

# SY19 Automne 2009

## TP 0

### 1 Loi normale bidimensionnelle

#### 1.1 Géométrie de la loi normale bidimensionnelle

On rappelle qu'une variable aléatoire  $\mathbf{X} \in \mathbb{R}^p$ , suivant une loi normale multidimensionnelle de vecteur moyenne  $\mu$  et de matrice de variance-covariance  $\Sigma$ , a pour fonction de densité :

$$f_{\mathbf{X}}(\mathbf{x}; \mu, \Sigma) = (2\pi)^{-p/2} |\Sigma|^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu)^\top \Sigma^{-1}(\mathbf{x} - \mu)\right). \quad (1)$$

Les courbes d'iso-densité de cette variable aléatoire sont des ellipsoïdes de centre  $\mu$ , dont la forme est définie par la matrice  $\Sigma$ .

La diagonalisation de la matrice de variance-covariance  $\Sigma$  peut s'écrire :

$$\Sigma = \lambda D^\top A D, \quad (2)$$

où  $\lambda$  est un paramètre scalaire,  $D$  est une matrice de changement de base, et  $A$  est une matrice diagonale telle que  $|A| = 1$ . En particulier, dans le cas bidimensionnel ( $p = 2$ ), on a :

$$D = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}, \quad A = \begin{pmatrix} a & 0 \\ 0 & 1/a \end{pmatrix}.$$

Le scalaire  $\theta$  mesure alors l'angle entre l'axe des abscisses et l'axe de l'ellipse orienté selon les premier et troisième quadrants. Le scalaire  $a$  caractérise le facteur d'aplatissement de l'ellipse ; le grand axe est orienté selon les I<sup>er</sup> et III<sup>e</sup> quadrants lorsque  $a > 1$ , et selon les quadrants II et IV lorsque  $a < 1$ . La figure 1 résume la signification de ces différents paramètres.

#### 1.2 Génération d'un échantillon de réalisations d'une variable multinormale

On souhaite générer un échantillon de réalisations d'une v.a.  $\mathbf{X} \in \mathbb{R}^p$  suivant une loi multinormale d'espérance  $\mu$  et de matrice de variance-covariance  $\Sigma$  fixées. Soit  $\mathbf{Z} \in \mathbb{R}^p$  une variable aléatoire formée par concaténation de  $p$  variables aléatoires indépendantes suivant chacune une loi normale  $\mathcal{N}(0, 1)$ .

1. Quelle est la loi de  $\mathbf{Z}$  ?
2. Soient un scalaire  $\lambda$ , une matrice orthogonale  $D$ , et une matrice diagonale  $A$  telle que  $|A| = 1$ , vérifiant l'équation (2). Étant donné un vecteur  $\mu$ , comment peut-on obtenir la variable  $\mathbf{X}$  à partir de  $\mathbf{Z}$  ? Justifier.
3. Que deviennent les calculs si les variables sont représentées sous forme de vecteurs lignes ? (Question subsidiaire)

Générer et représenter un échantillon de  $n = 1000$  réalisations d'une variable normale d'espérance  $\mu = (1, 2)$  et de matrice de variance-covariance

$$\Sigma = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}.$$

À partir de cet échantillon, estimer les paramètres  $\mu$  et  $\Sigma$  de la loi de  $\mathbf{X}$  ; comparer aux vraies valeurs des paramètres et commenter.

## 2 Mélanges de lois normales

### 2.1 Génération d'une classe suivant un mélange de lois normales

On souhaite à présent générer un échantillon de  $n_1 = 1000$  réalisations d'une variable multidimensionnelle  $\mathbf{Y}$ , dont la fonction de densité est définie à partir de *plusieurs* lois normales :

$$f_{\mathbf{Y}}(\mathbf{y}) = \pi_1 f_{\mathbf{X}}(\mathbf{y}; \mu_1, \Sigma_1) + \pi_2 f_{\mathbf{X}}(\mathbf{y}; \mu_2, \Sigma_2) + \pi_3 f_{\mathbf{X}}(\mathbf{y}; \mu_3, \Sigma_3),$$

où  $\pi_1 = 1/2$ ,  $\pi_2 = 1/4$  et  $\pi_3 = 1/4$ . Nous verrons par la suite en cours que la loi de la variable aléatoire  $\mathbf{Y}$  correspond à un mélange gaussien comportant trois composantes de fréquences  $\pi_1$ ,  $\pi_2$  et  $\pi_3$ . Les valeurs des paramètres caractérisant chacune de ces composantes sont consignées dans le tableau 1.

TABLE 1 – Caractéristiques des composantes de la fonction de densité  $f_{\mathbf{Y}}$ .

| composante                       | centre $\mu$    | angle $\theta$ | $\lambda$ | a   |
|----------------------------------|-----------------|----------------|-----------|-----|
| composante 1 $(\mu_1, \Sigma_1)$ | $(-3, 8)^\top$  | $-\pi/3$       | 2         | 1.5 |
| composante 2 $(\mu_2, \Sigma_2)$ | $(-5, 10)^\top$ | $-\pi/6$       | 1         | 1.5 |
| composante 3 $(\mu_3, \Sigma_3)$ | $(-1, 10)^\top$ | $\pi/6$        | 1         | 1.5 |

Générer un échantillon de  $n = 1000$  réalisations de  $\mathbf{Y}$ , en prenant soin de respecter les fréquences des différentes composantes. Afficher les réalisations de  $\mathbf{Y}$ , de même que les courbes de niveau de sa fonction de densité. En supposant que  $\mathbf{Y}$  suit une loi normale  $\mathcal{N}(\mu, \Sigma)$ , estimer les paramètres de cette loi.

