de la matrice $\mathbf{A} = \begin{bmatrix} 25 & 7 \\ 7 & 13 \end{bmatrix}$.

La décomposition de Cholesky d'une matrice **A** est le produit de matrices **LU** = **A** avec **L** une matrice triangulaire inférieure et $\mathbf{U} = \mathbf{L}'$. On écrit la décomposition comme $\begin{bmatrix} 25 & 7 \\ 7 & 13 \end{bmatrix} = \begin{bmatrix} \lambda_{11} & 0 \\ \lambda_{21} & \lambda_{22} \end{bmatrix} \cdot \begin{bmatrix} \lambda_{11} & \lambda_{21} \\ 0 & \lambda_{22} \end{bmatrix}$. Par multiplication directe, $25 = \lambda_{11}^2$ de sorte que $\lambda_{11} = 5$. Alors, $\lambda_{11}\lambda_{21} = 7$, de sorte que $\lambda_{21} = 7 / 5 = 1{,}4$. Finalement, $\lambda_{21}^2 + \lambda_{22}^2 = 13$, donc $\lambda_{22} = 3{,}322$.

## Exercice 10

Une matrice symétrique définie positive, **A**, peut aussi être écrite comme **A** = **UL**, avec **U** une matrice triangulaire supérieure et $\mathbf{L} = \mathbf{U}'$. Ce n'est néanmoins pas la décomposition de Cholesky. On obtient cette décomposition de la matrice dans l'exercice 9.

En utilisant la même logique que dans le problème précédent, $\begin{bmatrix} 25 & 7 \\ 7 & 13 \end{bmatrix} = \begin{bmatrix} \mu_{11} & \mu_{12} \\ 0 & \mu_{22} \end{bmatrix} \begin{bmatrix} \mu_{11} & 0 \\ \mu_{12} & \mu_{22} \end{bmatrix}$. En travaillant du bas vers la haut, $\mu_{22} = \sqrt{13} = 3{,}606$. Alors, $7 = \mu_{12}\mu_{22}$ de sorte que $\mu_{12} = 7 / \sqrt{13} = 1{,}941$. Finalement, $25 = \mu_{11}^2 + \mu_{12}^2$ de sorte que $\mu_{11}^2 = 25 - 49 / 13 = 21{,}23$, ou $\mu_{11} = 4{,}61$.

## Exercice 11

Quelle opération est réalisée en postmultipliant une matrice par une matrice diagonale ? Que dire de la prémultiplication ?

Les colonnes sont multipliées par l'élément de la diagonale correspondant. La prémultiplication multiplie les lignes par l'élément de la diagonale correspondant.

## Exercice 12

Est-ce que les formes quadratiques qui suivent sont positives pour toutes les valeurs de **x** ?

$$y = x_1^2 - 28x_1x_2 + (11x_2^2)$$

$$y = 5x_1^2 + x_2^2 + 7x_3^2 + 4x_1x_2 + 6x_1x_3 + 8x_2x_3$$

La première expression peut être écrite $\begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 1 & -14 \\ -14 & 11 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$. Le déterminant de la matrice est 121 – 196 = –75. Elle n'est alors pas définie positive. Ainsi, la première forme quadratique n'est pas nécessairement positive. La seconde forme repose sur la matrice $\begin{bmatrix} 5 & 2 & 3 \\ 2 & 1 & 4 \\ 3 & 4 & 7 \end{bmatrix}$. Il y a plusieurs façons de vérifier ce qui est demandé. Il est possible d'examiner les signes des minorants principaux, qui doivent être positifs. Les deux premiers termes sont 5 et 5(1) – 2(2)=1, mais le troisième, le déterminant, est –34. Par conséquent, la matrice n'est pas définie positive. Ses trois racines caractéristiques sont 11,1, 2,9 et –1. Ainsi, il existe des valeurs de $x_1$, $x_2$ et $x_3$ pour lesquelles la forme quadratique est négative.

## Exercice 13

Prouver que $tr(\mathbf{A}\otimes\mathbf{B}) = tr(\mathbf{A})tr(\mathbf{B})$.

Le $j$-ième bloc de la diagonale du produit est $a_{jj}\mathbf{B}$. Le $i$-ième élément de la diagonale est $a_{jj}b_{ii}$. Si on somme dans le $j$-ième bloc, on obtient $\sum_i a_{jj}b_{ii} = a_{jj}\sum_i b_{ii}$ .

La sommation vers le bas des blocs de la diagonale donne la trace,

$\sum_j a_{jj}\sum_i b_{ii} = tr(\mathbf{A})tr(\mathbf{B})$.

## Exercice 14

Une matrice, **A**, est *nilpotente* si $\lim_{k\to\infty} \mathbf{A}^k = \mathbf{0}$. Prouver qu'une condition nécessaire et suffisante pour qu'une matrice symétrique soit nilpotente est que toutes ses racines caractéristiques soient inférieures, en valeur absolue, à 1.

On utilise la décomposition spectrale pour écrire **A** comme $\mathbf{C}\Lambda\mathbf{C}'$ avec $\Lambda$ la matrice diagonale des racines caractéristiques. Alors la puissance $K$-ième de **A** est $\mathbf{C}\Lambda^K\mathbf{C}'$. La condition suffisante est évidente. Aussi, puisque $\lambda$ est plus grand que 1, $\Lambda^K$ doit croître fortement ; la condition est également nécessaire.

## Exercice 15

Calculer les racines caractéristiques de $\mathbf{A} = \begin{bmatrix} 2 & 4 & 3 \\ 4 & 8 & 6 \\ 3 & 6 & 5 \end{bmatrix}$.

Les racines sont déterminées par $|\mathbf{A} - \lambda\,\mathbf{I}| = 0$. Pour la matrice considérée, cela donne :

$$|\mathbf{A} - \lambda\mathbf{I}| = (2-\lambda)(8-\lambda)(5-\lambda) + 72 + 72 - 9(8-\lambda) - 36(2-\lambda) - 16(5-\lambda)$$
$$= -\lambda^3 + 15\lambda^2 - 5\lambda = -\lambda\,(\lambda^2 - 15\lambda + 5) = 0$$

Une solution est zéro. (On aurait pu le deviner). La deuxième colonne de la matrice correspond à deux fois la première, ainsi son rang n'est pas supérieur à 2. La matrice n'a donc pas plus de deux racines non nulles. Les deux autres racines sont $\left(15 \pm \sqrt{205}\right)/2$ = 0,341 et 4,659.

## Exercice 16

Supposer que $\mathbf{A} = \mathbf{A}(z)$ avec $z$ un scalaire. Qu'est-ce que $\partial\mathbf{x}'\mathbf{A}\mathbf{x} / \partial z$ ? On suppose maintenant que chaque élément de **x** est aussi une fonction de $z$. De nouveau, qu'est-ce que $\partial\mathbf{x}'\mathbf{A}\mathbf{x} / \partial z$ ?

La forme quadratique est $\sum_i \sum_j x_i x_j a_{ij}$, de sorte que

$\partial\mathbf{x}'\mathbf{A}(z)\mathbf{x} / \partial z = \sum_i \sum_j x_i x_j (\partial a_{ij} / \partial z) = \mathbf{x}'(\partial\mathbf{A}(z) / \partial z)\mathbf{x}$ où $\partial\mathbf{A}(z) / \partial z$ est une matrice des dérivées partielles. Maintenant, si chaque élément de **x** est aussi une fonction de $z$, alors :

$$\partial \mathbf{x}'\mathbf{A}\mathbf{x} / \partial z = \textstyle\sum_i \sum_j x_i x_j (\partial a_{ij} / \partial z) + \sum_i \sum_j (\partial x_i / \partial z) x_j a_{ij} + \sum_i \sum_j x_i (\partial x_j / \partial z) a_{ij}$$

$$= \mathbf{x}'(\partial \mathbf{A}(z) / \partial z)\mathbf{x} + (\partial \mathbf{x}(z) / \partial z)'\mathbf{A}(z)\mathbf{x}(z) + \mathbf{x}(z)'\mathbf{A}(z)(\partial \mathbf{x}(z) / \partial z)$$

Si **A** est symétrique, l'expression se simplifie un peu en

$'(\partial \mathbf{A}(z) / \partial z)\mathbf{x} + 2(\partial \mathbf{x}(z) / \partial z)'\mathbf{A}(z)\mathbf{x}(z)$.

## Exercice 17

Montrer que les solutions des équations du déterminant $|\mathbf{B} - \lambda\mathbf{A}| = 0$ et $|\mathbf{A}^{-1}\mathbf{B} - \lambda\mathbf{I}| = 0$ sont les mêmes. Comment les solutions de cette équation sont-elles liées à celles de l'équation $|\mathbf{B}^{-1}\mathbf{A} - \mu\mathbf{I}| = 0$ ?

Puisque **A** est supposée être non singulière, on peut écrire :

$$\mathbf{B} - \lambda\mathbf{A} = \mathbf{A}(\mathbf{A}^{-1}\mathbf{B} - \lambda\mathbf{I}). \text{ Alors, } |\mathbf{B} - \lambda\mathbf{A}| = |\mathbf{A}|\times|\mathbf{A}^{-1}\mathbf{B} - \lambda\mathbf{I}|$$

Le déterminant de **A** n'est pas zéro si **A** est non singulière, de sorte que les solutions des deux équations doivent être les mêmes. $\mathbf{B}^{-1}\mathbf{A}$ étant l'inverse de $\mathbf{A}^{-1}\mathbf{B}$, ses racines caractéristiques doivent être les réciproques de celles de $\mathbf{A}^{-1}\mathbf{B}$. Cela pourrait poser problème ici puisque ces deux matrices ne sont pas nécessairement symétriques, de sorte que les racines pourraient être complexes. Mais, pour l'application donnée, les deux matrices **A** et **B** sont symétriques et définies positives. On peut alors montrer que la solution est la même que la troisième équation du déterminant impliquant une matrice symétrique.

## Exercice 18

En utilisant la matrice **A** donnée dans l'exercice 9, trouver le vecteur **x** qui minimise $y = \mathbf{x}'\mathbf{A}\mathbf{x} + 2x_1 + 3x_2 - 10$. Quelle est la valeur de $y$ au minimum ? Maintenant, minimiser $y$ sous la contrainte $x_1 + x_2 = 1$. Comparer les deux solutions.

La solution qui minimise $y = \mathbf{x}'\mathbf{A}\mathbf{x} + \mathbf{b}'\mathbf{x} + d$ satisfait $\partial y \partial \mathbf{x} = 2\mathbf{A}\mathbf{x} + \mathbf{b} = \mathbf{0}$. Pour ce problème, $\mathbf{A} = \begin{bmatrix} 25 & 7 \\ 7 & 13 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ et $\mathbf{A}^{-1} = \begin{bmatrix} 13/276 & -7/276 \\ -7/276 & 25/276 \end{bmatrix}$, de sorte que la solution est $x_1 = -5 / 552 = -0{,}0090597$ et $x_2 = -61 / 552 = -0{,}110507$.

Le problème de maximisation sous contrainte peut être traité sous la forme d'un lagrangien : $L^* = \mathbf{x}'\mathbf{A}\mathbf{x} + \mathbf{b}'\mathbf{x} + d + \lambda(\mathbf{c}'\mathbf{x} - 1)$ où $\mathbf{c} = [1,1]'$. Les conditions nécessaires pour la solution sont :

$$\partial L^* / \partial \mathbf{x} = 2\mathbf{A}\mathbf{x} + \mathbf{b} + \lambda\mathbf{c} = \mathbf{0}$$

$$\partial L^* / \partial \lambda = \mathbf{c}'\mathbf{x} - 1 \qquad = 0$$

ou

$$\begin{bmatrix} 2\mathbf{A} & \mathbf{c} \\ \mathbf{c}' & 0 \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \lambda \end{bmatrix} = \begin{bmatrix} \mathbf{-b} \\ 1 \end{bmatrix}$$

En insérant **A, b** et **c**, on a la solution $\begin{bmatrix} 50 & 14 & 1 \\ 14 & 26 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \lambda \end{bmatrix} = \begin{bmatrix} -2 \\ -3 \\ 1 \end{bmatrix}$. La solution aux trois équations est obtenue en multipliant le vecteur de droite par l'inverse de la matrice de gauche. Les résultats sont 0,27083, 0,72917 et –25,75. La valeur de la fonction à la solution contrainte est 4,240, qui est plus grande que la valeur non contrainte –10,00787.

## Exercice 19

Quel est le jacobien des transformations suivantes ? (Une remarque, pour les lecteurs intéressés par la technique, concernant une erreur courante dans la littérature : un jacobien est un déterminant. Le terme « déterminant jacobien » est une redondance inutile.)

$$y_1 = x_1 / x_2$$

$$\ln y_2 = \ln x_1 - \ln x_2 + \ln x_3$$

$$y_3 = x_1 x_2 x_3$$

Les lettres capitales représentent des logarithmes. Alors, les trois transformations peuvent être écrites comme :

$$Y_1 = X_1 - X_2$$

$$Y_2 = X_1 - X_2 + X_3$$

$$Y_3 = X_1 + X_2 + X_3$$

Cette transformation linéaire est :

$$\mathbf{Y} = \begin{bmatrix} 1 & -1 & 0 \\ 1 & -1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \mathbf{X} = \mathbf{JX}$$

La transformation inverse est :

$$\mathbf{X} = \begin{bmatrix} 1 & -1/2 & 1/2 \\ 0 & -1/2 & 1/2 \\ 1 & 1 & 0 \end{bmatrix} \mathbf{Y} = \mathbf{J}^{-1}\mathbf{Y}$$

En fonction des variables originales, on a alors $x_1 = y_1(y_2 / y_3)^{1/2}$, $x_2 = (y_3 / y_2)^{1/2}$, et $x_3 = y_1 y_2$. La matrice des dérivées partielles peut être obtenue directement, mais un raccourci algébrique se révèle utile pour obtenir le jacobien. On remarque tout d'abord que $\partial x_i / \partial y_j = (x_i / y_j)(\partial \log x_i / \partial \log y_j)$. Par conséquent, les éléments des dérivées partielles des transformations inverses sont obtenus en multipliant la $i$-ième ligne par $x_i$, dans laquelle on substitue l'expression pour $x_i$ en termes de $y$, on multiplie alors la $j$-ième colonne par $(1 / y_j)$. Ainsi, le résultat de l'exercice 11 est utile ici. La matrice des dérivées partielles est :

$$\begin{bmatrix} \partial x_1/\partial y_1 & \partial x_1/\partial y_2 & \partial x_1/\partial y_3 \\ \partial x_2/\partial y_1 & \partial x_2/\partial y_2 & \partial x_2/\partial y_3 \\ \partial x_3/\partial y_1 & \partial x_3/\partial y_2 & \partial x_3/\partial y_3 \end{bmatrix} =$$

$$\begin{bmatrix} x_1 & 0 & 0 \\ 0 & x_2 & 0 \\ 0 & 0 & x_3 \end{bmatrix} \begin{bmatrix} 1 & -1/2 & 1/2 \\ 0 & -1/2 & 1/2 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1/y_1 & 0 & 0 \\ 0 & 1/y_2 & 0 \\ 0 & 0 & 1/y_3 \end{bmatrix}$$

Le déterminant de la matrice produit est le produit des trois déterminants. Le déterminant de la matrice du centre est $-1 / 2$. Les déterminants des matrices diagonales sont les produits des éléments de la diagonale. Par conséquent, le jacobien est $J = \text{abs}(|\partial \mathbf{x} / \partial \mathbf{y}'|) = ½(x_1x_2x_3) / (y_1y_2y_3) = 2(y_1 / y_2)$ (après avoir fait les substitutions pour $x_i$).

## Exercice 20

Prouver que l'échange des deux colonnes d'une matrice carrée inverse le signe de son déterminant. (Indice : utiliser une matrice permutation. Voir l'exercice 6.)

Échanger les deux premières colonnes d'une matrice équivaut à la multiplier par une matrice permutation $\mathbf{B} = [\mathbf{e}_2, \mathbf{e}_1, \mathbf{e}_3, \mathbf{e}_4, ...]$ où $\mathbf{e}_i$ est la $i$-ième colonne d'une matrice identité. Ainsi, le déterminant de la matrice est $|\mathbf{AB}| = |\mathbf{A}|\ |\mathbf{B}|$. La question porte sur le déterminant de $\mathbf{B}$. On suppose que $\mathbf{A}$ et $\mathbf{B}$ ont $n$ colonnes. Pour obtenir le déterminant de $\mathbf{B}$, on le développe simplement le long de la première ligne. Le seul terme non nul dans le déterminant est $(-1)|\mathbf{I}_{n-1}| = -1$, où $\mathbf{I}_{n-1}$ est la matrice identité $(n-1)\times(n-1)$. Cela complète la preuve.

## Exercice 21

Supposer que $\mathbf{x} = \mathbf{x}(z)$ avec $z$ un scalaire. Qu'est-ce que $\partial[(\mathbf{x}'\mathbf{A}\mathbf{x}) / (\mathbf{x}'\mathbf{B}\mathbf{x})] / z$ ?

