

Une introduction aux statistiques inférentielles

Solutions aux exercices

Christophe Lalanne

- It is well known that for two normal random variables, zero covariance implies independence. Why does this not apply to the following situation : $X \sim \mathcal{N}(0, 1)$, $\text{Cov}(X, X^2) = \mathbb{E}X^3 - \mathbb{E}X\mathbb{E}X^2 = 0 - 0 = 0$ but obviously X^2 is totally independent on X ?

Il est assez facile de montrer que l'indépendance de deux v.a. entraîne une covariance nulle :

$$\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) \stackrel{\text{indep.}}{=} \mathbb{E}(X)\mathbb{E}(Y) - \mathbb{E}(X)\mathbb{E}(Y) = 0.$$

La réciproque n'est vraie que si X et Y sont distribuées normalement de manière conjointe, ce que l'on peut vérifier en calculant la loi jointe et le produit des distributions marginales.

De l'expression ci-dessus on vérifie que, pour une v.a. X gaussienne, $\text{Cov}(X, X^2) = 0$. Dans cet exemple, une covariance nulle n'implique pas l'indépendance puisque la v.a. X^2 n'est pas gaussienne.

- Trouver les valeurs $\hat{\alpha}$ et $\hat{\beta}$ qui minimisent la somme des carrés

$$\sum_{i=1}^n (y_i - \alpha - \beta x_i)^2. \tag{1}$$

Les valeurs $\hat{\alpha}$ et $\hat{\beta}$ sont des estimés de l'ordonnée à l'origine et de la pente, respectivement, d'une droite de régression ajustée aux données $\{(x_i, y_i)\}_{i=1}^n$ par la méthode des moindres carrés (ordinaires). Plus formellement, les estimateurs peuvent s'exprimer comme

$$(\hat{\alpha}, \hat{\beta}) = \arg \min_{(\alpha, \beta)} \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2.$$

Il faut bien comprendre que $\hat{\alpha}$ et $\hat{\beta}$ sont des v.a. puisqu'elles s'expriment comme des fonctions des v.a. (x_i, y_i) . Les v.a. $\hat{\alpha}$ et $\hat{\beta}$ sont appelés les estimateurs des vrais paramètres (inconnus mais fixes) α et β .

Ces estimateurs peuvent être obtenus en différenciant la somme de carrés [1](#) par rapport à α et β et en cherchant à annuler la dérivée. On a ainsi

$$\begin{aligned} \frac{\partial \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2}{\partial \alpha} &= -2 \sum_{i=1}^n (y_i - \alpha - \beta x_i) = 0, \\ \alpha &= n^{-1} \sum_{i=1}^n y_i - n^{-1} \beta \sum_{i=1}^n x_i, \end{aligned}$$

et

$$\frac{\partial \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2}{\partial \beta} = -2 \sum_{i=1}^n (y_i - \alpha - \beta x_i) x_i = 0.$$

En introduisant la valeur de α calculée précédemment, on a alors

$$0 = \sum_{i=1}^n y_i x_i - n^{-1} \sum_{i=1}^n y_i \sum_{i=1}^n x_i + n^{-1} \beta \left(\sum_{i=1}^n x_i \right)^2 - \beta \sum_{i=1}^n x_i^2.$$

La résolution en β est immédiate :

$$\begin{aligned} \beta &= \frac{n^{-1} \sum_{i=1}^n y_i \sum_{i=1}^n x_i - \sum_{i=1}^n y_i x_i}{n^{-1} (\sum_{i=1}^n x_i)^2 - \sum_{i=1}^n x_i^2} \\ &= \frac{\sum_{i=1}^n y_i x_i - n^{-1} \sum_{i=1}^n y_i \sum_{i=1}^n x_i}{\sum_{i=1}^n x_i^2 - n^{-1} (\sum_{i=1}^n x_i)^2} \\ &= \frac{\sum_{i=1}^n y_i x_i - n \bar{y} \bar{x}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2} \\ &= \frac{s_{XY}}{s_{XX}}. \end{aligned}$$

En conclusion, la somme des carrés est minimisée pour $\alpha = \hat{\alpha} = \bar{y} - \hat{\beta} \bar{x}$ et $\beta = \hat{\beta} = \frac{s_{XY}}{s_{XX}}$.

