Vecteurs Aléatoires, Conditionnement, Simulation
TP noté

B. Delyon, V. Monbet

Master 1 - 2013-2014

A rendre, individuellement, au plus tard le 17 avril 2014 par mail

- un rapport au format pdf (nommé votrenom.pdf) dans lequel pour répondez aux questions en expliquant clairement votre démarche,
- les figures nécessaires aux réponses seront incluses dans le rapport
- les codes utilisés dans un seul fichier (nommé votrenom.R). La qualité des codes sera évaluée.

1 Questions de cours

Soit $Z \sim \mathcal{N}(0, \mathbb{I}_d)$ un vecteur aléatoire d-dimensionnel.

1. Donner la loi de $X = \mu + SZ$ avec $\mu \in \mathbb{R}^d$ et $S \in \mathbb{R}^{d\times d}$ une matrice définie positive.
2. Quelle est la loi de la première composante de $X = (X_1, \cdots, X_d)^T$?
3. Quelle est la loi des k premières composantes de $X = (X_1, \cdots, X_d)^T$?
4. Supposons que la moyenne μ est nulle. Quelle est la loi de $\sum_{i,j=1}^d K_{ij} X_i X_j$ avec $K = (SS^T)^{-1}$?
5. Supposons maintenant que $d = 2$. Quelle est la loi conditionnelle de X_1 sachant $X_2 = x_2$?
6. On considère un vecteur gaussien de dimension 3 de moyenne μ et de covariance Σ avec
$$\mu = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 1 & .7 & -.3 \\ .7 & 2 & 0 \\ -.3 & 0 & 3 \end{pmatrix}$$
7. Donner la loi de (X_2, X_3).
8. Expliciter les espérances conditionnelles $E(X_2|X_3)$, $E(X_1|X_3)$.

2 Prévision, krigeage

On dispose de données de température moyenne journalière en plusieurs villes de France au mois d'août 2003. On suppose que le vecteur aléatoire correspondant suit une loi de Gauss multivariée de moyenne $\mu_X \in \mathbb{R}^d$ telle que $\mu_X(i) = \mu$ quelque soit $i \in \{1, \cdots, d\}$ et de covariance Σ.

```
load("http://perso.univ-rennes1.fr/valerie.monbet/PrbM1/data_temperature.Rdata")
load("http://perso.univ-rennes1.fr/valerie.monbet/PrbM1/frontieres_france.Rdata")
load("http://perso.univ-rennes1.fr/valerie.monbet/PrbM1/grille_france.Rdata")
```

2.1 Modèle simplifié pour la covariance

1. Dans un premier temps, on choisit la forme paramétrique suivante pour la covariance.

$$C(x, y) = C_0 \exp\left(-\frac{d(x, y)}{\lambda}\right)$$

où $d(x, y)$ est la distance entre les points x et y, et λ est un paramètre à estimer. Une estimation de la variance donne $C_0 = 16.2$. Pour calculer la distance entre deux points on devra prendre en considération que les points sont définis par leur position géographique qui tient compte de la rotondité de la terre. On utilise la fonction `m_lldist_p` (fournie sur la page web). Avec quelle valeur λ_0 du paramètre réalise-t-on le maximum de vraisemblance?

2. Estimer, par bootstrap paramétrique, les distributions des l'estimateurs de μ et λ. Tracer un histogramme. Pour répondre à cette question vous écrirez deux fonctions. La première fera n tirages d'un vecteur Gaussien sachant sa dimension, sa moyenne et sa variance `mvrgauss <- function(n,mu,Sigma)`. La seconde `bootstrap` qui

   - prend en entrée les coordonnées des points x, y un vecteur θ des paramètres à estimer (ici $\theta = (\mu, \lambda)$), le nom de la fonction de covariance, le nombre N de tirages à réaliser ,
   - renvoie un échantillon de longueur N pour chaque paramètre,
   - fait N tirages aléatoires de taille d selon le modèle pour la valeur de θ donnée entrée, l'estimation de θ pour chaque échantillon de taille d

   La fonction `bootstrap` suivra le schéma suivant.

```
bootstrap = function(x,y,theta,covariance=cov_th_e,N){
...
# Calcul de la covariance
...
C[i,j] = covariance(d,theta[2:length(theta)]
...
# tirage aléatoire
X = mvrgauss(N,mu,Sigma)
# Estimation des paramètres
theta.boot = matrix(NA,length(theta),N)
```

```
for (k in 1:N) {
mu =
theta.boot[1,k] = mu
...
  par = optim(...)
  theta.boot[2:length(theta),k] = par$par # à adapter
}
return(theta.boot)
}
```

3. Calculer l'espérance et la variance conditionnelle de la température aux points (2.35,48.85), (-1.0,45) et (6.00,43.20), sachant les températures observées.

4. Tracer la carte de prédiction de la température pour la grille du fichier `grille_france.Rdata`. Tracer une carte de l'écart-type de l'erreur de prédiction. Commenter les résultats obtenus.

5. En construisant une méthode de validation croisée, donner une estimation de l'erreur de prédiction pour les points observés. Décrire la démarche. Analyser les résultats.

Covariance "complète"

Dans la partie précédente, on a considéré un modèle simplifié pour la covariance. On travaille maintenant avec un modèle plus classique.

$$C(x,y) = C_1 \mathbb{I}_{x=y} + C_2 \exp\left(-\frac{d(x,y)}{\lambda}\right) \mathbb{I}_{x \neq y}$$

avec C_1, C_2 et λ les paramètres inconnus.

1. On peut estimer C_1 par une méthode de moment. Estimer les paramètres C_2 et λ par maximum de vraisemblance, l'estimation de C_1 étant donnée par la méthode de moments.

2. Proposer une méthode de bootstrap paramétrique, pour estimer la distribution des estimateurs des paramètres. Tracer un histogramme. Pour répondre à cette question, vous utiliserez la fonction `bootstrap` écrite plus haut avec les bons appels.

3. Calculer l'espérance conditionnelle de la température aux points (2.35,48.85), (-1.0,45) et (6.00,43.20), sachant les températures observées. Comparer les résultats à ceux de la partie précédente.

4. En construisant une méthode de validation croisée, donner une estimation de l'erreur de prédiction pour les points observés.
   Comparer les résultats à ceux de la partie précédente.