Les dérivées requises sont données dans l'exercice 16. On pose $\mathbf{g} = \partial\mathbf{x} / \partial z$ ; et on note le numérateur et le dénominateur respectivement $a$ et $b$. Alors :

$$\partial(a / b) / \partial z = [b(\partial a / \partial z) - a(\partial b / \partial z)] / b^2 = [\mathbf{x}'\mathbf{B}\mathbf{x}(2\mathbf{x}'\mathbf{A}\mathbf{g}) - \mathbf{x}'\mathbf{A}\mathbf{x}(2\mathbf{x}'\mathbf{B}\mathbf{g})] / (\mathbf{x}'\mathbf{B}\mathbf{x})^2 = 2[\mathbf{x}'\mathbf{A}\mathbf{x} / \mathbf{x}'\mathbf{B}\mathbf{x}][\mathbf{x}'\mathbf{A}\mathbf{g} / \mathbf{x}'\mathbf{A}\mathbf{x} - \mathbf{x}'\mathbf{B}\mathbf{g} / \mathbf{x}'\mathbf{B}\mathbf{x}]$$

## Exercice 22

On suppose que $\mathbf{y}$ est un vecteur $n \times 1$ et $\mathbf{X}$ une matrice $n \times K$. La projection de $\mathbf{y}$ dans l'espace des colonnes de $\mathbf{X}$ est définie dans le manuel après l'équation (2-55), $\hat{\mathbf{y}} = \mathbf{X}\mathbf{b}$. On considère maintenant la projection de $\mathbf{y}^* = c\mathbf{y}$ dans l'espace des colonnes de $\mathbf{X}^* = \mathbf{X}\mathbf{P}$, où $c$ est un scalaire et $\mathbf{P}$ une matrice $K \times K$ non singulière. Trouver la projection de $\mathbf{y}^*$ dans l'espace des colonnes de $\mathbf{X}^*$. Prouver que le cosinus de l'angle entre $\mathbf{y}^*$ et sa projection dans l'espace des colonnes de $\mathbf{X}^*$ est le même que celui entre $\mathbf{y}$ et sa projection dans l'espace des colonnes de $\mathbf{X}$. Comment interpréter ce résultat ?

La projection de $\mathbf{y}^*$ dans l'espace des colonnes de $\mathbf{X}^*$ est $\mathbf{X}^*\mathbf{b}^*$ où $\mathbf{b}^*$ est la solution à l'ensemble des équations $\mathbf{X}^{*\prime}\mathbf{y}^* = \mathbf{X}^{*\prime}\mathbf{X}^*\mathbf{b}^*$ ou $\mathbf{P}'\mathbf{X}'(c\mathbf{y}) = \mathbf{P}'\mathbf{X}'\mathbf{X}\mathbf{P}\mathbf{b}^*$. Comme $\mathbf{P}$ est non singulière, $\mathbf{P}'$ a un inverse. En multipliant l'équation par $(\mathbf{P}')^{-1}$, on a $c\mathbf{X}'\mathbf{y} = \mathbf{X}'\mathbf{X}(\mathbf{P}\mathbf{b}^*)$ ou $\mathbf{X}'\mathbf{y} = \mathbf{X}'\mathbf{X}[(1 / c)\mathbf{P}\mathbf{b}^*]$. Donc, en termes du $\mathbf{y}$ original et $\mathbf{X}$, on voit que $\mathbf{b} = (1 / c)\mathbf{P}\mathbf{b}^*$, ce qui implique $\mathbf{b}^* = c\mathbf{P}^{-1}\ \mathbf{b}$. La projection est $\mathbf{X}^*\mathbf{b}^* = (\mathbf{X}\mathbf{P})(c\mathbf{P}^{-1}\mathbf{b}) = c\mathbf{X}\mathbf{b}$. On conclut donc que la projection de $\mathbf{y}^*$ dans l'espace des colonnes de $\mathbf{X}^*$ est un multiple $c$ de la projection de $\mathbf{y}$ dans l'espace de $\mathbf{X}$. Cela se comprend puisque, si $\mathbf{P}$ est une matrice non singulière, l'espace des colonnes de $\mathbf{X}^*$ est exactement le même que celui de $\mathbf{X}$. Le cosinus de l'angle entre $\mathbf{y}^*$ et sa projection est le même qu'entre $c\mathbf{y}$ et $c\mathbf{X}\mathbf{b}$. Bien sûr, c'est aussi le même qu'entre $\mathbf{y}$ et $\mathbf{X}\mathbf{b}$ puisque la longueur des deux vecteurs n'est pas liée au cosinus de l'angle entre eux. Ainsi, $\cos\theta = (c\mathbf{y})'(c\mathbf{X}\text{b})) / (\|c\mathbf{y}\|\times\|c\mathbf{X}\mathbf{b}\|) = (\mathbf{y}'\mathbf{X}\mathbf{b})) / (\|\mathbf{y}\|\times\|\mathbf{X}\mathbf{b}\|)$.

## Exercice 23

Pour la matrice $\mathbf{X}' = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 4 & -2 & 3 & -5 \end{bmatrix}$, calculer $\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ et $\mathbf{M} = (\mathbf{I} - \mathbf{P})$. Vérifier que $\mathbf{MP} = \mathbf{0}$. On pose $\mathbf{Q} = \begin{bmatrix} 1 & 3 \\ 2 & 8 \end{bmatrix}$. (Indice : montrer que **M** et **P** sont idempotentes.)

a. Calculer le **P** et le **M** fondés sur **XQ** au lieu de **X**.

b. Quelles sont les racines caractéristiques de **M** et **P** ?

En premier lieu :

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 4 & 0 \\ 0 & 54 \end{bmatrix}$$

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 1/4 & 0 \\ 0 & 1/54 \end{bmatrix}$$

$$\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \begin{bmatrix} 1 & 4 \\ 1 & -2 \\ 1 & 3 \\ 1 & -5 \end{bmatrix} \begin{bmatrix} 1/4 & 0 \\ 0 & 1/54 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 4 & -2 & 3 & -5 \end{bmatrix} = \frac{1}{108} \begin{bmatrix} 59 & 11 & 51 & -13 \\ 11 & 35 & 15 & 47 \\ 51 & 15 & 45 & -3 \\ -13 & 47 & -3 & 77 \end{bmatrix} = \mathbf{P}$$

$$\mathbf{M} = \mathbf{I} - \mathbf{P} = \frac{1}{108} \begin{bmatrix} 49 & -11 & -51 & 13 \\ -11 & 73 & -15 & -47 \\ -51 & -15 & 63 & 3 \\ 13 & -47 & 3 & 31 \end{bmatrix}$$

a. Il n'est pas nécessaire de recalculer les matrices **M** et **P** pour **XQ**, car ce sont les mêmes. La preuve est que la contrepartie de **P** est $(\mathbf{XQ})[(\mathbf{XQ})'(\mathbf{XQ})]^{-1}(\mathbf{XQ})' = \mathbf{XQ}[\mathbf{Q}'\mathbf{X}'\mathbf{XQ}]^{-1}\mathbf{Q}'\mathbf{X}' = \mathbf{XQQ}^{-1}(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{Q}')^{-1}\mathbf{Q}'\mathbf{X}' = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. La **M** matrice est aussi la même. C'est une application du résultat trouvé dans l'exercice précédent. La **P** matrice est la matrice projection et, comme on l'a vu, la projection dans l'espace de **X** est la même que la projection dans l'espace de **XQ**.

b. Puisque **M** et **P** sont idempotentes, leurs racines caractéristiques doivent toutes être 0 ou 1. La trace de la matrice est égale à la somme des racines, ce qui révèle la proportion de 1 et 0. Pour les matrices ci-dessus, les traces de **M** et **P** sont égales à 2, de sorte que chacune a deux racines unitaires et deux racines nulles.

# Exercice 24

On suppose que **A** est une matrice $n \times n$ de la forme $\mathbf{A} = (1 - \rho\mathbf{I}) + \rho\mathbf{ii}'$, où **i** est une colonne de 1 et $0 < \rho < 1$. Écrire le format de **A** explicitement pour $n = 4$. Trouver toutes les racines et tous les vecteurs caractéristiques de **A**. (Indice : il y a seulement deux racines caractéristiques distinctes, qui se produisent respectivement 1 et $n - 1$ fois. Chaque **c** d'un certain type est un vecteur caractéristique de **A**.)

Pour $n = 4$, $\mathbf{A} = \begin{bmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{bmatrix}$. Il y a plusieurs façons d'analyser cette matrice. On peut utiliser un raccourci simple. Les racines et vecteurs caractéristiques satisfont $[(1 - \rho)\mathbf{I} + \rho\mathbf{ii}']\mathbf{c} = \lambda\mathbf{c}$. On multiplie cela pour obtenir $(1 - \rho)\mathbf{c} + \rho\mathbf{ii}'\mathbf{c} = \lambda\mathbf{c}$ ou $\rho\mathbf{ii}'\mathbf{c} = [\lambda - (1 - \rho)]\mathbf{c}$. On pose $\mu = \lambda - (1 - \rho)$, de sorte que $\rho\mathbf{ii}'\mathbf{c} = \mu\mathbf{c}$. On a seulement besoin de trouver les racines caractéristiques de $\rho\mathbf{ii}'$, $\mu$. Les racines caractéristiques de la matrice originale sont juste $\lambda = \mu + (1 - \rho)$. Maintenant, $\rho\mathbf{ii}'$ est une matrice de rang 1, puisque chaque colonne est identique. Par conséquent, $n - 1$ des $\mu$ sont nuls. Ainsi, la matrice originale a $n - 1$ racines égales à $0 + (1 - \rho) = (1 - \rho)$. On peut trouver la dernière racine en remarquant que la somme des racines de $\rho\mathbf{ii}'$ est égale à la trace de $\rho\mathbf{ii}'$. Comme $\rho\mathbf{ii}'$ a seulement une racine non nulle, cette dernière est la trace, $n\rho$. Ainsi, la racine restante de la matrice originale est $(1 - \rho + n\rho)$. Les vecteurs caractéristiques satisfont l'équation $\rho\mathbf{ii}'\mathbf{c} = \mu\mathbf{c}$. Pour la racine non nulle, on a $\rho\mathbf{ii}'\mathbf{c} = n\rho\mathbf{c}$. On divise par $n\rho$ pour obtenir $\mathbf{i}(1 / n)\mathbf{i}'\mathbf{c} = \mathbf{c}$. Cette équation indique que, pour chaque élément dans le vecteur, $c_i = (1 / n)\sum_i c_i$. Cela implique que chaque élément dans le vecteur caractéristique correspondant à la racine $(1 - \rho + n\rho)$ soit le même ; en d'autres termes, **c** est un multiple d'une colonne de « 1 ». Ainsi, comme il a une longueur unitaire, le vecteur est $(1/\sqrt{n})\,\mathbf{i}$. Pour les racines zéro restantes, les vecteurs caractéristiques doivent satisfaire $\rho\mathbf{i}(\mathbf{i}'\mathbf{c}) = 0\mathbf{c} = \mathbf{0}$. Si le vecteur caractéristique n'est pas une colonne de « 0 », la seule façon d'obtenir une inégalité est d'égaliser $\mathbf{i}'\mathbf{c}$ à zéro. Par conséquent, pour les $n - 1$ vecteurs caractéristiques restants, on peut utiliser des vecteurs orthogonaux dont les éléments ont une somme nulle et dont les produits sont égaux à un. Il y a un nombre infini de tels vecteurs. Par exemple, on pose **D** un ensemble arbitraire de $n - 1$ vecteurs contenant $n$ éléments. On transforme les colonnes de **D** en déviations aux moyennes de leur propre colonne. Ainsi, on pose $\mathbf{F} = \mathbf{M}^0\mathbf{D}$ où $\mathbf{M}^0$ est défini dans la section 2.3.6. Maintenant, on pose $\mathbf{C} = \mathbf{F}(\mathbf{F}'\mathbf{F})^{-2}$. **C** une combinaison linéaire de colonnes de **F** : ses colonnes se somment à zéro. Les colonnes sont orthogonales et ont une longueur unitaire.

## Exercice 25

Trouver l'inverse de la matrice de l'exercice 24.

En utilisant l'indice, l'inverse est :

$$[(1-\rho)\mathbf{I}]^{-1} - \frac{[(1-\rho)\mathbf{I}]^{-1}[\rho\mathbf{ii}'][(1-\rho)\mathbf{I}]^{-1}}{1+\left(\sqrt{\rho}\mathbf{i}\right)'[(1-\rho)\mathbf{I}]^{-1}\left(\sqrt{\rho}\mathbf{i}\right)} = \frac{1}{1-\rho}\{\mathbf{I} - [\rho/(1-\rho+n\rho)]\mathbf{ii}'\}$$

## Exercice 26

Prouver que chaque matrice dans la séquence de matrices $\mathbf{H}_{i+1} = \mathbf{H}_i + \mathbf{d}_i\mathbf{d}_i'$, où $\mathbf{H}_0 = \mathbf{I}$, est définie positive. Pour une application, voir la section 5.5. Pour une extension, prouver que chaque matrice dans la séquence de matrices définie dans (5-22) est définie positive si $\mathbf{H}_0 = \mathbf{I}$.

Par substitutions répétées, on trouve $\mathbf{H}_{i+1} = \mathbf{I} + \sum_{j=1}^{i}\mathbf{d}_j\mathbf{d}_j'$. Une forme quadratique de $\mathbf{H}_{i+1}$ est donc :

$$\mathbf{x}'\mathbf{H}_{i+1}\mathbf{x} = \mathbf{x}'\mathbf{x} + \sum_{j=1}^{i}(\mathbf{x}'\mathbf{d}_j)(\mathbf{d}_j'\mathbf{x}) = \mathbf{x}'\mathbf{x} + \sum_{j=1}^{i}(\mathbf{x}'\mathbf{d}_j)^2$$

C'est positif pour tout $\mathbf{x}$. Une façon simple d'établir cela pour la matrice dans (5-22) est de remarquer que, malgré sa complexité, elle est de forme $\mathbf{H}_{i+1} = \mathbf{H}_i + \mathbf{d}_i\mathbf{d}_i' + \mathbf{f}_i\mathbf{f}_i'$. Si cela commence avec une matrice définie positive, telle que $\mathbf{I}$, alors un argument similaire permet d'établir le fait qu'elle est définie positive.

## Exercice 27

Quelle est la matrice inverse de $\mathbf{P} = \begin{bmatrix} \cos(x) & \sin(x) \\ -\sin(x) & \cos(x) \end{bmatrix}$ ? Quelles sont les racines caractéristiques de $\mathbf{P}$ ?

Le déterminant de $\mathbf{P}$ est $\cos^2(x) + \sin^2(x) = 1$, de sorte que l'inverse « inverse » juste les signes des éléments hors de la diagonale. Les deux racines sont les solutions de $|\mathbf{P} - \lambda\mathbf{I}| = 0$, qui est $\cos^2(x) + \sin^2(x) - 2\lambda\cos(x) + \lambda^2 = 0$. Cela se simplifie car $\cos^2(x) + \sin^2(x) = 1$. En utilisant la formule quadratique, alors :$\lambda = \cos(x) \pm (\cos^2(x) - 1)^{1/2}$. Mais, $\cos^2(x) - 1 = -\sin^2(x)$. Par conséquent, les solutions imaginaires de la forme quadratique résultante sont $\lambda_1,\lambda_2 = \cos(x) \pm i\sin(x)$.

# Annexe B

## Exercice 1

Combien de mains différentes de 5 cartes peuvent être tirées au poker avec un jeu de 52 cartes ?

Il y a $\binom{52}{5}$ = (52 $\times$ 51 $\times$ 51... $\times$ 1) / [(5 $\times$ 4 $\times$ 3 $\times$ 2 $\times$ 1)(47 $\times$ 46$\times$ ... $\times$ 1)] = 2 598 960 mains possibles.

## Exercice 2

Calculer la probabilité d'avoir un « 4 » dans une main au poker.

Il y a 48(13) mains possibles contenant un « 4 » et une autre carte quelconque parmi les 48 restantes. Ainsi, étant donné la réponse au problème précédent, la probabilité d'avoir une de ces mains est de 48(13) / 2 598 960 = 0,00024, ou moins d'une chance sur 4 000.

## Exercice 3

On suppose qu'un ticket de loterie coûte 1 euro par jeu. Un jeu correspond à un tirage de 6 nombres, entre 1 et 48, sans remise. Si vous devinez les six nombres, vous gagnez le prix. On suppose maintenant que $N$ = nombre de tickets vendus et $P$ = niveau du prix. $N$ et $P$ sont liés par les relations :

$$N = 5 + 1{,}2P$$

$$P = 1 + 0{,}4N$$

$N$ et $P$ sont en millions. Quelle est la valeur espérée d'un ticket dans ce jeu ? (Ne pas oublier que le prix est partagé avec les autres gagnants.)