- Soit $\mathcal{X}_* = \mathcal{H} \mathcal{X} \mathcal{D}^{-1/2}$, avec \mathcal{X} une matrice $(n \times p)$, \mathcal{H} une matrice de centrage et $\mathcal{D}^{-1/2} = \text{diag}(s_{11}^{-1/2}, \dots, s_{pp}^{-1/2})$. Montrer que \mathcal{X}_* est une matrice standardisée, où $\bar{x}_* = 0_p$ et $S_{\mathcal{X}_*} = \mathcal{R}_{\mathcal{X}}$, la matrice de corrélation de \mathcal{X} .

Le vecteur des moyennes, \bar{x}_* , peut s'exprimer comme

$$\begin{aligned} \bar{x}_* &= \mathbf{1}_n^T \mathcal{X}_* / n \\ &= \mathbf{1}_n^T \mathcal{H} \mathcal{X} \mathcal{D}^{-1/2} / n \\ &= \mathbf{1}_n^T (\mathbb{I}_n - n^{-1} \mathbf{1}_n \mathbf{1}_n^T) \mathcal{X} \mathcal{D}^{-1/2} / n \\ &= (\mathbf{1}_n^T - \mathbf{1}_n^T n^{-1} \mathbf{1}_n \mathbf{1}_n^T) \mathcal{X} \mathcal{D}^{-1/2} / n \\ &= (\mathbf{1}_n^T - \mathbf{1}_n^T) \mathcal{X} \mathcal{D}^{-1/2} / n \\ &= 0_p. \end{aligned}$$

De même, pour la matrice de covariance, $S_{\mathcal{X}}$, de \mathcal{X}_* , on a :

$$\begin{aligned}
\mathcal{S}_{\mathcal{X}} &= \mathbb{V}(\mathcal{H}\mathcal{X}\mathcal{D}^{-1/2}) \\
&= \mathbb{V}(\mathbb{I}_n\mathcal{X}\mathcal{D}^{-1/2}) + \mathbb{V}(n^{-1}\mathbf{1}_n\mathbf{1}_n^T\mathcal{X}\mathcal{D}^{-1/2}) \\
&= \mathcal{D}^{-1/2}(\mathbb{V}(\mathcal{X})\mathcal{D}^{-1/2}) \\
&= \mathcal{D}^{-1/2}\mathcal{S}_{\mathcal{X}}\mathcal{D}^{-1/2} \\
&= \mathcal{R}_{\mathcal{X}}.
\end{aligned}$$

Les formules ci-dessus montrent que la multiplication à gauche par la matrice de centrage \mathcal{H} permet de soustraire les moyennes des colonnes tandis que la multiplication à droite par la matrice $\mathcal{D}^{-1/2}$ permet de diviser chaque colonne par l'écart-type estimé.

- Un modèle linéaire peut s'exprimer sous la forme

$$Y = \mathcal{X}\beta + \varepsilon,$$

où \mathcal{X} est de plein rang et ε symbolise les erreurs aléatoires. Montrer que la solution des moindres carrés

$$\hat{\beta} = \arg \min_{\beta} (Y - \mathcal{X}\beta)^T (Y - \mathcal{X}\beta) = \arg \min_{\beta} \varepsilon^T \varepsilon,$$

peut s'exprimer sous la forme $\hat{\beta} = (\mathcal{X}^T \mathcal{X})^{-1} \mathcal{X}^T Y$. (voir aussi exercice 8)

On définit la fonction $f(\beta) = (Y - \mathcal{X}\beta)^T (Y - \mathcal{X}\beta)$, i.e.