### 2.2 Variation des proportions du mélange

On souhaite faire varier les proportions du modèle décrit ci-dessus. Générer un échantillon de  $n = 1000$  réalisations de  $\mathbf{Y}$ , en prenant les valeurs suivantes :

- $(\pi_1 = 0.6, \pi_2 = 0.2, \pi_3 = 0.2)$ ,
- $(\pi_1 = 0.8, \pi_2 = 0.1, \pi_3 = 0.1)$ ,
- $(\pi_1 = 0.98, \pi_2 = 0.01, \pi_3 = 0.01)$ .

En faisant toujours l'hypothèse de normalité de la variable aléatoire  $\mathbf{Y}$ , estimer les paramètres  $\mu$  et  $\Sigma$  pour chacun de ces échantillons. Comparer les quatre estimations obtenues (aux paragraphes 2.1 et 2.2) aux vraies valeurs des paramètres et commenter.

## 3 Informations d'ordre technique

### 3.1 Fonctions utiles

- `mvrnorm` (bibliothèque `MASS`) : générer un échantillon de réalisations d'une variable multidimensionnelle.
- `mvdnorm` (à récupérer sur [le site de SY19](#)) : calculer la densité d'une loi normale multidimensionnelle en un certain nombre de points.
- `mroot` (bibliothèque `mgcv`) : calcul de la racine carrée d'une matrice
- `ginv` (bibliothèque `MASS`) : calcul de l'inverse d'une matrice (la fonction est en fait plus générale)
- `help` : premiers secours

### 3.2 Affichage des valeurs d'une fonction $f : \mathbb{R}^2 \rightarrow \mathbb{R}$

Pour afficher la valeur prise par une fonction dans un domaine restreint de l'espace, on peut procéder en se fixant une grille de points de coordonnées  $(x_1, x_2)$  et en calculant la valeur de la fonction  $f(x_1, x_2)$  en chacun de ces points.

```
# formation de la grille de points
Xaff1 <- seq(from=-8,to=6,by=0.5)
naff1 <- length(Xaff1)
Xaff2 <- seq(from=-2,to=14,by=0.5)
naff2 <- length(Xaff2)
Xaff <- cbind(rep.int(Xaff1,times=rep(naff2,naff1)),rep(Xaff2,naff1))

# calcul des valeurs de la fonction
valeurs <- ma_fonction(Xaff, parametres)
```

Il est alors possible de représenter la fonction par ses courbes de niveau.

```
plot(X)
contour(Xaff1, Xaff2, matrix(valeurs,nrow=naff1,byrow=T),
add=T, drawlabels=FALSE)
```

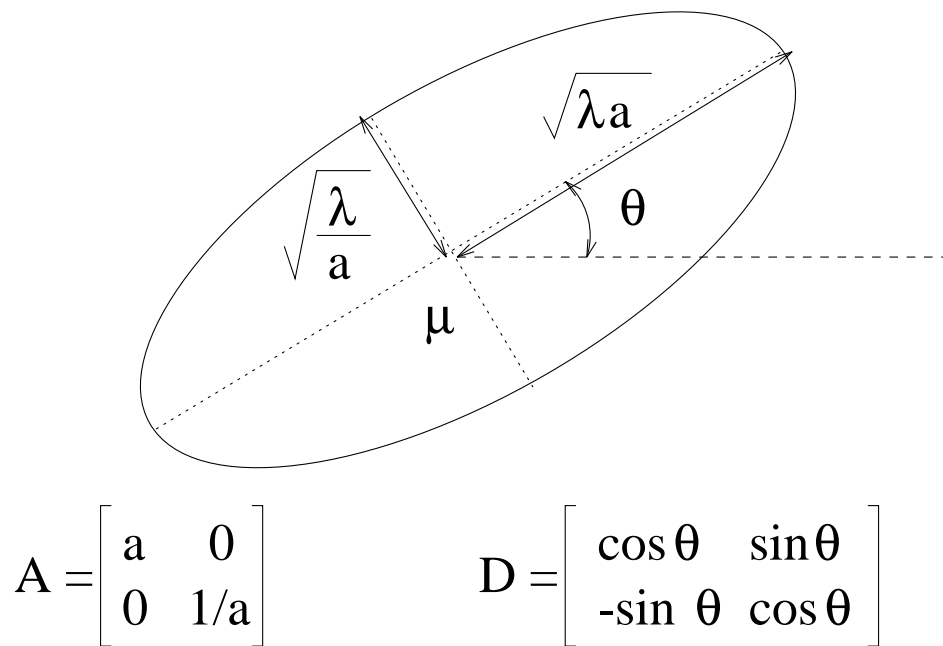


FIGURE 1 – Représentation géométrique d'une ellipse d'équidensité de la loi normale de moyenne  $\mu$  et de matrice de variance-covariance  $\Sigma = \lambda D^\top A D$ .