Le niveau du prix et le nombre de tickets vendus sont déterminés conjointement. Les solutions des deux équations sont $N = 11{,}92$ millions de tickets et $P = 5{,}77$ millions d'euros. Le nombre de combinaisons possibles des 48 nombres sans remise est :

$$\binom{48}{6} = (48 \times 47 \times 46... \times 1) / [(6 \times 5 \times 4 \times 3 \times 2 \times 1)(42 \times 41 \times ... \times 1)] = 12\ 271\ 512.$$

La probabilité de faire le bon choix est donc de $1 / 12271512 = 0{,}000000081$. Le nombre espéré de gagnants est la valeur espérée d'une variable binomiale avec $N$ essais et une probabilité de succès égale à $11{,}92 / 12{,}27 = 0{,}97$, soit environ 1. Ainsi, on ne s'attend pas à devoir partager le prix. La valeur espérée du ticket est Prob[gagne](5,77 millions – 1) + Prob[perd](–1) . –53 centimes.

## Exercice 4

Si $x$ a une distribution normale de moyenne 1 et d'écart-type 3, expliciter :

a. Prob[$|x| > 2$].

b. Prob[$x > -1 \mid x < 1{,}5$].

En utilisant la table normale :

a.

$$\begin{aligned} \text{Prob}[|x| > 2] &= 1 - \text{Prob}[|x| \leq 2] \\ &= 1 - \text{Prob}[-2 \leq x \leq 2] \\ &= 1 - \text{Prob}[(-2-1)/3 \leq z \leq (2-1)/3] \\ &= 1 - [F(1/3) - F(-1)] = 1 - 0{,}6306 + 0{,}1587 = 0{,}5281 \end{aligned}$$

b.

$$\begin{aligned} \text{Prob}[x > -1 \mid x < 1.5] &= \text{Prob}[-1 < x < 1.5] / \text{Prob}[x < 1{,}5] \\ \text{Prob}[-1 < x < 1{,}5] &= \text{Prob}[(-1-1)/3 < z < (1{,}5-1)/3)] \\ &= \text{Prob}[z < 1/6] - \text{Prob}[z < -2/3] \\ &= 0{,}5662 - 0{,}2525 = 0{,}3137 \end{aligned}$$

La probabilité conditionnelle est 0,3137 / 0,5662 = 0,5540.

## Exercice 5

Donner approximativement la probabilité qu'une variable aléatoire suivant une distribution chi-deux à 264 degrés de liberté soit inférieure à 297.

$z = [2(297)]^2 - [2(264) - 1]^2 = 1{,}4155$. La probabilité est donc approximativement 0,9215. Avec six chiffres après la virgule, l'approximation est 0,921539 alors que la valeur correcte est 0,921559.

## Exercice 6

**Inégalité de Chebychev**. Pour les distributions de probabilité suivantes, trouver la borne inférieure de la probabilité de l'évènement indiqué en utilisant l'inégalité de Chebychev (3–18). Trouver également la probabilité exacte en utilisant la table appropriée :

a. $x \sim$ normale$[0,3^2]$ et $-4 < x < 4$.

b. $x \sim$ chi-deux, 8 degrés de liberté, $0 < x < 16$.

L'inégalité donnée en (3–18) indique que $\text{Prob}[|x - \mu| \leq k\sigma] \geq 1 - 1 / k^2$. On remarque que le résultat ne donne pas d'informations si $k$ est inférieur ou égal à 1.

a. Le rang est 4 / 3 des écarts-types, de sorte que la borne inférieure est $1 - (3 / 4)^2$ ou 7 / 16 = 0,4375. De la table normale standard, la probabilité réelle est

   $1 - 2\text{Prob}[z < -4 / 3] = 0{,}8175$.

b. La moyenne de la distribution est 8 et l'écart-type est 4. Le rang est donc $\mu \pm 2\sigma$. Selon l'inégalité, la borne inférieure est $1 - (1 / 2)^2 = 0{,}75$. La probabilité réelle est le chi-deux cumulatif (8) à 16, ce qui est un peu plus grand que 0,95. (La valeur réelle est 0,9576.)

## Exercice 7

Étant donné la distribution de probabilité jointe suivante :

| | | X | | |
|---|---|---|---|---|
| | | 0 | 1 | 2 |
| | 0 | 0,05 | 0,1 | 0,03 |
| Y | 1 | 0,21 | 0,11 | 0,19 |
| | 2 | 0,08 | 0,15 | 0,08 |

a. Calculer les probabilités suivantes : Prob[$Y < 2$], Prob[$Y < 2, X > 0$], Prob[$Y = 1, X \geq 1$].

b. Trouver les distributions marginales de $X$ et $Y$.

c. Calculer $E[X]$, $E[Y]$, Var[$X$], Var[$Y$], Cov[$X$, $Y$], et $E[X^2Y^3]$.

d. Calculer Cov[Y, $X^2$].

e. Quelles sont les distributions conditionnelles de $Y$, étant donné que $X = 2$, et celles de $X$ étant donné que $Y > 0$ ?

f. Trouver $E[Y|X]$ et Var[$Y|X$]. Obtenir les deux parties de la décomposition de la variance $\text{Var}[Y] = E_x[\text{Var}[Y|X]] + \text{Var}_x[E[Y|X]]$.

On obtient tout d'abord les probabilités marginales. Pour la distribution jointe, on a :

$$X\text{: } P(0) = 0{,}34,\ P(1) = 0{,}36,\ P(2) = 0{,}30$$

$$Y\text{: } P(0) = 0{,}18,\ P(1) = 0{,}51,\ P(2) = 0{,}31$$

Ensuite :

a.

$$\text{Prob}[Y < 2] = 0{,}18 + 0{,}51 = 0{,}69$$

$$\text{Prob}[Y < 2, X > 0] = 0{,}1 + 0{,}03 + 0{,}11 + 0{,}19 = 0{,}43$$

$$\text{Prob}[Y = 1, X \geq 1] = 0{,}11 + 0{,}19 = 0{,}30$$

b. Elles sont données ci-dessus.

c.

$$E[X] = 0(0{,}34) + 1(0{,}36) + 2(0{,}30) = 0{,}96$$

$$E[Y] = 0(0{,}18) + 1(0{,}51) + 2(0{,}31) = 1{,}13$$

$$E[X^2] = 0^2(0{,}34) + 1^2(0{,}36) + 2^2(0{,}30) = 1{,}56$$

$$E[Y^2] = 0^2(0{,}18) + 1^2(0{,}51) + 2^2(0{,}31) = 1{,}75$$

$$\text{Var}[X] = 1{,}56 - 0{,}96^2 = 0{,}6384$$

$$\text{Var}[Y] = 1{,}75 - 1{,}13^2 = 0{,}4731$$

$$E[XY] = 1(1)(0{,}11)+1(2)(0{,}15)+2(1)(0{,}19)+2(2)(0{,}08) = 1{,}11$$

$$\text{Cov}[X, Y] = 1{,}11 - 0{,}96(1{,}13) = 0{,}0252$$

$$E[X^2Y^3] = 0{,}11 + 8(0{,}15) + 4(0{,}19) + 32(0{,}08) = 4{,}63$$

d.

$$\text{E}[YX^2] = 1(12)0{,}11+1(22)0{,}19+2(12)0{,}15+2(22)0{,}08 = 1{,}81$$

$$\text{Cov}[Y, X^2] = 1{,}81 - 1{,}13(1{,}56) = 0{,}0472$$

e.

$$\text{Prob}[Y = 0 * X = 2] = 0{,}03 / 0{,}3 = 0{,}1$$

$$\text{Prob}[Y = 1 * X = 2] = 0{,}19 / 0{,}3 = 0{,}633$$

$$\text{Prob}[Y = 1 * X = 2] = 0{,}08 / 0{,}3 = 0{,}267$$

$$\text{Prob}[X = 0 * Y > 0] = (0{,}21 + 0{,}08) / (0{,}51 + 0{,}31) = 0{,}3537$$

$$\text{Prob}[X = 1 * Y > 0] = (0{,}11 + 0{,}15) / (0{,}51 + 0{,}31) = 0{,}3171$$

$$\text{Prob}[X = 2 * Y > 0] = (0{,}19 + 0{,}08) / (0{,}51 + 0{,}31) = 0{,}3292$$

f.

$$E[Y * X{=}0] = 0(0{,}05 / 0{,}34)+1(0{,}21 / 0{,}34)+2(0{,}08 / 0{,}34) = 1{,}088$$

$$E[Y^2 * X{=}0] = 1^2(0{,}21 / 0{,}34)+2^2(0{,}08 / 0{,}34) = 1{,}559$$

$$\text{Var}[Y* X{=}0] = 1{,}559 - 1{,}088^2 = 0{,}3751$$

$$E[Y*X{=}1] = 0(0{,}1 / 0{,}36)+1(0{,}11 / 0{,}36)+2(0{,}15 / 0{,}36) = 1{,}139$$

$$E[Y^2*X{=}1] = 1^2(0{,}11 / 0{,}36)+2^2(0{,}15 / 0{,}36) = 1{,}972$$

$$\text{Var}[Y*X{=}1] = 1{,}972 - 1{,}139^2 = 0{,}6749$$

$$E[Y*X{=}2] = 0(0{,}03 / 0{,}30)+1(0{,}19 / 0{,}30)+2(0{,}08 / 0{,}30) = 1{,}167$$

$$E[Y^2*X{=}2] = 1^2(0{,}19 / 0{,}30)+2^2(0{,}08 / 0{,}30) = 1{,}700$$

$$\text{Var}[Y*X{=}2] = 1{,}700 - 1{,}167^2 = 0{,}6749 = 0{,}3381$$

$$E[\text{Var}[Y*X]] = 0{,}34(0{,}3751)+0{,}36(0{,}6749)+0{,}30(0{,}3381) = 0{,}4719$$

$$\text{Var}[E[Y*X]] = 0{,}34(1{,}088^2)+0{,}36(1{,}139^2)+0{,}30(1{,}167^2) - 1{,}13^2$$
$$= 1{,}2781 - 1{,}2769 = 0{,}0012$$

$$E[\text{Var}[Y*X]] + \text{Var}[E[Y*X]] = 0{,}4719 + 0{,}0012 = 0{,}4731 = \text{Var}[Y]$$

## Exercice 8

**Prédicteur du minimum du carré des erreurs à la moyenne**. Pour la distribution jointe de l'exercice 7, calculer $E[y - E[y|x]]^2$. Trouver ensuite le $a$ et le $b$ qui minimisent la fonction $E[y - a - bx]^2$. Étant donné les solutions, vérifier que

$$E[y - E[y|x]]^2 \leq E[y - a - bx]^2.$$

Le résultat est fondamental dans la théorie des moindres carrés.

$$
\begin{array}{llccc}
 & & (x=0) & (x=1) & (x=2) \\
E[y - \mathrm{E}[y|x]]^2 = & (y{=}0) & 0{,}05(0 - 1{,}088)^2 & + 0{,}10(0 - 1{,}139)^2 & + 0{,}03(0 - 1{,}167)^2 \\
 & (y{=}1) & + 0{,}21(1 - 1{,}088)^2 & + 0{,}11(1 - 1{,}139)^2 & + 0{,}19(1 - 1{,}167)^2 \\
 & (y{=}2) & + 0{,}08(2 - 1{,}088)^2 & + 0{,}15(2 - 1{,}139)^2 & + 0{,}08(2 - 1{,}167)^2
\end{array}
$$

$$= 0{,}4719 = E[\mathrm{Var}[y|x]].$$

Les conditions nécessaires pour minimiser la fonction par rapport à $a$ et $b$ sont :

$$\partial E[y - a - bx]^2 / \partial a = 2E\{[y - a - bx](-1)\} = 0$$

$$\partial E[y - a - bx]^2 / \partial b = 2E\{[y - a - bx](-x)\} = 0$$

En divisant d'abord par –2 et en prenant ensuite les espérances, on obtient :

$$E[y] - a - bE[x] = 0$$

$$E[xy] - aE[x] - b\mathrm{E}[x^2] = 0.$$

On résout la première égalité, $a = E[y] - bE[x]$, et on intègre l'expression trouvée dans la seconde pour obtenir :

$$E[xy] - E[x](E[y] - bE[x]) - bE[x^2] = 0$$

ou

$$(E[xy] - E[x]E[y]) = b(E[x^2] - (E[x])^2)$$

ou

$$b = \mathrm{Cov}[x, y] / \mathrm{Var}[x] = -0{,}0708 / 0{,}4731 = -0{,}150$$

et

$$a = E[y] - bE[x] = 1{,}13 - (-0{,}1497)(0{,}96) = 1{,}274$$

La fonction linéaire comparée à la moyenne conditionnelle donne :

$X = 0$

$x = 1$

$x = 2$

| | | | |
|---|---|---|---|
| $E[y\|x]$ | 1,088 | 1,139 | 1,167 |
| $a + bx$ | 1,274 | 1,124 | 0,974 |

On répète le calcul ci-dessus en utilisant $a + bx$ au lieu de $E[y|x]$ :

| | | $(x = 0)$ | $(x = 1)$ | $(x = 2)$ |
|---|---|---|---|---|
| $E[y - a - bx]^2 =$ | $(y = 0)$ | $0{,}05(0 - 1{,}274)^2$ | $+ 0{,}10(0 - 1{,}124)^2$ | $+ 0{,}03(0 - 0{,}974)^2$ |
| | $(y = 1)$ | $+ 0{,}21(1 - 1{,}274)^2$ | $+ 0{,}11(1 - 1{,}124)^2$ | $+ 0{,}19(1 - 0{,}974)^2$ |
| | $(y = 2)$ | $+ 0{,}08(2 - 1{,}274)^2$ | $+ 0{,}15(2 - 1{,}124)^2$ | $+ 0{,}08(2 - 0{,}974)^2$ |

$= 0{,}4950 > 0{,}4719$

## Exercice 9

On suppose que $x$ a une distribution exponentielle, $f(x) = \theta e^{-\theta x}$, $x \geq 0$. Trouver la moyenne, la variance, l'asymétrie et l'aplatissement de $x$. (Indice : les deux derniers sont définis dans la section 3.3.)

Afin de trouver les moments centrés, on utilise les moments bruts :

$$E[x^{r]} = \int_0^\infty \theta x^r e^{-\theta\ x} dx$$

Ils peuvent être obtenus en utilisant l'intégrale gamma. En faisant les substitutions appropriées, on a :

$$E[x^r] = [\theta\Gamma(r + 1)] / \theta^{r+1} = r! / \theta^r$$

Les quatre premiers moments sont : $E[x] = 1 / \theta$, $E[x^2] = 2 / \theta^2$, $E[x^3] = 6 / \theta^3$ et $E[x^4] = 24 / \theta^4$. La moyenne est donc $1 / \theta$ et la variance est $2 / \theta^2 - (1 / \theta)^2 = 1 / \theta^2$. Pour les coefficients d'asymétrie et d'aplatissement, on a :

$$E[x - 1 / \theta]^3 = E[x^3] - 3E[x^2] / \theta + 3E[x] / \theta^2 - 1 / \theta^3 = 2 / \theta^3$$

Le coefficient d'asymétrie normalisé est 2. Le coefficient d'aplatissement est :

$$E[x - 1 / \theta]^4 = E[x^4] - 4E[x^3] / \theta + 6E[x^2] / \theta^2 - 4E[x] / \theta^3 + 1 / \theta^4 = 9 / \theta^4$$

Le coefficient d'aplatissement est donc 6.

## Exercice 10

On suppose que $x$ a la distribution de probabilité discrète suivante :

| X | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Prob$[X = x]$ | 0,1 | 0,2 | 0,4 | 0,3 |

Trouver les moyenne et variance exactes de $X$. On suppose maintenant que $Y = 1 / X$. Trouver les moyenne et variance exactes de $Y$. Trouver les moyenne et variance exactes des approximations linéaire et quadratique de $Y = f(X)$. Est-ce que la moyenne et la variance de l'approximation quadratique sont plus proches de la vraie moyenne que celles de l'approximation linéaire ?