$$f(\beta) = Y^T Y - 2\beta^T \mathcal{X}^T Y + \beta^T \mathcal{X}^T \mathcal{X} \beta.$$

Le minimum de $f(\beta)$ peut être obtenu en cherchant le zéro de la dérivée

$$\frac{\partial f(\beta)}{\partial \beta} = \frac{\partial Y^T Y - 2\beta^T \mathcal{X}^T Y + \beta^T \mathcal{X}^T \mathcal{X} \beta}{\partial \beta} = -2\mathcal{X}^T Y + 2\mathcal{X}^T \mathcal{X} \beta = 0.$$

Il en découle que la solution, $\hat{\beta}$, doit satisfaire $\hat{\beta} = (\mathcal{X}^T \mathcal{X})^{-1} \mathcal{X}^T Y$.

Vérifions que nous avons bien trouvé un minimum en calculant la dérivée seconde de $f(\beta)$ au point $\hat{\beta}$:

$$\frac{\partial^2 f(\beta)}{\partial \beta \partial \beta^T} = \frac{\partial (-2\mathcal{X}^T Y + 2\mathcal{X}^T \mathcal{X} \beta)}{\partial \beta} = 2\mathcal{X}^T \mathcal{X}.$$

La matrice \mathcal{X} est de plein rang, donc la matrice $\mathcal{X}^T \mathcal{X}$ est définie positive et par conséquent, $\hat{\beta}$ est bien le minimum de la fonction du carré des résidus $f(\beta)$.

- Supposons un vecteur aléatoire Y de distribution $Y \sim \mathcal{N}_p(0, \mathcal{I})$. Le transformer pour créer le vecteur $X \sim \mathcal{N}(\mu, \Sigma)$ avec $\mu = (3, 2)^T$ et $\Sigma = \begin{pmatrix} 1 & -1.5 \\ -1.5 & 4 \end{pmatrix}$. Comment peut-on implémenter la formule résultante sur un ordinateur ?

Considérons la transformation suivante :

$$X = \mu + \Sigma^{1/2}Y.$$

On sait qu'un vecteur aléatoire gaussien soumis à une transformation linéaire est encore gaussien. D'après les règles de manipulation des matrices des espérances et des variances pour les v.a. transformées, on sait que $\mathbb{E}X = \mu + \Sigma^{1/2}\mathbb{E}Y = \mu$ et $\mathbb{V}X = \Sigma^{1/2}\mathbb{V}Y(\Sigma^{1/2})^T = \Sigma$.

Sur un ordinateur, la matrice $\Sigma^{1/2}$ peut être calculée à partir de Σ en utilisant une décomposition spectrale :

$$\Sigma^{1/2} = \begin{pmatrix} -0.38 & 0.92 \\ 0.92 & 0.38 \end{pmatrix} \begin{pmatrix} 4.62 & 0 \\ 0 & 0.38 \end{pmatrix}^{1/2} = \begin{pmatrix} 0.84 & -0.54 \\ -0.54 & 1.95 \end{pmatrix}.$$

Il suffit ensuite d'appliquer la formule précédente qui transforme Y en X .

- Montrer que si $X \sim \mathcal{N}_p(0, \Sigma)$, alors la variable $U = (X - \mu)^T \Sigma^{-1} (X - \mu)$ suit une loi χ_p^2 .

Si l'on a un vecteur aléatoire $X \sim \mathcal{N}_p(0, \Sigma)$ tel que $\Sigma > 0$, le vecteur à p dimensions

$$(Y_1, \dots, Y_p)^T = Y = \Sigma^{-1/2}(X - \mu)$$

suit une loi normale multivariée de vecteur de moyennes $\mathbb{E}Y = 0_p$ et de matrice de covariance $\mathbb{V}Y = \mathcal{I}_p$.

La transformation linéaire $\Sigma^{-1/2}(X - \mu)$ est appelée transformation de Mahalanobis.

