Université de Montréal

Approche bayésienne de la construction d'intervalles de crédibilité simultanés à partir de courbes simulées

par

Marc-Élie Lapointe

Département de mathématiques et de statistique

Faculté des arts et des sciences

Mémoire présenté à la Faculté des études supérieures

en vue de l'obtention du grade de

Maître ès sciences (M.Sc.)
en Statistique

juillet 2015

Université de Montréal

Faculté des études supérieures

Ce mémoire intitulé

Approche bayésienne de la construction d'intervalles de crédibilité simultanés à partir de courbes simulées

présenté par

Marc-Élie Lapointe

a été évalué par un jury composé des personnes suivantes :

Alejandro Murua

(président-rapporteur)

Jean-François Angers

(directeur de recherche)

Christian Léger

(membre du jury)

Mémoire accepté le:

8 juillet 2015

SOMMAIRE

Ce mémoire porte sur la simulation d'intervalles de crédibilité simultanés dans un contexte bayésien. Dans un premier temps, nous nous intéresserons à des données de précipitations et des fonctions basées sur ces données : la fonction de répartition empirique et la période de retour, une fonction non linéaire de la fonction de répartition. Nous exposerons différentes méthodes déjà connues pour obtenir des intervalles de confiance simultanés sur ces fonctions à l'aide d'une base polynomiale et nous présenterons une méthode de simulation d'intervalles de crédibilité simultanés. Nous nous placerons ensuite dans un contexte bayésien en explorant différents modèles de densité *a priori*. Pour le modèle le plus complexe, nous aurons besoin d'utiliser la simulation Monte-Carlo pour obtenir les intervalles de crédibilité simultanés *a posteriori*. Finalement, nous utiliserons une base non linéaire faisant appel à la transformation angulaire et aux splines monotones pour obtenir un intervalle de crédibilité simultané valide pour la période de retour.

Mots-clés : intervalle de confiance simultané, intervalle de crédibilité simultané, période de retour, statistique bayésienne, Monte-Carlo, splines monotones.

SUMMARY

This master's thesis addresses the problem of the simulation of simultaneous credible intervals in a Bayesian context. First, we will study precipation data and two functions based on these data : the empirical distribution function and the return period, a non-linear function of the empirical distribution. We will review different methods already known to obtain simultaneous confidence intervals of these functions with a polynomial basis and we will present a method to simulate simultaneous credible intervals. Second, we will explore some models of prior distributions and in the more complex one, we will need the Monte-Carlo method to simulate simultaneous posterior credible intervals. Finally, we will use a non-linear basis based on the angular transformation and on monotone splines to obtain valid simultaneous credible intervals for the return period.

Keywords : simultaneous confidence interval, simultaneous credible interval, return period, Bayesian statistics, Monte-Carlo, monotone splines.

TABLE DES MATIÈRES

LISTE DES FIGURES

LISTE DES TABLEAUX

REMERCIEMENTS

Je tiens d'abord à exprimer ma gratitude envers mon directeur de recherche, M. Jean-François Angers, pour son appui et sa grande disponibilité qui m'ont guidé dans mes premières expériences de recherche.

Aussi, je veux remercier David Labonté et Dimitri Zuchowski, mes enseignants de mathématiques au cégep, qui ont su me transmettre leur passion des mathématiques.

Je profite aussi de cette section pour souligner le support et l'influence positive de ma famille et de mes amis tout au long de mes études, particulièrement mon épouse Karine Coulombe.

Je remercie aussi le CRNSG et le FRQNT qui m'ont soutenu financièrement tout au long de ma maîtrise.

Finalement, je tiens à remercier les membres du jury, M. Christian Léger et M. Alejandro Murua, pour leurs commentaires qui ont grandement amélioré la qualité de ce mémoire.

INTRODUCTION

Plusieurs décisions sont prises en se basant sur des résultats scientifiques. Comme ces résultats sont souvent eux-mêmes basés sur un échantillon d'observations fini, les résultats ne sont que des estimations des vrais paramètres. En fait, nous pouvons voir tout résultat provenant d'un calcul à partir d'un échantillon comme une statistique. La statistique est donc un estimateur de la vraie valeur du paramètre qui intéresse le scientifique. Il est alors nécessaire de prendre en compte la variabilité de la statistique en étudiant l'intervalle de confiance de celle-ci pour tirer des conclusions valides sur les résultats obtenus. Pour ce faire, il faut voir la statistique comme une variable aléatoire et supposer une distribution sur cette variable.