On développe en premier lieu un certain nombre de moments dont on a besoin :

$$E[x] = 0{,}1(1) + 0{,}2(2) + 0{,}4(3) + 0{,}3(4) = 2{,}9 = \mu$$

$$E[x^2] = 0{,}1(1) + 0{,}2(4) + 0{,}4(9) + 0{,}3(16) = 9{,}3$$

$$\mathrm{Var}[x] = 9{,}3 - 2{,}9^2 = 0{,}89 = \sigma^2$$

Pour une utilisation ultérieure, on obtient aussi :

$$E[x - \mu]^3 = 0{,}1(1 - 2{,}9)^3 + \ldots -0{,}432$$

$$E[x - \mu]^4 = 0{,}1(1 - 2{,}9)^4 + \ldots = 1{,}8737.$$

L'approximation est $y = 1 / x$. Les moyenne et variance exactes sont :

$$E[y] = 0{,}1(1) + 0{,}2(1 / 2) + 0{,}4(1 / 3) + 0{,}3(1 / 4) = 0{,}40833$$

$$\mathrm{Var}[y] = 0{,}1(12) + 0{,}2(1 / 4) + 0{,}4(1 / 9) + 0{,}3(1 / 16) - 0{,}40833^2 = 0{,}04645$$

L'approximation linéaire de Taylor en séries autour de $\mu$ est $y \approx 1 / \mu + (-1 / \mu^2)(x - \mu)$. La moyenne de l'approximation linéaire est $1 / \mu = 0{,}3448$ alors que sa variance est $(1 / \mu^4)\mathrm{Var}[x\text{-}\mu] = \sigma^2 / \mu^4 = 0{,}01258$. L'approximation quadratique est :

$$y \approx 1 / \mu + (-1 / \mu^2)(x - \mu) + (1 / 2)(2 / \mu^3)(x - \mu)^2 = 1 / \mu - (1 / \mu^2)(x - \mu) + (1 / \mu^3)(x - \mu)^2$$

La moyenne de cette approximation est $E[y] \approx 1 / \mu + \sigma^2 / \mu^3 = 0{,}3813$, et la variance est approximée par la variance du terme de droite,

$$\begin{aligned}&(1 / \mu^4)\mathrm{Var}[x - \mu] + (1 / \mu^6)\mathrm{Var}[x - \mu]^2 - (2 / \mu^5)\mathrm{Cov}[(x - \mu), (x - \mu)^2] \\ &= (1 / \mu^4)\sigma^2 + (1 / \mu^6)(E[x - \mu]^4 - \sigma^4] - (2 / \mu^5)E[x - \mu]^3 \\ &= 0{,}01498\end{aligned}$$

Aucune approximation ne donne une estimation proche de la variance. On remarque que dans les deux cas, il serait possible d'évaluer simplement les approximations aux quatre valeurs de $x$, et de calculer directement les moyennes et les variances. L'avantage de cette approche est qu'elle peut être appliquée quand il y a beaucoup de valeurs de $x$. Elle est nécessaire quand la distribution de $x$ est continue.

## Exercice 11

**Extrapolation de la table du chi-deux.** Afin d'obtenir un point de pourcentage, entre deux valeurs, dans la table du chi-deux, on extrapole linéairement entre les *réciproques* des degrés de liberté. La distribution du chi-deux est définie pour des valeurs non entières des paramètres des degrés de liberté, mais la table ne contient pas de valeurs critiques pour les valeurs non entières. En utilisant une extrapolation linéaire, trouver la valeur critique à 99 % pour une variable du chi-deux avec un paramètre de degré de liberté égal à 11,3.

Les valeurs critiques à 99 % pour les degrés de liberté 11 et 12 sont 24,725 et 26,217. Pour extrapoler linéairement entre ces deux valeurs afin d'obtenir la valeur correspondant à 11,3 degrés de liberté, on utilise :

$$c = 26{,}217 + \frac{(111{,}3 - 1/12)}{(1/11 - 1/12)}\ (24{,}725 - 26{,}217) = 25{,}2009$$

## Exercice 12

On suppose que $x$ a une distribution normale standard. Quelle est la fonction de densité de la variable aléatoire suivante ?

$$y = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, 0 < y < \frac{1}{\sqrt{2\pi}}$$

[Indice : vous connaissez la distribution de $z = x^2$ en (3-30). Résoudre le problème en termes de $y = g(z)$.]

On sait que $z = x^2$ est distribuée selon un chi-deux à 1 degré de liberté. On cherche la densité de $y = ke^{-z/2}$ où $k = (2\pi)^{-2}$. La transformation inverse est $z = 2\ln k - 2\ln y$, de sorte que le jacobien est $|-2 / y| = 2 / y$. La densité de $z$ est celle d'une gamma de paramètres 1 / 2 et 1 / 2. Ainsi :

$$f(z) = \frac{(1/2)^{1/2}}{\Gamma(1/2)} e^{-z/2} z^{-1/2}, z > 0$$

On remarque que $\Gamma(1 / 2) = \sqrt{\pi}$ . En faisant les substitutions pour $z$ et en multipliant par le jacobien, on obtient :

$$f(y) = \frac{(1/2)^{1/2}}{\Gamma(1/2)} \frac{2}{y} e^{(-1/2)(2\ln k - 2\ln y)} (2\ln k - 2\ln y)^{-1/2}$$

Le terme exponentiel se réduit à $y / k$. Le facteur d'échelle est égal à $2k / y$. Donc, la densité est simplement :

$$f(y) = 2(2\ln k - 2\ln y)^{-1/2} = \sqrt{2}\,(\ln k - \ln y)^{-1/2} = \{2 / [\ln(1 / (y(2\pi)^{1/2}))]\},\ 0 < y < (2\pi)^{-1/2}$$

## Exercice 13

**La transformation de probabilité fondamentale**. On suppose que la variable aléatoire continue $x$ a une distribution cumulative $F(x)$. Quelle est la distribution de probabilité de la variable aléatoire $y = F(x)$ ? (Remarque : ce résultat forme la base de la simulation de tirages à partir de nombreuses distributions continues.)

La transformation inverse est $x(y) = F^{-1}(y)$, de sorte que le jacobien est $dx / dy = F^{-1\prime}(y) = 1 / f(x(y))$ où $f(.)$ est la densité de $x$. La densité de $y$ est $f(y) = f\,[F^{-1}(y)] \times 1 / f\,(x(y)) = 1,\ 0 \leq y \leq 1$. Ainsi, $y$ a une distribution continue uniforme. On remarque alors que pour obtenir un échantillon aléatoire de la distribution, on peut tirer un échantillon $y_1$, ..., $y_n$ de la distribution de $y$, et ensuite obtenir $x_1 = x_1(y_1)$, ... $x_n = x_n(y_n)$.

## Exercice 14

**Générateurs de nombres aléatoires**. On suppose que $x$ est distribuée uniformément entre 0 et1, de sorte que $f(x) = 1,\ 0 \leq x \leq 1$. On pose $\theta$ une constante positive. Quelle est la densité de $y = -(1 / \theta)\ln x$ ? (Indice : voir la section 3.5.) Est-ce que cela suggère un moyen de simuler des tirages de cette distribution si l'on a un générateur de nombres aléatoires qui donne les tirages de la distribution uniforme ? Pour continuer, suggérer un moyen de simuler des tirages d'une distribution logistique, $f(x) = e^{-x} / (1+e^{-x})^2$.

La transformation inverse est $x = e^{-\theta y}$, de sorte que le jacobien est $dx / dy = \theta e^{-\theta y}$. Comme $f(x) = 1$, ce jacobien est aussi la densité de $y$. On peut simuler des tirages $y$ de toute distribution exponentielle de paramètre $\theta$ en tirant des observations $x$ de la distribution uniforme et en calculant $y = -(1 / \theta)\ln x$. De la même façon, pour la distribution logistique, la fonction cumulative est $F(x) = 1 / (1 + e^{-x})$. Ainsi, des tirages de $y$ d'une distribution uniforme peuvent être vus comme des tirages de $F(x)$. On peut alors obtenir $x$ comme $x = \ln[F(x) / (1 - F(x)] = \ln[y / (1 - y)]$.

# Exercice 15

On suppose que $x_1$ et $x_2$ sont distribuées selon des lois normales standard indépendantes. Quelle est la distribution jointe de $y_1 = 2 + 3x_1 + 2x_2$ et $y_2 = 4 + 5x_1$ ? On suppose que vous êtes capables d'obtenir deux échantillons d'observations de distributions normales standard. Comment obtenir un échantillon d'observations d'une distribution normale bivariée de moyennes 1 et 2, de variances 4 et 9 et de covariance 3 ?

On peut écrire la paire de transformations comme :

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix} + \begin{bmatrix} 3 & 2 \\ 5 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathbf{b} + \mathbf{Ax}$$

Comme $\mathbf{x} \sim N[\mathbf{0}, \mathbf{I}]$, on a donc $\mathbf{y} \sim N[\mathbf{b} + \mathbf{A0}, \mathbf{AIAN}]$ où :

$$E[\mathbf{y}] = \mathbf{b} + \mathbf{A0} = \mathbf{b} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}, \ \mathrm{Var}[\mathbf{y}] = \mathbf{AA}' = \begin{bmatrix} 13 & 15 \\ 15 & 25 \end{bmatrix}$$

Pour la seconde partie du problème, en utilisant les résultats ci-dessus, on a besoin de **A** et **b** tels que

$\mathbf{b} + \mathbf{A0} = (1,2)'$ et $\mathbf{AA}' = \begin{bmatrix} 4 & 3 \\ 3 & 9 \end{bmatrix}$. Le vecteur est simplement $\mathbf{b} = (1,2)'$. Afin de trouver les éléments de **A**, plusieurs étapes sont à réaliser. La factorisation de Cholesky utilisée dans l'exercice 9 est probablement la plus simple. On pose $y_1 = 1 + 2x_1$. Ainsi, $y_1$ a une moyenne 1 et une variance 4, comme demandé. On pose maintenant $y_2 = 2 + w_1x_1 + w_2x_2$. La covariance entre $y_1$ et $y_2$ est $2w_1$, puisque $x_1$ et $x_2$ ne sont pas corrélées. Ainsi, $2w_1 = 3$, ou $w_1 = 1{,}5$. On obtient également $\mathrm{Var}[y_2] = w_1^2 + w_2^2 = 9$, de sorte que $w_2^2 = 9 - 1{,}5^2 = 6{,}75$. La matrice de transformation est donc $\mathbf{A} = \begin{bmatrix} 2 & 0 \\ 1{,}5 & 2{,}598 \end{bmatrix}$. C'est la factorisation de Cholesky de $\mathbf{AA}'$ que l'on désire. Cela donne une méthode simple pour trouver la matrice **A** requise pour un certain nombre de variables. Une méthode alternative serait d'utiliser les racines et les vecteurs caractéristiques de $\mathbf{AA}'$.

# Exercice 16

La densité de la distribution normale standard, notée $\phi(x)$, est donnée en (3-28). La fonction fondée sur la $i$-ième dérivée de la densité, désignée par $H_i = [(-1)^i d^i\phi(x) / dx^i] / \phi(x)$, $i = 0, 1, 2, \ldots$ est appelée *le polynôme de Hermite.* Par définition, $H_0 = 1$.

a. Trouver les trois polynômes de Hermite suivants.

b. Une équation utile dans ce contexte est l'équation différentielle suivante :

$$d^r\phi(x) / dx^r + xd^{r-1}\phi(x) / dx^{r-1} + (r-1)d^{r-2}\phi(x) / dx^{r-2} = 0$$

Utiliser ce résultat et celui de la première question pour trouver $H_4$ et $H_5$.

Le résultat important à utiliser dans les dérivations est $d\phi(x) / \mathrm{d}x = -x\phi(x)$. Donc,

$$d^2\phi(x) / dx^2 = (x^2 - 1)\phi(x)$$

et

$$d^3\phi(x) / dx^3 = (3x - x^3)\phi(x)$$

Les polynômes sont $H_1 = x$, $H_2 = x^2 - 1$, et $H_3 = x^3 - 3x$.

Pour la question (b), on résout :

$$d^r\phi(x) / dx^r = -xd^{r-1}\phi(x) / dx^{r-1} - (r-1)d^{r-2}\phi(x) / dx^{r-2}.$$

Par conséquent,

$$d^4\phi(x) / dx^4 = -x(3x - x^3)\phi(x) - 3(x^2 - 1)\phi(x) = (x^4 - 6x^2 + 3)\phi(x)$$

et

$$d^5\phi(x) / dx^5 = (-x^5 + 10x^3 - 15x)\phi(x)$$

Ainsi,

$$H_4 = x^4 - 6x^2 + 3 \text{ et } H_5 = x^5 - 10x^3 + 15x$$

## Exercice 17

**Polynômes orthogonaux**. Les polynômes de Hermite sont orthogonaux si $x$ a une distribution normale standard. C'est-à-dire $E[H_iH_j] = 0$ si i ≠ j. Le prouver pour les $H_1$, $H_2$, et $H_3$ obtenus ci-dessus.

$$E[H_1(x)H_2(x)] = E[x(x^2 - 1)] = E[x^3 - x] = 0$$

Puisque la distribution normale est symétrique, alors :

$$E[H_1(x)H_3(x)] = E[x(x^3 - 3x)] = E[x^4 - 3x^2] = 0$$

Le quatrième moment de la distribution normale standard est 3 fois la variance. Finalement,

$$E[H_2(x)H_3(x)] = E[(x^2 - 1)(x^3 - 3x)] = E[x^5 - 4x^3 + 3x] = 0$$

parce que tous les moments d'ordre impair de la distribution normale sont nuls. (Le résultat général permettant d'étendre ce qui vient d'être trouvé est que dans un produit de polynômes de Hermite, si la somme des indices est impaire (respectivement paire), le produit est une somme de puissances impaires (respectivement paires) de $x$. Cela donne une méthode permettant déterminer les moments plus hauts de la distribution normale s'ils sont nécessaires. (Par exemple, $E[H_1H_3] = 0$ implique que $E[x^4] = 3E[x^2]$.)

## Exercice 18

Si $x$ et $y$ ont pour moyennes $\mu_x$ et $\mu_y$, pour variances $\sigma_x^2$ et $\sigma_y^2$ et pour covariance $\sigma_{xy}$, quelle est l'approximation de la matrice de covariance des deux variables aléatoires $f_1 = x / y$ et $f_2 = xy$ ?

Les éléments de $\mathbf{J\Sigma J}'$ sont :

$$1{,}1 = \frac{\sigma_x^2}{\mu_y^2} + \frac{\sigma_y^2\mu_2^x}{\mu_y^4} - \frac{2\sigma_{xy}\mu_x}{\mu_y^3}$$

$$(1{,}2) = \sigma_x^2 - \sigma_y^2\ \mu_x^2\ /\ \mu_y^4$$

$$(2{,}2) = \sigma_x^2\ \mu_y^4\ + \sigma_y^2\ \mu_x^2\ + 2\sigma_{xy}\mu_x\mu_y$$

## Exercice 19

**Moments factoriels.** Pour trouver les moments d'une distribution de Poisson par exemple, le moment factoriel est un dispositif utile. (La distribution de Poisson est donnée dans l'exemple 1 du chapitre 3) La densité est :

$$f(x) = e^{-\lambda}\lambda^x / x!,\ x = 0, 1, 2, ...$$

Pour trouver la moyenne, on peut utiliser :

$$E[x] = \textstyle\sum_{x=0}^{\infty} xf(x) = \sum_{x=0}^{\infty} xe^{-\lambda}\lambda^x / x!$$

$$= \textstyle\sum_{x=1}^{\infty} e^{-\lambda}\lambda^{x-1}/(x-1)!$$

$$= \lambda \textstyle\sum_{y=0}^{\infty} e^{-\lambda}\lambda^y / y!$$

$$= \lambda$$

car la somme des probabilités est égale à 1. Pour trouver la variance, on étend cette méthode en trouvant $E[x(x-1)]$, et on procède de la même façon pour les autres moments. Utiliser cette méthode pour trouver la variance et le troisième moment centré de la distribution de Poisson. (Remarque : cette procédure est utilisée pour transformer le factoriel dans le dénominateur de la probabilité.)

En utilisant la même technique,

$$E[x(x–1)] = \sum_{x=0}^{\infty} x(x-1) f(x) = \sum_{x=0}^{\infty} x(x-1)e^{-\lambda}\lambda^{x} / x!$$

$$= \sum_{x=2}^{\infty} e^{-\lambda}\lambda^{x-2} /(x-2)!$$

$$= \lambda^2 \sum_{y=0}^{\infty} e^{-\lambda}\lambda^{y} / y!$$

$$= \lambda^2$$

$$= \mathrm{E}[x^2] - \mathrm{E}[x]$$

Ainsi,

$$E[x^2] = \lambda^2 + \lambda$$

Comme $E[x] = \lambda$, il s'ensuit que $\mathrm{Var}[x] = (\lambda^2 + \lambda) - \lambda^2 = \lambda$. Suivant la même méthode, ce qui précède donne :

$$E[x(x-1)(x-2)] \qquad = E[x^3] - 3E[x^2] + 2E[x] = \lambda^3$$

Donc,

$$E[x^3] = \lambda^3 + 3(\lambda + \lambda^2) - 2\lambda = \lambda^3 + 3\lambda^2 + \lambda$$

Puis,

$$E[x - E[x]]^3 = E[x^3] - 3\lambda E[x^2] + 3\lambda^2 E[x] - \lambda^{3} = \lambda$$

## Exercice 20

Si $x$ a une distribution normale de moyenne μ et d'écart-type σ, quelle est la distribution de probabilité de $y = e^x$ ?

Si $y = e^x$, alors $x = \ln y$ et le jacobien est $dx / dy = 1 / y$. En faisant la substitution,

$$f(y) = \frac{1}{\sigma y\sqrt{2\pi}} e^{-\frac{1}{2}[(\ln y - \mu)/\sigma]^2}$$

C'est la densité de la distribution log-normale.