Par conséquent, la v.a.

$$U = (X - \mu)^T \Sigma^{-1} (X - \mu) = Y^T Y = \sum_{i=1}^p Y_i^2$$

est une somme de carrés de v.a. gaussiennes centrées et indépendantes ; elle suit donc une loi du chi-deux à p degrés de liberté.

- Supposons que X soit de moyenne nulle et de covariance $\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$. Soit $Y = X_1 + X_2$. Écrire Y comme une transformation linéaire, c'est-à-dire trouver la matrice de transformation \mathcal{A} . Calculer ensuite $\mathbb{V}(Y)$.

On a clairement

$$Y = X_1 + X_2 = \mathcal{A}X = \begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

et $\mathcal{A}\mathbb{X} = \mathbb{E}\{(\mathcal{A}X - \mathbb{E}\mathcal{A}X)(\mathcal{A}X - \mathbb{E}\mathcal{A}X)^T\} = \mathcal{A}\{\mathbb{E}(X - \mathbb{E}X)(X - \mathbb{E}X)^T\}\mathcal{A}^T = \mathcal{A}\mathbb{V}X\mathcal{A}^T$.

Par conséquent,

$$\mathbb{V}X = \mathcal{A}\Sigma\mathcal{A}^T = \begin{pmatrix} 1 & 1 \end{pmatrix} \Sigma \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 3.$$

Une autre possibilité consiste à écrire

$$\mathbb{V}Y = \mathbb{V}(X_1 + X_2) = \mathbb{V}X_1 + 2 \text{Cov}(X_1, X_2) + \mathbb{V}X_2 = 3.$$

- Calculer la moyenne et la variance de l'estimateur $\hat{\beta} = (\mathcal{X}^T \mathcal{X})^{-1} \mathcal{X}^T Y$ dans le modèle linéaire $Y = \mathcal{X}\beta + \varepsilon$, où $\mathbb{E}(\varepsilon) = 0_n$ et $\mathbb{V}(\varepsilon) = \sigma^2 \mathcal{I}_n$.

L'estimateur $\hat{\beta} = (\mathcal{X}^T \mathcal{X})^{-1} \mathcal{X}^T Y$ du paramètre inconnu β a été dérivé dans l'exercice 4. Il s'ensuit que

$$\mathbb{E}\hat{\beta} = (\mathcal{X}^T \mathcal{X})^{-1} \mathcal{X}^T \mathbb{E}Y = (\mathcal{X}^T \mathcal{X})^{-1} \mathcal{X}^T (\mathcal{X}\beta + \mathbb{E}\varepsilon) = \beta$$

puisque par hypothèse $\mathbb{E}\varepsilon = 0_n$.

En ce qui concerne la variance, on a

$$\begin{aligned} \mathbb{V}\hat{\beta} &= \mathbb{V}\{(\mathcal{X}^T \mathcal{X})^{-1} \mathcal{X}^T Y\} \\ &= (\mathcal{X}^T \mathcal{X})^{-1} \mathcal{X}^T \mathbb{V}Y \mathcal{X} (\mathcal{X}^T \mathcal{X})^{-1} \\ &= (\mathcal{X}^T \mathcal{X})^{-1} \mathcal{X}^T \sigma^2 \mathcal{I}_n \mathcal{X} (\mathcal{X}^T \mathcal{X})^{-1} \\ &= \sigma^2 (\mathcal{X}^T \mathcal{X})^{-1}, \end{aligned}$$

en utilisant le fait que $\mathbb{V}Y = \mathbb{V}\varepsilon = \sigma^2 \mathcal{I}_n$.