Lorsqu'un résultat scientifique est une courbe, une fonction d'un ou plusieurs paramètres et non seulement une valeur à une dimension, il faut construire un intervalle de confiance sur l'ensemble de la courbe. Nous parlerons alors d'intervalle de confiance simultané. Il faut donc d'abord trouver une distribution de cette courbe, mais il n'est pas aussi simple d'obtenir un intervalle de confiance sur une courbe que sur une valeur simple puisque dans ce cas nous avons affaire à une valeur de dimension infinie. Des solutions, pour résoudre ce problème et obtenir des intervalles de confiance théoriques simultanés, existent sous l'hypothèse de normalité et nous les exposerons.

Lorsque l'on veut ajouter de l'information *a priori* sur les données, l'approche bayésienne est très utile. Au lieu de supposer que les paramètres de la

distribution de notre statistique sont uniquement déterminés par notre échantillon, nous intégrons de l'information que l'on connait déjà sur notre paramètre. Par exemple, dans le cas d'une courbe, nous pouvons savoir avant de regarder notre échantillon qu'elle devrait être croissante, bornée entre certaines valeurs, etc. Ces informations *a priori* peuvent être intégrées directement à travers le modèle que l'on utilise ou en supposant que les paramètres de notre statistique suivent eux-mêmes une certaine distribution. Cela a plusieurs avantages dont nous discuterons plus en détail dans le corps de l'ouvrage.

Pour les modèles les plus simples, les intervalles de confiance bayésiens, que l'on appellera intervalles de crédibilité, s'obtiennent facilement puisqu'il est possible d'obtenir les distributions *a posteriori* (distributions incluant l'information *a priori* et l'information provenant de notre échantillon) et donc d'utiliser directement les solutions mentionnées précédemment. Par contre, pour les modèles les plus complexes, il n'existe pas de solution simple pour obtenir un intervalle de crédibilité sur la courbe puisque nous n'avons pas la distribution *a posteriori*. Nous allons ici faire appel à une méthode de simulation pour obtenir des intervalles de crédibilité simulés comparables aux intervalles de confiance théoriques.

Dans ce mémoire, nous étudierons un jeu de données particulier pour exposer les différentes approches. Les données étudiées seront des précipitations et nous nous intéresserons surtout aux évènements extrêmes qui sont appelés à augmenter avec les changements climatiques. Cela nous permettra d'explorer l'estimation de fonctions non linéaires comme la période de retour en passant par l'estimation de la fonction de répartition et ainsi d'étudier une base non linéaire.

Nous verrons d'abord dans le chapitre 1 les données particulières que nous utiliserons et les fonctions que nous étudierons, soit la période de retour et la fonction de répartition empirique. Par la suite, nous ferons un rappel de la

régression linéaire et nous présenterons les approches classiques pour obtenir des intervalles de confiance simultanés. Finalement, nous verrons comment estimer une courbe à l'aide de la régression linéaire avec base polynomiale ainsi que la façon d'obtenir un intervalle de confiance pour cette courbe. Nous comparerons brièvement les différentes approches classiques et verrons que la base polynomiale a des limites importantes pour l'estimation de la fonction de répartition.

Dans le chapitre 2, nous ferons une introduction à l'approche bayésienne et nous étudierons des modèles de plus en plus complexes d'estimation de fonctions qui supposent des distributions *a priori* sur différents paramètres. Les modèles étudiés seront en ordre : *a priori* de type G, matrice de covariance quelconque, inverse-gamma et inverse-gamma avec inverse-Wishart. Pour le modèle étudié le plus complexe, nous introduirons la méthode de simulation de Monte-Carlo.

Finalement, le chapitre 3 traitera de l'estimation de fonction à l'aide d'une base non linéaire. La transformation angulaire sera introduite et les splines de régression monotones seront présentés. Cela nous mènera à examiner la régression avec contraintes et les distributions *a priori* pour des paramètres avec contraintes. Nous testerons notre méthode sur des données simulées à partir d'une distribution uniforme puis nous terminerons avec une présentation des intervalles de crédibilité obtenus en combinant l'utilisation d'une base non linéaire et un modèle bayésien nécessitant la simulation Monte-Carlo.