# Exercice 21

Si $y$ a une distribution log-normale, quelle est la distribution de probabilité de $y^2$ ?

On pose $z = y^2$. Alors, $y = \sqrt{z}$ et d$y$ / dz = 1 / (2$\sqrt{z}$ ). En les insérant dans la densité ci-dessus, on trouve :

$$f(z) = \frac{1}{\sigma\sqrt{2\pi}}\frac{1}{\sqrt{z}}\frac{1}{2\sqrt{z}}e^{-\frac{1}{2}\left[\left(\frac{1}{2}\ln z-\mu\right)/\sigma\right]^2}, z > 0$$

$$= \frac{1}{(2\sigma)z\sqrt{2\pi}}e^{-\frac{1}{2}[(\ln z-2\mu)/(2\sigma)]^2}, z > 0$$

Ainsi, $z$ a une distribution log-normale de paramètres 2μ et 2σ. En généralisant, si $y$ a une distribution log-normale de paramètres μ et σ, $y^r$ a une distribution log-normale de paramètres $r$μ et $r$σ.

# Exercice 22

On suppose que $y$, $x_1$, et $x_2$ ont une distribution jointe normale de paramètre $\boldsymbol{\mu}' = [1, 2, 4]$ et de matrice de covariance $\boldsymbol{\Sigma} = \begin{bmatrix} 2 & 3 & 1 \\ 3 & 5 & 2 \\ 1 & 2 & 6 \end{bmatrix}$.

a. Calculer l'ordonnée à l'origine et la pente de la fonction $E[y{*}x_1]$, Var[$y{*}x_1$], et le coefficient de détermination dans la régression.

b. Calculer l'ordonnée à l'origine et la pente de la fonction de la moyenne conditionnelle, $E[y{*}x_1, x_2]$. Qu'est-ce que $E[y{*}x_{1=}2{,}5, x_{2=}3{,}3]$ ? Qu'est-ce que Var[$y{*}x_{1=}2{,}5, x_{2=}3{,}3$] ?

En premier lieu, pour les variables distribuées normalement, on a :

$$E[y{*}\mathbf{x}] = \mu_y + \mathrm{Cov}[y, \mathbf{x}]\{\mathrm{Var}[\mathbf{x}]\}^{-1}(\mathbf{x} - \mathbf{:}_x)$$

et

$$\mathrm{Var}[y{*}\mathbf{x}] \qquad = \mathrm{Var}[y] - \mathrm{Cov}[y, \mathbf{x}]\{\mathrm{Var}[\mathbf{x}]\}^{-1}\mathrm{Cov}[\mathbf{x}, y]$$

et

$$COD = \mathrm{Var}[E[y{*}\mathbf{x}]] \,/\, \mathrm{Var}[y]$$

$$= \mathrm{Cov}[y, \mathbf{x}]\{\mathrm{Var}[\mathbf{x}]\}^{-1}\mathrm{Cov}[\mathbf{x}, y] \,/\, \mathrm{Var}[y]$$

On insère juste les chiffres ci-dessus pour obtenir les résultats.

$$E[y*x_1] = 1 + (3 / 5)(x_1 - 2) = -0{,}2 + 0{,}6x_1$$

$$\mathrm{Var}[y*x_1] = 2 - 3(1 / 5)3 = 1 / 5 = 0{,}2$$

$$COD = 0{,}6^2(5) / 2 = 0{,}9$$

$$\mathrm{E}[y*x_1, x_2] = 1 + \begin{bmatrix} 3 & 1 \end{bmatrix} \begin{bmatrix} 5 & 2 \\ 2 & 6 \end{bmatrix}^{-1} \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$

$$= -0{,}4615 + 0{,}6154x_1 - 0{,}03846x_2$$

$$\mathrm{Var}[y*x_1, x_2] = 2 - (0{,}6154, -0{,}03846)(3, 1)' = 0{,}1923$$

$$E[y*x_1{=}2{,}5,\ x_2{=}3{,}3] = 1{,}3017$$

La variance conditionnelle n'est pas une fonction de $x_1$ ou $x_2$.

## Exercice 23

Quelle est la densité de $y = 1 / x$ si $x$ a une distribution chi-deux ?

La densité d'une variable chi-deux est une variable gamma de paramètres 1 / 2 et $n$ / 2 avec $n$ les degrés de liberté de la variable chi-deux. Ainsi,

$$f(x) = \frac{(1/2)^{n/2}}{\Gamma(n/2)} e^{-\frac{1}{2}x} x^{\frac{n}{2}-1}, x > 0$$

Si $y = 1 / x$ alors $x = 1 / y$ et $|dx / dy| = 1 / y^2$. Donc, après avoir multiplié par le jacobien,

$$f(y) = \frac{(1/2)^{n/2}}{\Gamma(n/2)} e^{-\frac{1}{2y}} \left(\frac{1}{y}\right)^{\frac{n}{2}+1}, y > 0$$

## Exercice 24

**Fonction génératrice des probabilités**. Pour une variable aléatoire discrète, $x$, la fonction $E[t^x] = \sum_{x=0}^{\infty} t^x \mathrm{Prob}[X = x]$ est appelée la *fonction génératrice des probabilités* parce que, dans la fonction, le coefficient de $t^i$ est $\mathrm{Prob}[X{=}i]$. Supposer que $x$ soit le nombre de répétitions d'une expérience jusqu'à ce que le premier succès ait lieu avec une probabilité de succès égale à $\pi$. La densité de $x$ est la *distribution géométrique* :

$$\mathrm{Prob}[X = x] = (1 - \pi)^{x-1}\pi$$

Quelle est la fonction génératrice des probabilités ?

$$E[t^x] = \sum_{x=0}^{\infty} t^x (1-\pi)^{x-1}\pi$$

$$= \frac{\pi}{(1-\pi)} \sum_{x=0}^{\infty} [t(1-\pi)]^x$$

$$= \frac{\pi}{(1-\pi)} \frac{1}{1-t(1-\pi)}$$

## Exercice 25

**Fonction génératrice des moments**. Pour la variable aléatoire $X$, avec une fonction de densité $f(x)$, si la fonction $M(t) = E[e^{tx}]$ existe, il s'agit de la fonction génératrice des moments. En supposant qu'elle existe, on peut montrer que $d^rM(t) / dt^r|_{t=0} = E[x^r]$. Trouver la fonction génératrice des moments pour :

a. La distribution exponentielle de l'exercice 9.

b. La distribution de Poisson de l'exercice 19.

Pour la variable continue de la question (a), $f(x) = \theta\exp(-\theta x)$, $M(t) = \int_0^{\infty} e^{tx}\theta e^{-\theta\ x}dx = \int_0^{\infty} \theta e^{-(\theta\ -t)x}dx$.

C'est θ fois une intégrale gamma avec $p = 1$, $c = 1$, et $a = (\theta - t)$. Donc,

$$M(t) = \theta / (\theta - t)$$

Pour la distribution de Poisson :

$$M(t) = \sum_{x=0}^{\infty} e^{tx} e^{-\lambda}\lambda^x / x! = \sum_{x=0}^{\infty} e^{-\lambda} (\lambda e^t)^x / x!$$

$$= \sum_{x=0}^{\infty} e^{-\lambda} e^{\lambda e^t} e^{-\lambda e^t} (\lambda e^t)^x / x!$$

$$= e^{-\lambda+\lambda e^t} \sum_{x=0}^{\infty} e^{-\lambda e^t} (\lambda e^t)^x / x!$$

La somme est la somme de probabilités d'une distribution de Poisson, de paramètre $\lambda e^t$, qui est égale à 1, de sorte que le terme avant le signe de la somme est la fonction génératrice des moments, $M(t) = \exp[\lambda(e^t - 1)]$.

# Annexe C

## Exercice 1

L'échantillon qui suit a été tiré d'une distribution normale de moyenne μ et d'écart-type σ :

$$x = 1{,}3,\ 2{,}1,\ 0{,}4,\ 1{,}3,\ 0{,}5,\ ,2,\ 1{,}8,\ 2{,}5,\ 1{,}9,\ 3{,}2$$

Calculer la moyenne, la médiane, la variance et l'écart-type de l'échantillon.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = 1{,}52$$

$$s^2 = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n-1} = 0{,}9418$$

$$s = 0{,}97$$

Médiane = 1,55, à mi-chemin entre 1,3 et 1,8.

## Exercice 2

En utilisant les données de l'exercice précédent, tester les hypothèses suivantes :

a. $\mu > 2$

b. $\mu < 0{,}7$

c. $\sigma^2 = 0{,}5$

d. En utilisant un test du ratio de vraisemblance, tester l'hypothèse suivante : $\mu = 1{,}8$, $\sigma^2 = 0{,}8$.

a. On rejette l'hypothèse si 1,52 est trop petit relativement à la valeur supposée de 2. Puisque les données sont tirées d'une distribution normale, on peut utiliser un *t* test pour tester l'hypothèse. Le *t*-ratio est :

$$t[9] = (1{,}52 - 2) / [0{,}97 / \sqrt{10}] = -1{,}472$$

La valeur critique à 95 % de la *t* distribution pour un test unilatéral est –1,833. Donc, on ne rejette pas l'hypothèse au seuil de 95 %.

b. On rejette l'hypothèse si 1,52 est excessivement large relativement à la moyenne supposée de 0,7. Le *t*-ratio est $t[9] = (1{,}52 - 0{,}7) / [0{,}97 / \sqrt{10}] = 2{,}673$. En utilisant la même valeur critique que dans le problème précédent, on rejette cette hypothèse.

c. La statistique $(n - 1)s^2 / \sigma^2$ est distribuée selon un chi-deux à 9 degrés de liberté. Elle est égale à 9(0,94) / 0,5 = 16,920. Les valeurs critiques au seuil de 95 % de la table du chi-deux pour un test bilatéral sont 2,70 et 19,02. Ainsi, on ne rejette pas l'hypothèse.

d. La log-vraisemblance pour un échantillon d'une distribution normale est :

$$\ln L = -(n / 2)\ln(2\pi) - (n / 2)\ln\sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{n}(x_i - \mu)^2$$

Les valeurs de l'échantillon sont :

$$\hat{\mu} = \bar{x} = 1,52, \quad \hat{\sigma}^2 = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n} = 0,8476$$

La log-vraisemblance maximisée pour l'échantillon est –13,363. Un raccourci utile pour calculer la log-vraisemblance aux valeurs supposées est $\sum_{i=1}^{n}(x_i - \mu)^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2 + n\left(\bar{x} - \mu\right)^2$. Pour la valeur supposée $\mu$ = 1,8, on a $\sum_{i=1}^{n}(x_i - 1,8)^2 = 9,26$.

La log-vraisemblance est $-5(\ln(2\pi) - 5(\ln 0,8) - (1 / 1,6)9,26 = -13,861$. La statistique du ratio de vraisemblance est $-2(\ln L_r - \ln L_u) = 0,996$. La valeur critique pour un chi-deux à 2 degrés de liberté est 5,99. On ne rejette pas l'hypothèse.

## Exercice 3

Supposer que l'échantillon suivant est tiré d'une distribution normale de moyenne $\mu$ et d'écart-type $\sigma$ : $y$ = 3,1, –0,1, 0,3, 1,4, 2,9, 0,3, 2,2, 1,5, 4,2, 0,4. Tester l'hypothèse selon laquelle la moyenne de la distribution qui a généré ces données est la même que celle qui a produit les données de l'exercice 1. Tester l'hypothèse en supposant que les variances sont les mêmes. Tester l'hypothèse selon laquelle les variances sont les mêmes en utilisant un $F$ test et un test du ratio de vraisemblance. (Ne pas faire l'hypothèse que les moyennes sont les mêmes.)

Si les variances sont les mêmes,

$$\bar{x}_1 \sim N[\mu_1, \sigma_1^2 / n_1] \text{ et } \bar{x}_2 \sim N[\mu_2, \sigma_2^2 / n_2]$$

$$\bar{x}_1 - \bar{x}_2 \sim N[\mu_1 - \mu_2, \sigma^2\{(1/n_1) + (1/n_2)\}],$$

$$(n_1 - 1)s_1^2 / \sigma^2 \sim \chi^2[n_1 - 1] \text{ et } (n_2 - 1)s_2^2 / \sigma^2 \sim \chi^2[n_2 - 1]$$

$$(n_1 - 1)s_1^2 / \sigma^2 + (n_2 - 1)s_2^2 / \sigma^2 \sim \chi^2[n_1 + n_2 - 2]$$

Ainsi, la statistique

$$t = \frac{\left\{\left(\overline{x_1} - \overline{x_2}\right) - \left(\mu_1 - \mu_2\right)\right\} / \sqrt{\sigma^2\left[(1/n_1) + (1/n_2)\right]}}{\sqrt{\left\{(n_1 - 1)s_1^2 / \sigma^2 + (n_2 - 1)s_2^2 / \sigma^2\right\} / (n_1 + n_2 - 2)}}$$

est le ratio d'une variable normale standard et de la racine carrée d'une variable chi-deux divisée par ses degrés de liberté. Cette statistique suit une distribution $t$ avec $n_1 + n_2 - 2$ degrés de liberté. Sous l'hypothèse selon laquelle les moyennes sont égales, la statistique est :

$$t = \frac{\left(\overline{x_1} - \overline{x_2}\right) / \sqrt{(1/n_1) + (1/n_2)}}{\sqrt{\left\{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2\right\} / (n_1 + n_2 - 2)}}$$

Les statistiques d'échantillon sont :

$$n_1 = 10,\ \overline{x}_1 = 1{,}52,\ s_1^2 = 0{,}9418$$

$$n_2 = 10,\ \overline{x}_2 = 1{,}62,\ s_2^2 = 2{,}0907$$

de sorte que $t[18] = 0{,}1816$. Cette valeur est très faible, de sorte qu'on ne rejette pas l'hypothèse de moyennes égales.

Pour un échantillon aléatoire tiré de deux distributions normales, sous l'hypothèse de variances égales, la statistique $F[n_1 - 1, n_2 - 1] = \dfrac{\left[(n_1 - 1)s_1^2 / \sigma^2\right] / (n_1 - 1)}{\left[(n_2 - 1)s_2^2 / \sigma^2\right] / (n_2 - 1)}$ est le ratio de deux variables chi-deux indépendantes, chacune étant divisée par leurs degrés de liberté. Il s'agit d'une $F$ distribution avec $n_1 - 1$ et $n_2 - 1$ degrés de liberté. Si $n_1 = n_2$, la statistique se réduit à $F[n_1 - 1, n_2 - 1] = s_1^2 / s_2^2$. Étant donné l'objectif assigné, il est plus pratique de placer la variance la plus grande dans le dénominateur. Ainsi, pour les données d'échantillon, $F[9, 9] = 2{,}0907 / 0{,}9418 = 2{,}2199$. La valeur critique au seuil de 95 % d'une $F$ table est 3,18. Ainsi, on ne rejette pas l'hypothèse de variances égales.

Le test du ratio de vraisemblance est fondé sur la statistique de test $\lambda = -2(\ln L_r - \ln L_u)$. La log-vraisemblance pour l'échantillon, joint de 20 observations, est la somme de deux log-vraisemblances séparées si les échantillons sont supposés indépendants. L'insertion des estimations de la vraisemblance maximum est un raccourci utile pour calculer la log-vraisemblance. À l'estimation de la vraisemblance maximum, $\ln L = (-n / 2)[1 + \ln(2\pi) + \ln \hat{\sigma}^2]$. Ainsi, la log-vraisemblance pour l'échantillon est $\ln L_2 = (-5 / 2)[1 + \ln(2\pi) + \ln((9 / 10)2{,}0907)] = -17{,}35007$. (On rappelle qu'on ne corrige pas les degrés de liberté pour l'estimateur de la variance.). La fonction de log-vraisemblance non contrainte est donc –13,363 + (–17,35001) = –30,713077. Pour calculer la fonction de log-vraisemblance contrainte, on a besoin de l'estimateur groupé qui ne suppose pas que les moyennes sont identiques. Il s'agit de :

$$\hat{\sigma}^2 = [(n_1 - 1)\, s_1^2 + (n_2 - 1)\, s_2^2\,] \,/\, [n_1 + n_2]$$
$$= [9(0{,}9418) + 9(2{,}0907)] \,/\, 20 = 1{,}36463$$

Ainsi, la log-vraisemblance contrainte est :

$$\ln L_r = (-20 \,/\, 2)[1 + \ln(2\pi) + \ln(1{,}36463)] = -31{,}4876.$$

L'opposé du double de la différence est $\lambda = -2[-31{,}4876 - (-30{,}713077)] = 1{,}541$ qui est distribué comme un chi-deux à 1 degré de liberté. La valeur critique est 3,84 On ne rejette donc pas l'hypothèse.

## Exercice 4

Une méthode courante pour simuler des tirages aléatoires d'une distribution normale standard est de calculer la somme de 12 tirages d'une distribution uniforme [0,1] et de soustraire 6. Peut-on justifier cette procédure ?