- Calculer les moments conditionnels $\mathbb{E}(X_2 \mid x_1)$ et $\mathbb{E}(X_1 \mid x_2)$ pour la fonction de densité bi-dimensionnelle suivante :

$$f(x_1, x_2) = \begin{cases} \frac{1}{2}x_1 + \frac{3}{2}x_2 & 0 \leq x_1, x_2 \leq 1 \\ 0 & \text{sinon} \end{cases}$$

Les densités marginales de X_1 et X_2 , pour $0 \leq x_1, x_2 \leq 1$, sont :

$$f_{X_1}(x_1) = \int_0^1 f(x_1, x_2) dx_2 = \left[\frac{1}{2}x_1 x_2 + \frac{3}{4}x_2^2 \right]_0^1 = \frac{1}{2}x_1 + \frac{3}{4}$$

et

$$f_{X_2}(x_2) = \int_0^1 f(x_1, x_2) dx_1 = \left[\frac{1}{4}x_1^2 + \frac{3}{2}x_1 x_2 \right]_0^1 = \frac{1}{4} + \frac{3}{2}x_2.$$

Les espérances conditionnelles, toujours pour $0 \leq x_1, x_2 \leq 1$, peuvent être calculées comme suit :

$$\begin{aligned}
\mathbb{E}(X_2 | X_1 = x_1) &= \int_0^1 x_2 f(x_2 | x_1) dx_2 \\
&= \int_0^1 x_2 \frac{f(x_1, x_2)}{f_{X_1}(x_1)} dx_2 \\
&= \int_0^1 x_2 \left(\frac{\frac{1}{2}x_1 + \frac{3}{2}x_2}{\frac{1}{2}x_1 + \frac{3}{4}} \right) dx_2 \\
&= \left[\frac{\frac{x_1 x_2^2}{4} + \frac{x_2^3}{2}}{\frac{3}{4} + \frac{x_1}{2}} \right]_0^1 \\
&= \frac{x_1 + 2}{3 + 2x_1}
\end{aligned}$$

et

$$\begin{aligned}
\mathbb{E}(X_1 | X_2 = x_2) &= \int_0^1 x_1 f(x_1 | x_2) dx_1 \\
&= \int_0^1 x_1 \frac{f(x_1, x_2)}{f_{X_2}(x_2)} dx_1 \\
&= \int_0^1 x_1 \left(\frac{\frac{1}{2}x_1 + \frac{3}{2}x_2}{\frac{3}{2}x_2 + \frac{1}{4}} \right) dx_1 \\
&= \left[\frac{\frac{x_1^3}{6} + \frac{3x_1^2 x_2}{4}}{\frac{1}{4} + \frac{3x_2}{2}} \right]_0^1 \\
&= \frac{2 + 9x_2}{3 + 18x_2}.
\end{aligned}$$

- Montrer que $\mathbb{E}(X_2) = \mathbb{E}\{\mathbb{E}(X_2 | X_1)\}$, où $\mathbb{E}(X_2 | X_1)$ désigne l'espérance conditionnelle de X_2 connaissant X_1 .

Comme $\mathbb{E}(X_2 | X_1 = x_1)$ est une fonction de x_1 , il est clair que $\mathbb{E}(X_2 | X_1)$ est un vecteur aléatoire (fonction du vecteur X_1 , lui-même aléatoire).

Si l'on considère que le vecteur $X = (X_1, X_2)^T$ a pour densité $f(x_1, x_2)$, alors

$$\begin{aligned}
\mathbb{E}\{\mathbb{E}(X_2 | X_1)\} &= \int \left\{ \int x_2 f(x_2 | x_1) dx_2 \right\} f(x_1) dx_1 \\
&= \int \left\{ \int x_2 \frac{f(x_2, x_1)}{f(x_1)} dx_2 \right\} f(x_1) dx_1 = \int \int x_2 f(x_2, x_1) dx_2 dx_1 \\
&= \mathbb{E}X_2.
\end{aligned}$$

- Trouver la fonction de densité de probabilité associée au vecteur aléatoire $Y = \mathcal{A}X$ où $\mathcal{A} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$, sachant que X possède la fonction de densité définie à l'exercice 9.

La fonction de densité de Y est donnée par

$$f_Y(y) = \text{abs}(|\mathcal{J}|)f_X\{u(y)\},$$

où $u(\cdot)$ est la transformée inverse, i.e. $X = u(Y)$, et \mathcal{J} est le jacobien de $u(\cdot)$. Dans ce cas, $X = u(Y) = \mathcal{A}^{-1}Y = \mathcal{J}Y$.

Il suffit de résoudre $y_1 = x_1 + x_2$ et $y_2 = x_1 - x_2$ pour x_1 et x_2 :

$$\begin{aligned} x_1 &= u_1(y_1, y_2) = (y_1 + y_2)/2 \\ x_2 &= u_2(y_1, y_2) = (y_1 - y_2)/2 \end{aligned}$$

et on en déduit que le jacobien de $u(\cdot)$ est

$$\mathcal{J} = \begin{pmatrix} \frac{\partial u_1(y)}{\partial y_1} & \frac{\partial u_1(y)}{\partial y_2} \\ \frac{\partial u_2(y)}{\partial y_1} & \frac{\partial u_2(y)}{\partial y_2} \end{pmatrix} = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{pmatrix}.$$

Ensuite, on a $|\mathcal{J}| = -\frac{1}{2}$ et $\text{abs}(|\mathcal{J}|) = \frac{1}{2}$, et donc on obtient la densité du vecteur transformé Y :

$$\begin{aligned} f_Y(y) &= \frac{1}{2}f_X\{u(y)\} = \frac{1}{2}f_X\left\{\begin{pmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}\right\} \\ &= \frac{1}{2}f_X\left\{\frac{1}{2}(y_1 + y_2), \frac{1}{2}(y_1 - y_2)\right\} \end{aligned}$$

pour $0 \leq u_1(y_1, y_2), u_2(y_1, y_2) \leq 1$ et $f_Y(y) = 0$ sinon.

En substituant cela dans la fonction de densité de X , on obtient

$$f_Y(y) = \begin{cases} \frac{1}{2} \left[\frac{1}{2} \left\{ \frac{1}{2}(y_1 + y_2) \right\} + \frac{3}{2} \left\{ \frac{1}{2}(y_1 - y_2) \right\} \right] & 0 \leq y_1 + y_2 \leq 2, \\ 0 & \text{sinon} \end{cases}$$

et, en utilisant de l'algèbre simple pour déterminer la région sur laquelle la densité de $f_Y(y)$ est supérieure à 0, on a finalement :

$$f_Y(y) = \begin{cases} \frac{1}{2}y_1 - \frac{1}{4}y_2 & 0 \leq y_1 \leq 2, \quad |y_2| \leq 1 - |1 - y_1| \\ 0 & \text{sinon} \end{cases}$$

- Montrer que la fonction

$$f_Y(y) = \begin{cases} \frac{1}{2}y_1 + \frac{1}{4}y_2 & 0 \leq y_1 \leq 2, \quad |y_2| \leq 1 - |1 - y_1| \\ 0 & \text{sinon} \end{cases}$$

est bien une densité de probabilité.

La surface sur laquelle la fonction ci-dessus n'est pas nulle est représentée dans la **figure 1**.

Afin de montrer que $f_Y(y)$ est bien une densité de probabilité (bi-dimensionnelle), il faut vérifier qu'elle n'est pas négative et que son intégrale vaut 1.

On voit aisément que la fonction $f_Y(y)$ n'est jamais négative à l'intérieur du carré représenté dans la **figure 1** puisque $y_1 \geq 0$ et $y_1 \geq y_2$ impliquent que $y_1/2 - y_2/4 > 0$.