La distribution uniforme a pour moyenne 2 et pour variance 1 / 12. Par conséquent, la statistique $12(\bar{x} - 1/2) = \sum_{i=1}^{12} x_i - 6$ est équivalente à $z = \sqrt{n}\,(\bar{x} - \mu)/\sigma$. Lorsque $n \to \infty$, elle converge vers une variable normale standard. L'expérience suggère qu'un échantillon de 12 est suffisamment important pour approximer ce résultat. Néanmoins, les générateurs de nombres aléatoires, développés plus récemment, utilisent généralement des procédures différentes fondées sur l'erreur de troncature qui se produit dans la représentation des nombres réels dans un calculateur numérique.

## Exercice 5

En s'appuyant sur un échantillon de 65 observations d'une distribution normale, on obtient une *médiane* de 34 et un écart-type de 13,3. Établir un intervalle de confiance pour la moyenne. (Indice : utiliser la distribution asymptotique.) Comparer l'intervalle de confiance trouvé à celui qui aurait été obtenu si 34 avait été la moyenne de l'échantillon et pas la médiane.

La variance de la médiane est $\pi\sigma^2/(2n)$. En utilisant la distribution normale asymptotique à la place de la $t$ distribution, l'intervalle de confiance est $34 - 1{,}96(13{,}3^2\pi / 130)^2 \le \mu \le 34 + 1{,}96(13{,}3^2\pi / 130)^2$ ou $29{,}95 \le \mu \le 38{,}052$. Si l'estimateur avait été la moyenne au lieu de la médiane, la variance asymptotique appropriée aurait été $\sigma^2 / n$, qu'on estimerait avec $13{,}3^2 / 65 = 2{,}72$ comparé à 4,274 pour la médiane. L'intervalle de confiance serait (30,77, 37,24), ce qui est un peu plus étroit.

## Exercice 6

La variable aléatoire $x$ a une distribution continue $f(x)$ et une fonction de répartition $F(x)$. Quelle est la distribution de probabilité du maximum de l'échantillon ? (Indice : dans un échantillon aléatoire de $n$ observations, $x_1, x_2, ..., x_n$, si $z$ est le maximum, alors chaque observation dans l'échantillon est inférieure ou égale à $z$. Utiliser une fonction de densité.

Si $z$ est le maximum, alors chaque observation de l'échantillon est inférieure ou égale à z. La probabilité de cet évènement est $\text{Prob}[x_1 \leq z, x_2 \leq z, ..., x_n \leq z] = F(z)F(z)...F(z) = [F(z)]^n$. La densité est la dérivée, $n[F(z)]^{n-1}f(z)$.

## Exercice 7

On suppose que la distribution de $x$ est $f(x) = 1 / \theta,\ 0 \leq x \leq \theta$. Dans un échantillon aléatoire tiré de cette distribution, prouver que le maximum de l'échantillon est un estimateur convergent de θ. Remarque : on peut prouver que le maximum est l'estimateur de θ de la vraisemblance maximum. Mais, les propriétés habituelles ne s'appliquent pas. Pourquoi ? (Indice : essayer de vérifier que la dérivée première espérée de la log-vraisemblance par rapport à θ est zéro.)

En utilisant le résultat du problème précèdent, la densité du maximum est :

$$n[z / \theta]^{n-1}(1 / \theta),\ 0 < z < \theta$$

Donc, la valeur espérée est $E[z] = \int_0^\theta z^n dz = [\theta^{n+1} / (n+1)][n / \theta^n] = n\theta / (n + 1)$. On trouve la variance de façon similaire, $E[z^2] = \int_0^\theta z^2 n(z / n)^{n-1}(1 / \theta)dz = n\theta^2 / (n+2)$ de sorte que $\text{Var}[z] = E[z^2] - (E[z])^2 = n\theta^2 / [(n + 1)^2(n + 2)]$. En utilisant la convergence en moyenne quadratique, on voit que $\lim_{n\to\infty} E[z] = \theta$ et $\lim_{n\to\infty} \text{Var}[z] = 0$, ainsi plim $z = \theta$.

## Exercice 8

Dans un échantillon aléatoire tiré d'une distribution exponentielle, $f(x) = \frac{1}{\theta} e^{\frac{-x}{\theta}}$, $x > 0$, $\theta > 0$, trouver l'estimateur de la vraisemblance maximum de θ et établir la distribution asymptotique de cet estimateur.

La log-vraisemblance est $\ln L = -n\ln\theta - (1 / \theta)\sum_{i=1}^{n} x_i$. L'estimateur de la vraisemblance maximum est obtenu comme la solution de $\partial \ln L / \partial\theta = -n / \theta + (1 / \theta^2) \sum_{i=1}^{n} x_i = 0$, ou

$\hat{\theta}_{ML} = (1 / n) \sum_{i=1}^{n} x_i = \bar{x}$. La variance asymptotique de l'EMV est $\{-E[\partial^2 \ln L / \partial\theta^2]\}^{-1}$ $= \{-E[n / \theta^2 - (2 / \theta^3) \sum_{i=1}^{n} x_i ]\}^{-1}$. Pour trouver la valeur espérée de cette variable aléatoire, on a besoin de $E[x_i] = \theta$. Par conséquent, la variance asymptotique est $\theta^2 / n$. La distribution asymptotique est normale avec une moyenne $\theta$ et cette variance.

## Exercice 9

On considère un échantillon de 500 observations tirées d'une distribution normale de moyenne $\mu$ et d'écart-type $\sigma$ dans lequel 35 % des observations sont inférieures à 2,1 et que 55 % des observations sont inférieures à 3,6. Estimer $\mu$ et $\sigma$.

Si 35 % des observations sont inférieures à 2,1, on en déduit que :

$$\Phi[(2{,}1 - \mu) / \sigma] = 0{,}35, \text{ ou } (2{,}1 - \mu) / \sigma = -0{,}385 \Rightarrow 2{,}1 - \mu = -0{,}385\sigma$$

De la même façon :

$$\Phi[(3{,}6 - \mu) / \sigma] = 0{,}55, \text{ ou } (3{,}6 - \mu) / \sigma = 0{,}126 \Rightarrow 3{,}6 - \mu = 0{,}126\sigma$$

La solution jointe est $\hat{\mu} = 3{,}2301$ et $\hat{\sigma} = 2{,}9354$. Cela peut ne pas sembler évident, mais on peut dériver aussi les écarts-types asymptotiques de ces estimations en les construisant comme des estimateurs de la méthode des moments. On observe tout d'abord que les deux estimations sont fondées sur les estimateurs des moments des probabilités. On note $x_i$ une des 500 observations tirées de la distribution normale. Ensuite, les deux proportions sont obtenues de la façon suivante : on pose $z_i(2{,}1) = \mathbf{1}[x_i < 2{,}1]$ et $z_i(3{,}6) = \mathbf{1}[x_i < 3{,}6]$ des fonctions indicatrices. Alors, la proportion 35 % est obtenue comme $\bar{z}(2{,}1)$ et 0,55 est $\bar{z}(3{,}6)$. Ainsi, les deux proportions sont simplement les moyennes des fonctions des observations de l'échantillon. Chaque $z_i$ est un tirage d'une distribution de Bernoulli avec une probabilité de succès $\pi(2{,}1) = \Phi((2{,}1 - \mu) / \sigma)$ pour $z_i(2{,}1)$, et $\pi(3{,}6) = \Phi((3\ 6 - \mu) / \sigma)$ pour $z_i(3{,}6)$. Par conséquent, $E[\bar{z}(2{,}1)] = \pi(2{,}1)$ et $E[\bar{z}(3{,}6)] = \pi(3{,}6)$. Les variances, dans chaque cas, sont $\mathrm{Var}[\bar{z}(.)] = 1 / n[\pi(.)(1 - \pi(.))]$. La covariance des deux moyennes des échantillons pose quelques difficultés, mais on peut la déduire des résultats de l'échantillon aléatoire. $\mathrm{Cov}[\bar{z}(2{,}1), \bar{z}(3{,}6)]] = 1 / n$ $\mathrm{Cov}[z_i(2{,}1), z_i(3{,}6)]$, et, comme dans un échantillonnage aléatoire, les moments de l'échantillon convergent vers leurs contreparties de la population, $\mathrm{Cov}[z_i(2{,}1), z_i(3{,}6)] = \mathrm{plim}\ [\{(1 / n) \sum_{i=1}^{n} z_i(2{,}1) z_i(3{,}6)\} - \pi(2{,}1)\pi(3{,}6)]$. Mais, $z_i(2{,}1)z_i(3{,}6)$ doit égaliser $[z_i(2{,}1)]^2$ qui, à son tour, doit être égal à $z_i(2{,}1)$. Il s'ensuit alors que $\mathrm{Cov}[z_i(2{,}1), z_i(3{,}6)] = \pi(2{,}1)[1 - \pi(3{,}6)]$. Par conséquent, la matrice de covariance asymptotique des deux proportions d'échantillons est :

$$Asy.Var[p(2,1), p(3,6)] = \Sigma \quad = \frac{1}{n}\begin{bmatrix} \pi(2,1)(1-\pi(2,1)) & \pi(2,1)(1-\pi(3,6)) \\ \pi(2,1)(1-\pi(3,6)) & \pi(3,6)(1-\pi(3,6)) \end{bmatrix}$$

Si on insère les estimations des échantillons, on obtient :

$$Est.Asy.Var[p(2,1), p(3,6)] = \mathbf{S} = \begin{bmatrix} 0,000455 & 0,000315 \\ 0,000315 & 0,000495 \end{bmatrix}$$

Enfin, les estimations de μ et de σ sont obtenues comme fonctions de $p(2,1)$ et $p(3,6)$, en utilisant la méthode des moments. Les équations du moment sont :

$$m_{2,1} = \left[\frac{1}{n}\sum_{i=1}^{n} z_i(2,1)\right] - \Phi\left[\frac{2,1-\mu}{\sigma}\right] = 0$$

$$m_{3,6} = \left[\frac{1}{n}\sum_{i=1}^{n} z_i(3,6)\right] - \Phi\left[\frac{3,6-\mu}{\sigma}\right] = 0$$

On pose maintenant $\mathbf{\Gamma} = \begin{bmatrix} \partial m_{2,1} / \partial\mu & \partial m_{2,1} / \partial\sigma \\ \partial m_{3,6} / \partial\mu & \partial m_{3,61} / \partial\sigma \end{bmatrix}$ et on note **G** l'estimation de l'échantillon de Γ. Alors, l'estimateur de la matrice de covariance asymptotique de $(\hat{\mu}, \hat{\sigma})$ est $[\mathbf{G}\mathbf{S}^{-1}\mathbf{G}']^{-1}$. Il reste à établir les dérivées, qui sont juste $\partial m_{2,1} / \partial\mu = (1 / \sigma)\phi((2,1 - \mu) / \sigma)$ et $\partial m_{2,1} / \partial\sigma = (2,1 - \mu) / \sigma[\mathrm{M}m_{2,1} / \mathrm{M}\sigma]$ et de la même façon pour $m_{3,6}$. En intégrant les estimations de l'échantillon, on obtient $\mathbf{G} = \begin{bmatrix} 0,37046 & -0,14259 \\ 0,39579 & 0,04987 \end{bmatrix}$. Enfin, en multipliant les matrices et en calculant les inverses nécessaires, on a $[\mathbf{G}\mathbf{S}^{-1}\mathbf{G}']^{-1} = \begin{bmatrix} 0,10178 & -0,12492 \\ -0,12492 & 0,16973 \end{bmatrix}$. La distribution asymptotique est normale, comme d'habitude. Fondé sur les résultats, un intervalle de confiance à 95 % pour μ est $3,2301 \pm 1,96(0,10178)^2 = 2,6048$ à 3,8554.

# Exercice 10

La variable aléatoire $x$ a la distribution suivante : $f(x) = e^{-\lambda}\lambda^x / x!$, $x = 0, 1, 2, \ldots$ L'échantillon aléatoire suivant est tiré : 1, 1, 4, 2, 0, 0, 3, 2, 3, 5, 1, 2, 1, 0, 0. Mettre en œuvre un test de Wald de l'hypothèse « λ= 2 ».

Pour un échantillon aléatoire tiré d'une distribution de Poisson, l'estimateur du maximum de vraisemblance de λ est $\bar{x} = 25 / 15$. La dérivée seconde de la log-vraisemblance est $-\sum_{i=1}^{n} x_i \ / \lambda^2$, de sorte que la variance asymptotique est $\lambda / n$. La statistique de Wald est :

$$W = \frac{\left(\bar{x} - 2\right)^2}{\hat{\lambda}/n} = [(25 / 15 - 2)^2] / [(25 / 15) / 15] = 1{,}0$$

La valeur critique à 95 % d'une distribution chi-deux à 1 degré de liberté est 3,84. L'hypothèse n'est pas rejetée. Alternativement, on pourrait estimer la variance avec $s^2 / n = 2{,}38 / 15 = 0{,}159$. Alors, la statistique de Wald est $(1{,}6 - 2)^2 / 0{,}159 = 1{,}01$. La conclusion est la même.

## Exercice 11

**Test de la normalité**. Une méthode suggérée pour tester si la distribution fondant un échantillon est normale, consiste à comparer la statistique $L = \mathrm{n}\{\text{asymétrie}^2 / 6 + (\text{aplatissement} - 3)^2 / 24\}$) à la distribution chi-deux à 2 degrés de liberté. En utilisant les données de l'exercice 1, mettre en œuvre ce test.

Le coefficient d'asymétrie est 0,14192 et l'aplatissement est 1,8447. (Il s'agit des troisième et quatrième moments divisés par la troisième et la quatrième puissances de l'écart-type de l'échantillon.) En les intégrant dans la formule donnée ci-dessus, on obtient $L = 10\{0{,}14192^2 / 6 + (1{,}8447 - 3)^2 / 24\} = 0{,}59$. La valeur critique de la distribution chi-deux à 2 degrés de liberté (95 %) est 5,99. Ainsi, l'hypothèse de normalité ne peut pas être rejetée.

## Exercice 12

On suppose que la distribution jointe de deux variables aléatoires $x$ et $y$ est :

$$f(x, y) = \theta e^{-(\beta+\theta)y} (\beta y)^x / x! \ \beta, \theta \ 0, y \ \$ \ 0, x = 0, 1, 2, ...$$

a. Trouver les estimateurs du maximum de vraisemblance de β et θ ainsi que leur distribution asymptotique jointe.

b. Trouver l'estimateur du maximum de vraisemblance de $\theta / (\beta+\theta)$ et sa distribution asymptotique.

c. Prouver que $f(x)$ est de forme $f(x) = \gamma(1 - \gamma)^x$, $x = 0, 1, 2, ...$

   Trouver alors l'estimateur du maximum de vraisemblance de γ et sa distribution asymptotique.

d. Prouver que f(y*x) est de forme $\lambda e^{-\lambda y}(\lambda y)^{x} / x!$. Prouver que $f(y|x)$ s'intègre à 1. Trouver l'estimateur du maximum de vraisemblance de $\lambda$ et sa distribution asymptotique. (Indice : dans la distribution conditionnelle, considérer les $x$ comme des constantes.)

e. Prouver que $f(y) = \theta e^{-\theta y}$. Trouver alors l'estimateur du maximum de vraisemblance de $\theta$ et sa variance asymptotique.

f. Prouver que $f(x|y) = e^{-\beta y} (\beta y)^{x} / x!$. À partir de cette distribution, quel est l'estimateur du maximum de vraisemblance de $\beta$ ?

La log-vraisemblance est :

$$\ln L = n\ln\theta - (\beta+\theta)\textstyle\sum_{i=1}^{n} y_i + \ln\beta \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} x_i \log y_i - \sum_{i=1}^{n} \log(x_i\,!)$$

Les dérivées premières et secondes sont :

$$\partial \ln L / \partial\theta = n / \theta - \textstyle\sum_{i=1}^{n} y_i$$

$$\partial \ln L / \partial\beta = -\textstyle\sum_{i=1}^{n} y_i + \sum_{i=1}^{n} x_i / \beta$$

$$\partial^2 \ln L / \partial\theta^2 = -n / \theta^2$$

$$\partial^2 \ln L / \partial\beta^2 = -\textstyle\sum_{i=1}^{n} x_i / \beta^2$$

$$\partial^2 \ln L / \partial\beta\partial\theta = 0$$

Par conséquent, les estimateurs du maximum de vraisemblance sont $\hat{\theta} = 1 / \overline{y}$ et $\hat{\beta} = \overline{x} / \overline{y}$ et la matrice de covariance asymptotique est l'inverse de $E\begin{bmatrix} n/\theta^2 & 0 \\ 0 & \sum_{i=1}^{n} x_i / \beta^2 \end{bmatrix}$. Afin de compléter la dérivation, on a besoin de la valeur espérée de $\sum_{i=1}^{n} x_i = nE[x_i]$. Pour obtenir $E[x_i]$, il est nécessaire d'avoir la distribution marginale de $x_i$, qui est $f(x) = \int_0^{\infty} \theta e^{-(\beta+\theta)y} (\beta y)^x / x!\,dy = \beta^x (\theta / x!) \int_0^{\infty} e^{-(\beta+\theta)y} y^x dy$. C'est $\beta^x(\theta / x!)$ fois une intégrale gamma : $f(x) = \beta^x(\theta / x!)[\Gamma(x+1)] / (\beta+\theta)^{x+1}$. Comme $\Gamma(x + 1) = x!$, l'expression se simplifie :

$$f(x) = [\theta / (\beta + \theta)][\beta / (\beta + \theta)]^x$$

Ainsi, $x$ a une distribution géométrique de paramètre $\pi = \theta / (\beta + \theta)$. (C'est la distribution du nombre de tentatives indépendantes jusqu'au premier succès, la probabilité de succès étant égale à $1 - \pi$.) Enfin, on a besoin de la valeur espérée de $x_i$, qui est

$E[x] = [\theta / (\beta + \theta)] \sum_{x=0}^{\infty} x[\beta / (\beta + \theta)]^x = \beta / \theta$. La matrice de covariance asymptotique demandée est alors $\begin{bmatrix} n/\theta^2 & 0 \\ 0 & n(\beta/\theta)/\beta^2 \end{bmatrix}^{-1} = \begin{bmatrix} \theta^2/n & 0 \\ 0 & \beta\theta/n \end{bmatrix}$.