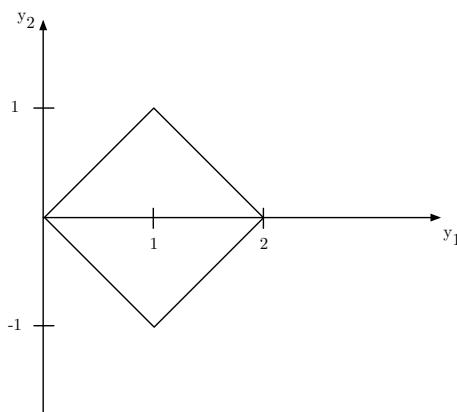


Figure 1 Domaine de définition de la densité $f_Y(y_1, y_2)$.

Il reste à vérifier que $\int f_Y(y) dy = 1$, ce que l'on effectue ci-dessous :

$$\begin{aligned}
 \iint f_Y(y_1, y_2) dy_2 dy_1 &= \int_0^1 \int_{-y_1}^{y_1} f_Y(y) dy_2 dy_1 + \int_1^2 \int_{y_1-2}^{2-y_1} f_Y(y) dy_2 dy_1 \\
 &= \int_0^1 \int_{-y_1}^{y_1} \left(\frac{1}{2} y_1 - \frac{1}{4} y_2 \right) dy_2 dy_1 + \int_1^2 \int_{y_1-2}^{2-y_1} \left(\frac{1}{2} y_1 - \frac{1}{4} y_2 \right) dy_2 dy_1 \\
 &= \int_0^1 \left[\frac{1}{2} y_1 y_2 - \frac{1}{8} y_2^2 \right]_{-y_1}^{y_1} dy_1 + \int_1^2 \left[\frac{1}{2} y_1 y_2 - \frac{1}{8} y_2^2 \right]_{y_1-2}^{2-y_1} dy_1 \\
 &= \int_0^1 y_1^2 dy_1 + \int_1^2 (-y_1^2 + 2y_1) dy_1 \\
 &= \left[\frac{1}{3} y_1^3 \right]_0^1 + \left[-\frac{1}{3} y_1^3 + y_1^2 \right]_1^2 = \frac{1}{3} + \frac{2}{3} = 1.
 \end{aligned}$$

- Déterminer la distribution du vecteur aléatoire $Y = \mathcal{A}X$ avec $\mathcal{A} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$, où $X = (X_1, X_2)^T$ possède une distribution bi-normale.

Le vecteur aléatoire Y suit une loi bi-normale puisqu'il est défini à partir d'une transformation linéaire de vecteurs gaussiens.

La distribution normale est entièrement déterminée par sa moyenne et sa matrice de covariance pour lesquelles on a

$$\mathbb{E}Y = \mathbb{E}\mathcal{A}X = \mathcal{A}\mathbb{E}X = \mathcal{A}0_2 = 0_2$$

et

$$\mathbb{V}Y = \mathbb{V}\mathcal{A}X = \mathcal{A}\mathbb{V}X\mathcal{A}^T = \mathcal{A}\mathcal{I}_2\mathcal{A}^T = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}.$$

Donc, Y_1 et Y_2 ne sont pas corrélés et, pour une des v.a. gaussiennes conjointement, une covariance nulle implique l'indépendance.

La densité du vecteur aléatoire X ,

$$f_X(x_1, x_2) = \frac{1}{2\pi} \exp \left\{ \frac{1}{2} (x_1 \ x_2) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \right\}$$

est évidemment constante sur le cercle centré en $(0, 0)^T$ puisque ces valeurs ne changent que lorsque la forme quadratique $(x_1 \ x_2) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = x_1^2 + x_2^2$ change. On remarquera qu'un cercle de diamètre r est défini comme l'ensemble des points $x = (x_1, x_2)^T \in \mathbb{R}^2$ satisfaisant $x_1^2 + x_2^2 = r^2$.

La densité du vecteur transformé Y est également constante sur le cercle, mais la distribution est plus dispersée. La transformation $Y = \mathcal{A}X$ correspond à une rotation suivie d'une multiplication par un facteur $\sqrt{2}$.