L'estimateur du maximum de vraisemblance de $\theta / (\beta + \theta)$ est :

$$\widehat{\theta/(\beta+\theta)} = (1 / \bar{y}) / [\bar{x} / \bar{y} + 1 / \bar{y}] = 1 / (1 + \bar{x})$$

Sa variance asymptotique est obtenue en utilisant la variance d'une fonction non linéaire :

$$V = [\beta / (\beta + \theta)]^2(\theta^2 / n) + [-\theta / (\beta + \theta)]^2(\beta\theta / n) = \beta\theta^2 / [n(\beta + \theta)^3]$$

La variance asymptotique aurait pu être aussi obtenue comme $[-1 / (1 + E[x])^2]^2$Asy. Var$[\bar{x}]$.)

Pour la question (c), on remarque juste que $\gamma = \theta / (\beta + \theta)$. Pour un échantillon d'observations de $x$, la log-vraisemblance est :

$$\ln L = n\ln\gamma + \ln(1 - \gamma) \sum_{i=1}^{n} x_i$$

$$\partial \ln L / d\gamma = n / \gamma - \sum_{i=1}^{n} x_i \ / (1 - \gamma)$$

Une solution est obtenue en remarquant tout d'abord qu'elle est telle que : $(1 - \gamma) / \gamma = \bar{x} = 1 / \gamma - 1$. La solution pour $\gamma$ est donc $\hat{\gamma} = 1 / (1 + \bar{x})$. Bien entendu, c'est ce qui a été trouvé à la question (b).

Pour la question (d), $f(y|x) = \dfrac{f(x,y)}{f(x)} = \dfrac{\theta e^{-(\beta+\theta)y}(\beta y)^x(\beta+\theta)^x(\beta+\theta)}{x! \quad \theta \quad \beta x}$. En éliminant les termes et en regroupant ceux qui restent, on a $f(y|x) = (\beta+\theta)[(\beta+\theta)y]^x e^{-(\beta+\theta)y} / x!$ de sorte que la densité a la forme requise avec $\lambda = (\beta + \theta)$. L'intégrale est $\left\{[\lambda^{x+1}]/x!\right\}\int_0^\infty e^{-\lambda y} y^x dy$. Il s'agit d'une intégrale gamma égale à $\Gamma(x + 1) / \lambda^{x+1}$, qui est la réciproque du premier scalaire, de sorte que le produit est égal à 1. La log-vraisemblance est :

$$\ln L = n\ln\lambda - \lambda \sum_{i=1}^{n} y_i + \ln\lambda \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \ln x_i!$$

$$\partial \ln L / \partial\lambda = (\sum_{i=1}^{n} x_i + n) / \lambda - \sum_{i=1}^{n} y_i .$$

$$\partial^2 \ln L / \partial\lambda^2 = -(\sum_{i=1}^{n} x_i + n) / \lambda^2$$

Donc, l'estimateur du maximum de vraisemblance de $\lambda$ est $(1 + \bar{x}) / \bar{y}$ et la variance asymptotique, conditionnellement à $x$, est Asy. Var.$\left[\hat{\lambda}\right] = (\lambda^2 / n) / (1 + \bar{x})$.

Pour la question (e), on peut obtenir $f(y)$ en sommant sur $x$ dans la densité jointe. Tout d'abord, on écrit la densité jointe comme $f(x,y) = \theta e^{-\theta y} e^{-\beta y} (\beta y)^x / x!$. La somme est donc $f(y) = \theta e^{-\theta y} \sum_{x=0}^{\infty} e^{-\beta y} (\beta y)^x / x!$. La somme est celle des probabilités d'une distribution de Poisson, de sorte qu'elle est égale à 1. On obtient le résultat voulu. L'estimateur du maximum de vraisemblance de θ et sa variance asymptotique sont dérivés de :

$$\ln L = n \ln\theta - \theta \sum_{i=1}^{n} y_i$$

$$\partial \ln L / \partial\theta = n / \theta - \sum_{i=1}^{n} y_i$$

$$\partial^2 \ln L / \partial\theta^2 = -n / \theta^2$$

Par conséquent, l'estimateur du maximum de vraisemblance est $1 / \overline{y}$ et sa variance asymptotique est $\theta^2 / n$. Comme on a trouvé $f(y)$ en factorisant $f(x, y)$ en $f(y)f(x|y)$ (évident, étant donné nos résultats), la réponse suit immédiatement. Diviser juste l'expression utilisée dans la question (e) par $f(y)$. C'est une distribution de Poisson de paramètre $\beta y$. La fonction de log-vraisemblance et sa dérivée première sont :

$$\ln L = -\beta \sum_{i=1}^{n} y_i + \ln \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} x_i \ln y_i - \sum_{i=1}^{n} \ln x_i!$$

$$\partial \ln L / \partial\beta = -\sum_{i=1}^{n} y_i + \sum_{i=1}^{n} x_i / \beta$$

d'où :

$$\hat{\beta} = \overline{x} / \overline{y}$$

## Exercice 13

On suppose que $x$ a une distribution de Weibull, $f(x) = \alpha\beta x^{\beta-1}\exp(-\alpha x^{\beta})$, $x, \alpha, \beta > 0$.

a. Écrire la fonction de log-vraisemblance pour un échantillon aléatoire de $n$ observations.

b. Trouver les équations de vraisemblance pour l'estimation du maximum de vraisemblance de α et β. On remarque que la première donne une solution explicite pour α en fonction des données et de β. Mais, après l'avoir intégrée dans la seconde, on obtient seulement une solution implicite pour β. Comment obtenir les estimateurs du maximum de vraisemblance ?

c. Établir la matrice des dérivées secondes de la log-vraisemblance par rapport à α et β. Les espérances exactes des éléments impliquant β reposent sur les dérivées de la fonction gamma et sont complexes analytiquement. Bien entendu, votre résultat exact donne un estimateur empirique. Comment estimer la matrice de covariance asymptotique des estimateurs de la question (b) ?

d. Prouver que $\alpha\beta\mathrm{Cov}[\ln x, x^{\beta}] = 1$. (Indice : utiliser le fait que les dérivées premières espérées de la fonction de log-vraisemblance sont nulles.)

La log-vraisemblance et ses dérivées premières sont

$$\log L = n\log\alpha + n\log\beta + (\beta - 1)\sum_{i=1}^{n}\log x_i - \alpha\sum_{i=1}^{n} x_i^{\beta}$$

$$\partial\log L / \partial\alpha = n / \alpha - \sum_{i=1}^{n} x_i^{\beta}$$

$$\partial\log L / \partial\beta = n / \beta + \sum_{i=1}^{n}\log x_i - \alpha\sum_{i=1}^{n}(\log x_i)x_i^{\beta}$$

Comme la première équation de vraisemblance implique qu'au maximum, $\hat{\alpha} = n / \sum_{i=1}^{n} x_i^{\beta}$, une approche serait de balayer les valeurs possibles de β et de calculer les valeurs de α correspondantes. Deux complications pratiques viennent de l'intervalle large de β et des valeurs de départ à utiliser pour la recherche.

Les dérivées secondes sont :

$$\partial^2\ln L / \partial\alpha^2 = -n / \alpha^2$$

$$\partial^2\ln L / \partial\beta^2 = -n / \beta^2 - \alpha\sum_{i=1}^{n}(\log x_i)^2 x_i^{\beta}$$

$$\partial^2\ln L / \partial\alpha\partial\beta = -\sum_{i=1}^{n}(\log x_i)x_i^{\beta}$$

Si l'on disposait des estimations, la façon la plus simple d'estimer les valeurs espérées du hessien serait d'évaluer les expressions ci-dessus avec les estimations du maximum de vraisemblance, et de calculer ensuite l'inverse négatif. En premier lieu, comme la valeur espérée de $\partial\ln L / \partial\alpha$ est zéro, il s'ensuit que $\mathrm{E}[x_i^{\beta}] = 1 / \alpha$. Maintenant,

$$E[\partial\ln L / \partial\beta] = n / \beta + E[\sum_{i=1}^{n}\log x_i] - \alpha E[\sum_{i=1}^{n}(\log x_i)x_i^{\beta}] = 0$$

également. On divise par $n$, et on utilise le fait que chaque terme dans une somme a la même espérance pour obtenir :

$$1 / \beta + E[\ln x_i] - E[(\ln x_i)x_i^{\beta}] / E[x_i^{\beta}] = 0$$

Maintenant, on multiplie par $E[x_i^{\beta}]$ pour obtenir $E[x_i^{\beta}] = E[(\ln x_i)x_i^{\beta}] - E[\ln x_i]E[x_i^{\beta}]$

ou $1 / (\alpha\beta) = \mathrm{Cov}[\ln x_i, x_i^{\beta}]$

# Exercice 14

Les données suivantes ont été générées par une distribution de Weibull à l'exercice 13 :

1,3043 0,49254 1,2742 1,4019 0,32556 0,29965 0,26423

1,0878 1,9461 0,47615 3,6454 0,15344 1,2357 0,96381

0,33453 1,1227 2,0296 1,2797 0,96080 2,0070

a. Trouver les estimations du maximum de vraisemblance de $\alpha$ et $\beta$ et estimer leur matrice de covariance asymptotique.

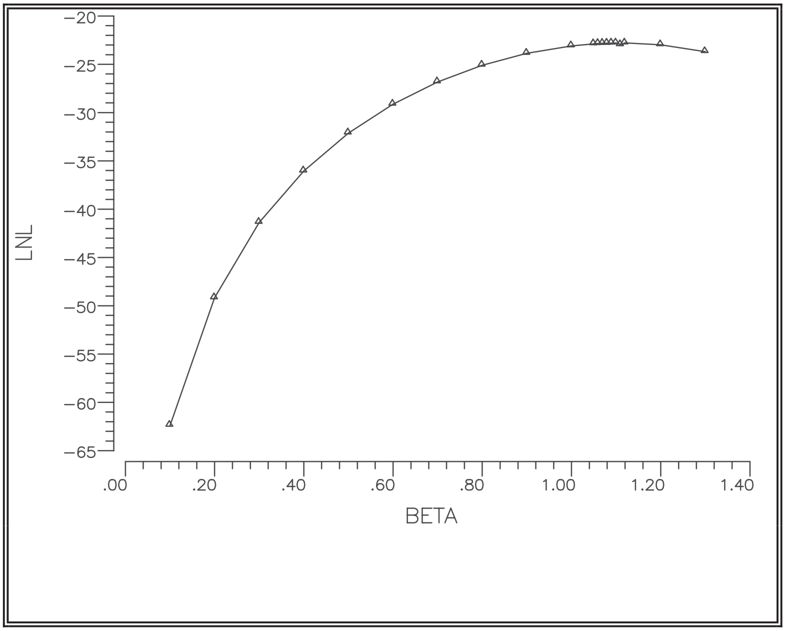
b. Mettre en œuvre un test de Wald de l'hypothèse « $\beta = 1$ ».

c. Trouver l'estimation du maximum de vraisemblance de $\alpha$ sous l'hypothèse « $\beta = 1$ ».

d. En utilisant les résultats des questions (a) et (c) mettre en œuvre un test du ratio de vraisemblance de l'hypothèse « $\beta = 1$ ».

e. Mettre en œuvre un test du multiplicateur de Lagrange de l'hypothèse « $\beta = 1$ ».

Comme il a été suggéré dans le problème précédent, on peut tout d'abord se concentrer sur $\alpha$. De $\partial \log L / \partial\alpha = 0$, on trouve qu'au maximum, $\alpha = 1 / [(1/n) \ \sum_{i=1}^{n} x_i^{\beta} \ ]$. On examine ensuite les différentes valeurs de $\beta$ pour chercher la valeur qui maximise $\log L$, où on substitue cette expression à chaque occurrence de $\alpha$. Des valeurs de $\beta$ et la log-vraisemblance correspondante sont données et représentées sur la figure ci-dessous.




| $\beta$ | $\log L$ |
|---|---|
| 0,1 | –62,386 |
| 0,2 | –49,175 |
| 0,3 | –41,381 |
| 0,4 | –36,051 |
| 0,5 | –32,122 |
| 0,6 | –29,127 |
| 0,7 | –26,829 |
| 0,8 | –25,098 |
| 0,9 | –23,101 |
| 1,05 | –22,891 |
| 1,06 | –22,863 |
| 1,07 | –22,841 |
| 1,08 | –22,823 |
| 1,09 | –22,809 |
| 1,10 | –22,800 |
| 1,11 | –22,796 |
| 1,12 | –22,797 |
| 1,2 | –22,984 |
| 1,3 | –23,693 |

Le maximum se produit en $\beta = 1{,}11$. La valeur de $\alpha$ correspondante est 1,179. L'opposé de la matrice des dérivées secondes, à ces valeurs, et son inverse sont $\mathbf{I}\left(\hat{\alpha},\hat{\beta}\right) = \begin{bmatrix} 25.55 & 9.6506 \\ 9.6506 & 27.7552 \end{bmatrix}$ et $\mathbf{I}^{-1}\left(\hat{\alpha},\hat{\beta}\right) = \begin{bmatrix} .04506 & -.2673 \\ -.2673 & .04148 \end{bmatrix}$.

La statistique de Wald de l'hypothèse « $\beta = 1$ » est $W = (1{,}11 - 1)^2 / 0{,}041477 = 0{,}276$. La valeur critique pour le test de taille 0,05 est 3,84, de sorte que l'on ne rejette pas l'hypothèse.

Si $\beta = 1$, alors $\hat{\alpha} = n / \sum_{i=1}^{n} x_i = 0{,}88496$. La distribution est géométrique si $\beta = 1$, de sorte que la log-vraisemblance contrainte est $\log L_r = n\log\alpha - \alpha \sum_{i=1}^{n} x_i = n(\log\alpha - 1)$ à l'EMV. $\log L_r$ en $\alpha = 0{,}88496$ est $-22{,}44435$.

La statistique du ratio de vraisemblance est :

$$-2\log\lambda = 2(23{,}10068 - 22{,}44435) = 1{,}3126.$$

De nouveau, c'est une valeur faible. Pour obtenir la statistique du multiplicateur de Lagrange, on calcule :

$$\begin{bmatrix} \partial \log L / \partial\alpha & \partial \log L / \partial\beta \end{bmatrix} \begin{bmatrix} -\partial^2 \log L / \partial\alpha^2 & -\partial^2 \log L / \partial\alpha\partial\beta \\ -\partial^2 \log L / \partial\alpha\partial\beta & -\partial^2 \log L / \partial\beta^2 \end{bmatrix}^{-1} \begin{bmatrix} \partial \log L / \partial\alpha \\ \partial \log L / \partial\beta \end{bmatrix}$$

aux estimations contraintes de $\alpha = 0{,}88496$ et $\beta = 1$. En faisant les substitutions adéquates dans l'expression ci-dessus, on a à ces valeurs :

$$\partial \log L / \partial\alpha = 0$$

$$\partial \log L / \partial\beta = n + \sum_{i=1}^{n} \log x_i - \frac{1}{\bar{x}} \sum_{i=1}^{n} x_i \log x_i = 9{,}400342$$

$$\partial^2 \log L / \partial\alpha^2 = -n\bar{x}^2 = -25{,}54955$$

$$\partial^2 \log L / \partial\beta^2 = -n - \frac{1}{\bar{x}} \sum_{i=1}^{n} x_i (\log x_i)^2 = -30{,}79486$$

$$\partial^2 \log L / \partial\alpha\partial\beta = -\sum_{i=1}^{n} x_i \log x_i = -8{,}265$$

L'élément en bas à droite dans la matrice inverse est 0,041477. La statistique LM est donc $(9{,}40032)^2 0{,}041477 = 2{,}9095$. C'est bien inférieur à la valeur critique de la distribution du chi-deux, de sorte que l'hypothèse n'est pas rejetée sur la base de ces trois tests.

## Exercice 15

On cherche à établir un intervalle de confiance pour la variance d'une distribution normale. L'intervalle est formé en trouvant $c_{bas}$ et $c_{haut}$ tels que :

$$\text{Prob}[c_{bas} < \chi^2[n-1] < c_{haut}] = 1 - \alpha$$

Les points limites de l'intervalle de confiance sont alors $(n-1)s^2 / c_{haut}$ et $(n-1)s^2 / c_{bas}$. Comment trouver l'intervalle le plus étroit ? On cherche simplement à minimiser la largeur de l'intervalle, $c_{haut} - c_{bas}$ sous contrainte que la probabilité contenue dans l'intervalle est $(1-\alpha)$. Prouver que pour la distribution symétrique comme pour la distribution asymétrique, l'intervalle le plus étroit est tel que la densité est la même aux deux limites.

Le problème général est de minimiser $Haut - Bas$ sous contrainte $F(Haut) - F(Bas) = 1 - \alpha$, où $F(.)$ est la distribution chi-deux appropriée. On peut traiter ce problème sous la forme d'un lagrangien :

$$\text{Min}_{H,B}\, L_* = H - B + \lambda\{(F(H) - F(B)) - (1-\alpha)\}$$

Les conditions nécessaires sont :

$$\partial L_* / \partial H = 1 + \lambda f(H) = 0$$

$$\partial L_* / \partial B = -1 - \lambda f(B) = 0$$

$$\partial L_* / \partial \lambda = (F(H) - F(B)) - (1-\alpha) = 0$$

Des deux premières équations, il est évident, qu'au minimum, $f(H)$ doit être égale à $f(B)$.

# Annexe E

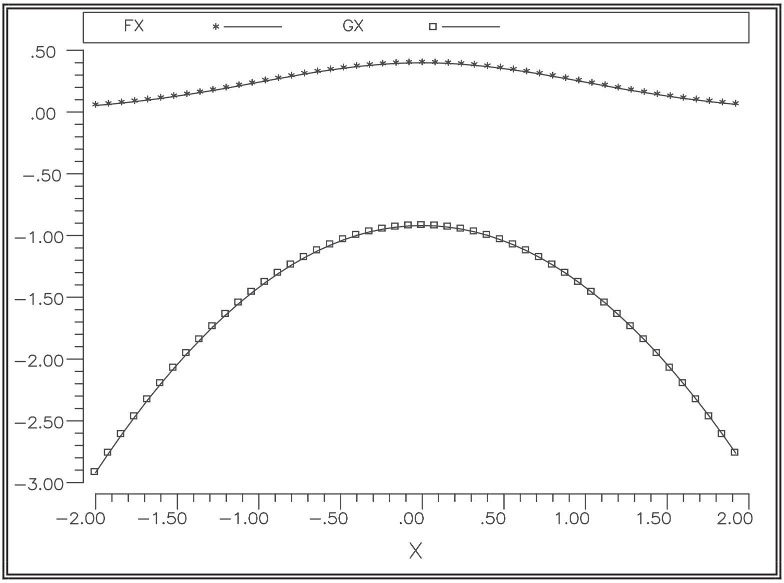
## Exercice 1

Montrer comment maximiser la fonction $f(\beta) = \frac{1}{\sqrt{2\pi}} e^{-(\beta-c)^2/2}$ par rapport à β pour une constante, $c$, en utilisant la méthode de Newton. Montrer que maximiser $\log f(\beta)$ conduit à la même solution. Représenter graphiquement $f(\beta)$ et $\log f(\beta)$.

La condition nécessaire pour maximiser $f(\beta)$ est :

$$df(\beta) / d\beta = \frac{1}{\sqrt{2\pi}} e^{-(\beta-c)^2/2} [-(\beta - c)] = 0 = -(\beta - c)\mathrm{f}(\beta)$$

La fonction exponentielle n'est jamais nulle, la seule solution ayant la condition nécessaire est alors $\beta = c$. La dérivée seconde est $\mathrm{d}^2 f(\beta) / \mathrm{d}\beta^2 = -(\beta - c)\mathrm{d}f(\beta) / \mathrm{d}\beta - \mathrm{f}(\beta) = [(\beta - c)^2 - 1]f(\beta)$. À la valeur stationnaire $b = c$, la dérivée seconde est négative, de sorte que c'est un maximum. On considère à la place la fonction $g(\beta) = \log f(\beta) = -(1 / 2)\ln(2\pi) - (1 / 2)(\beta - c)^2$. Le premier terme constant n'est, bien entendu, pas pertinent pour la solution, et la partie quadratique est toujours négative sauf au point $\beta = c$. Donc, il est évident que cette fonction a la même valeur maximisatrice que $f(\beta)$. Formellement, $dg(\beta) / d\beta = -(\beta - c) = 0$ en $\beta = c$, et $d^2 g(\beta) / \mathrm{d}\beta^2 = -1$, de sorte que c'est en fait le maximum. Un graphique des deux fonctions est donné ci-après.



On remarque que la fonction transformée est toujours concave, alors que la fonction originale présente deux points d'inflexion, et qu'elle est d'abord convexe, puis concave, puis convexe.

## Exercice 2

Prouver que la méthode de Newton pour minimiser la somme des carrés des résidus dans le modèle de régression linéaire converge vers le minimum en une itération.

La fonction à maximiser est $f(\beta) = (\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta)$. Les dérivées requises sont $\partial f(\beta) / \partial\beta = -\mathbf{X}'(\mathbf{y} - \mathbf{X}\beta)$ et $\partial^2 f(\beta) / \partial\beta\partial\beta\partial = \mathbf{X}'\mathbf{X}$. On commence maintenant l'itération de Newton en un point arbitraire, $\beta^0$. L'itération est définie en (12-17) :

$$\beta^1 = \beta^0 - (\mathbf{X}'\mathbf{X})^{-1}\{-\mathbf{X}'(\mathbf{y} - \mathbf{X}\beta^0)\} = \beta^0 + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\beta^0 = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{b}$$

Donc, quelle que soit la valeur de départ choisie, la valeur suivante est le vecteur de coefficients des moindres carrés.

## Exercice 3

Pour le modèle de régression de Poisson, $\text{Prob}[Y_i = y_i|\mathbf{x}_i] = \dfrac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}$ avec $\lambda_i = e^{\boldsymbol{\beta}'\mathbf{x}_i}$. La fonction de log-vraisemblance est $\ln L = \sum_{i=1}^{n} \log\text{Prob}[Y_i = y_i|\mathbf{x}_i]$.

a. Insérer l'expression de $\lambda_i$ pour obtenir la fonction de log-vraisemblance en fonction des données observées.

b. Dériver les conditions du premier ordre pour maximiser cette fonction par rapport à $\beta$.

c. Dériver la matrice des dérivées secondes de cette fonction-critère par rapport à $\beta$. Est-ce que cette matrice est définie négative ?

d. Définir les calculs pour utiliser la méthode de Newton afin d'obtenir les estimations des paramètres inconnus.

e. Écrire l'ensemble complet des étapes dans un algorithme afin d'obtenir les estimations des paramètres de ce modèle. Inclure dans votre algorithme un test de convergence des estimations fondé sur le critère suggéré de Belsley.

f. Comment obtenez-vous les valeurs de départ pour les itérations ?

g. Les données suivantes sont générées par un modèle de régression de Poisson avec $\log\lambda = \alpha + \beta x$.

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *y* | 6 | 7 | 4 | 10 | 10 | 6 | 4 | 7 | 2 | 3 | 6 | 5 | 3 | 3 | 4 |
| *x* | 1,5 | 1,8 | 1,8 | 2,0 | 1,3 | 1,6 | 1,2 | 1,9 | 1,8 | 1,0 | 1,4 | 0,5 | 0,8 | 1,1 | 0,7 |

Utiliser les résultats des questions (a) à (f) pour calculer les estimations du maximum de vraisemblance de $\alpha$ et $\beta$. Obtenir aussi les estimations de la matrice de covariance asymptotique de vos estimations.

La log-vraisemblance est :

$$\log L = \sum_{i=1}^{n} [-\lambda_i + y_i \ln\lambda_i - \ln y_i!] = -\sum_{i=1}^{n} e^{\boldsymbol{\beta}'\mathbf{x}_i} + \sum_{i=1}^{n} y_i\left(\boldsymbol{\beta}'\mathbf{x}_i\right) - \sum_{i=1}^{n} \log y_i!$$

$$= -\sum_{i=1}^{n} e^{\boldsymbol{\beta}'\mathbf{x}_i} + \boldsymbol{\beta}'\sum_{i=1}^{n} \mathbf{x}_i y_i - \sum_{i=1}^{n} \log y_i!$$

La condition nécessaire est :

$$\partial \ln L / \partial\boldsymbol{\beta} = -\sum_{i=1}^{n} \mathbf{x}_i e^{\boldsymbol{\beta}'\mathbf{x}_i} + \sum_{i=1}^{n} \mathbf{x}_i y_i = \mathbf{0} \text{ ou } \mathbf{X}'\mathbf{y} = \sum_{i=1}^{n} \mathbf{x}_i \lambda_i$$

Il est utile de remarquer que, comme $E[y_i{*}\mathbf{x}_i] = \lambda_i = e^{\boldsymbol{\beta}'\mathbf{x}_i}$, la condition du premier ordre est équivalente à $\sum_{i=1}^{n} \mathbf{x}_i y_i = \sum_{i=1}^{n} \mathbf{x}_i E[y_i{*}\mathbf{x}_i]$ ou $\mathbf{X}'\mathbf{y} = \mathbf{X}'E[\mathbf{y}]$, ce qui est logique. Il est possible d'écrire la condition du premier ordre comme $\partial \ln L / \partial\boldsymbol{\beta} = \sum_{i=1}^{n} \mathbf{x}_i(y_i - \lambda_i) = \mathbf{0}$ qui est similaire à la contrepartie de la régression classique, si l'on écrit $(y_i - \lambda_i) = (y_i - E[y_i{*}\mathbf{x}_i])$ comme un résidu. La matrice des dérivées secondes est $\partial \ln L / \partial\boldsymbol{\beta}\partial\boldsymbol{\beta}' = -\sum_{i=1}^{n}\left(e^{\boldsymbol{\beta}'\mathbf{x}_i}\right)\mathbf{x}_i\mathbf{x}_i' = -\sum_{i=1}^{n} \lambda_i \mathbf{x}_i\mathbf{x}_i'$. C'est une matrice définie négative. Pour le prouver, on remarque tout d'abord que $\lambda_i$ doit toujours être positif. Alors, on note $\boldsymbol{\Omega}$ une matrice diagonale dont le *i*-ième élément de la diagonale est $\sqrt{\lambda_i}$, et on note $\mathbf{Z} = \boldsymbol{\Omega}\mathbf{X}$. Alors, $\partial \ln L / \partial\boldsymbol{\beta}\partial\boldsymbol{\beta}' = -\mathbf{Z}'\mathbf{Z}$ qui est clairement définie négative. Cela implique que la fonction de log-vraisembance est globalement concave et qu' l'on trouve son maximum, de façon simple et fiable, en utilisant la méthode de Newton.

L'itération de la méthode de Newton est définie dans (5-17). Il est possible de l'appliquer directement dans ce problème. Les calculs à mener afin de mettre en œuvre la méthode de Newton pour maximiser ln*L* se déroulent de la façon suivante :

1. Obtenir les valeurs de départ pour les paramètres. Comme la fonction de log-vraisemblance est globalement concave, les valeurs utilisées importent généralement peu. La plupart des applications utilisent simplement zéro. Une suggestion qui apparaît dans la littérature est $\boldsymbol{\beta}^0 = \left[\sum_{i=1}^{n} q_i \mathbf{x}_i\mathbf{x}_i'\right]^{-1}\left[\sum_{i=1}^{n} q_i \mathbf{x}_i y_i\right]$ avec $q_i = \log(\max(1, y_i))$.
2. L'itération est calculée comme $\hat{\boldsymbol{\beta}}_{t+1} = \hat{\boldsymbol{\beta}}_t + \left[\sum_{i=1}^{n} \hat{\lambda}_i \mathbf{x}_i\mathbf{x}_i'\right]^{-1}\left[\sum_{i=1}^{n} \mathbf{x}_i(y_i - \hat{\lambda}_i)\right]$.

3. À chaque fois qu'on calcule $\hat{\beta}_{t+1}$, on doit vérifier la convergence. Il y a plusieurs façons de le faire :

   a. Gradient : est-ce que les éléments de $\partial \ln L / \partial \beta$ sont petits ?

   b. Changement : est-ce que $\hat{\beta}_{t+1} - \hat{\beta}_t$ est petit ?

   c. Fonction taux de change : vérifier la taille de :

$$\delta_t = \left[\sum_{i=1}^n \mathbf{x}_i(y_i - \hat{\lambda}_i)\right]' \left[\sum_{i=1}^n \hat{\lambda}_i \mathbf{x}_i \mathbf{x}_i'\right]^{-1} \left[\sum_{i=1}^n \mathbf{x}_i(y_i - \hat{\lambda}_i)\right]$$

   avant de calculer $\hat{\beta}_{t+1}$. Cette mesure décrit ce qui arrive à la fonction à la valeur suivante de $\beta$. C'est le critère de Belsley.

4. Quand la convergence a été atteinte, la matrice de covariance asymptotique des estimations est estimée avec la matrice inverse utilisée dans les itérations.

   En utilisant les données du problème, les résultats des calculs décrits sont :

| Iter. | $\alpha$ | $\beta$ | $\ln L$ | $\partial \ln L/\partial\alpha$ | $\partial \ln L/\partial\beta$ | Change |
|---|---|---|---|---|---|---|
| 0 | 0 | 0 | -102,387 | 65 | 95,1 | 296,261 |
| 1 | 1,37105 | 2,17816 | -1 442,38 | -1 636,25 | -2 788,5 | 1 526,36 |
| 2 | 0,619874 | 2,05865 | -461,989 | -581,966 | -996,711 | 516,92 |
| 3 | 0,210347 | 1,77914 | -141,022 | -195,953 | -399,751 | 197,652 |
| 4 | 0,351893 | 1,26291 | -51,2989 | -57,9294 | -102,847 | 30,616 |
| 5 | 0,824956 | 0,698768 | -33,5530 | -12,8702 | -23,1932 | 2,75855 |
| 6 | 1,05288 | 0,453352 | -32,0824 | -1,28785 | -2,29289 | 0,032399 |
| 7 | 1,07777 | 0,425239 | -32,0660 | -0,016067 | -0,028454 | 0,0000051 |
| 8 | 1,07808 | 0,424890 | -32,0660 | 0 | 0 | 0 |

Aux valeurs finales, l'inverse opposé de la matrice des dérivées secondes est :

$$\left[\sum_{i=1}^n \hat{\lambda}_i \mathbf{x}_i \mathbf{x}_i'\right]^{-1} = \begin{bmatrix} 0{,}151044 & -0{,}095961 \\ -0{,}095961 & 0{,}0664665 \end{bmatrix}$$

## Exercice 4

Utiliser l'intégration de Monte-Carlo pour représenter la fonction $g(r) = E[x^r * x > 0]$ pour la distribution normale standard.

La valeur espérée de la distribution normale tronquée est :

$$E[x^r \mid x > 0] = \int_0^\infty x^r f(x \mid x > 0)dx = \frac{\int_0^\infty x^r \phi(x)dx}{\int_0^\infty \phi(x)dx} = \frac{2}{\sqrt{\pi}} \int_0^\infty x^r e^{-\frac{x^2}{2}} dx$$

Pour évaluer cette espérance, on tire tout d'abord un échantillon de 1 000 observations de la distribution normale standard tronquée en utilisant (5-1). Pour celle-ci, $\mu = 0$, $\sigma = 1$, $P_L = \Phi((0 - 0) / 1) = 2$, et $P_U = \Phi((+4 - 0) / 1) = 1$. Donc, les tirages sont obtenus en transformant les tirages de U(0, 1) (notée $F_i$) en $x_i = \Phi[2(1 + F_i)]$. Comme $0 < F_i < 1$, l'argument entre crochets doit être plus grand que 2, de sorte que $x_i > 0$, ce qui était attendu. En utilisant les mêmes 1 000 tirages à chaque fois (de façon à lisser la courbe sur la figure), on représente ensuite les valeurs de $\bar{x}_r = \frac{1}{1\,000} \sum_{i=1}^{1\,000} x_i^r$, r = 0, 0,2, .,4, 0,6, ..., 5,0. Comme expérience supplémentaire, on génère un second échantillon de 1 000 observations en les tirant d'une distribution normale standard et en écartant, et en les retirant les valeurs qui ne sont pas positives. Les moyennes et écarts-types des deux échantillons sont (0,8097, 0,6170) pour le premier et (0,8059, 0,6170) pour le second. Le tirage du second échantillon prend approximativement deux fois plus de temps. Pourquoi ?