Chapitre 1

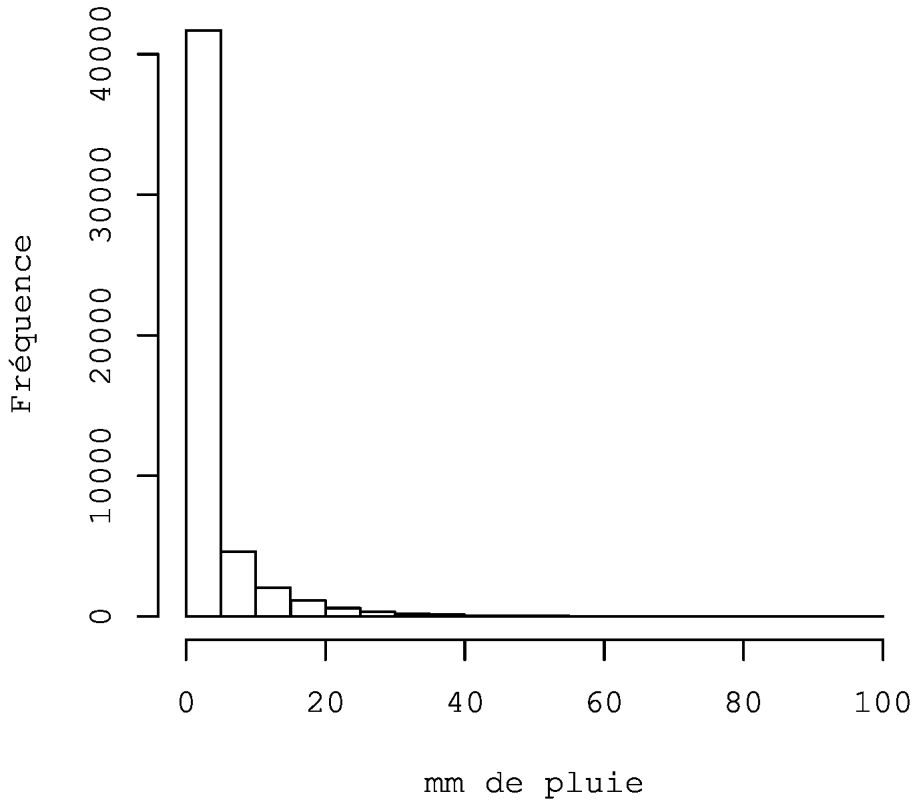
ESTIMATION DE FONCTIONS

1.1. DONNÉES

Pour ce projet, nous travaillerons plus spécifiquement avec des données de précipitations qui sont le nombre total de millimètres de précipitation reçue par jour à Dorval. Ces données sont en partie observées et en partie simulées sur 140 ans. Les 40 premières années sont observées (1960-2000) et les 100 années suivantes ont été simulées par le consortium de recherche OURANOS. Nous avons donc plus de 50000 points sur lesquels travailler. L'histogramme des données (voir la figure 1.1) révèle que plus de la moitié des journées n'ont aucune précipitation. Nous remarquons aussi que plus une journée a de précipitation, plus elle est rare. Autrement dit, la fonction qui lie la quantité de pluie reçue en une journée à la fréquence de cet évènement semble monotone ce qui est conforme avec notre intuition. Dans la prochaine section, nous formalisons cette intuition.

1.2. PÉRIODE DE RETOUR

La période de retour des évènements extrêmes est ce qui nous intéresse dans le cas présent : nous voulons savoir à quelle fréquence les précipitations très abondantes arrivent et donner un intervalle de confiance à cette période de retour. La période de retour d'un évènement Y est définie comme l'espérance



(a) Histogramme.

FIGURE 1.1. Histogramme des données de précipitations.

du temps passant entre un évènement Y et un autre évènement au moins aussi intense que Y. Pour des précipitations, nous dirons typiquement : 100 mm de pluie en une journée est un évènement qui arrive une fois aux 50 ans. Nous voulons alors dire que l'espérance du temps d'attente entre deux jours où il tombera au moins 100 mm est 50 ans.

Définissons les choses clairement. Nous posons l'évènement Y : qu'il tombe plus de a mm de pluie en un jour. Nous aurons alors que Y suit une loi de Bernouilli à paramètre p_a où p_a est la probabilité qu'il tombe plus de a mm de pluie pour un jour donné. Notre but est d'estimer l'espérance du nombre de

jours séparant deux évènements Y. En fait, nous remarquons que le nombre de jours séparant deux évènements suit une loi géométrique : à partir d'un premier évènement, nous refaisons des épreuves de Bernouilli chaque jour jusqu'à obtenir un autre évènement Y. Cela est vrai si l'on suppose que les évènements Y sont indépendants, hypothèse que nous adopterons comme dans Eagleson (1978). L'espérance est donc

$$\begin{aligned}\sum_{k=1}^{\infty} k p_a (1-p_a)^{k-1} &= p_a \frac{d}{dp_a} \sum_{k=1}^{\infty} -(1-p_a)^k \\ &= p_a \frac{d}{dp_a}\left(\frac{-1}{1-1+p_a}+1\right) = p_a \frac{d}{dp_a}\left(\frac{-1+p_a}{p_a}\right) = p_a \frac{1}{p_a^2} = \frac{1}{p_a}.\end{aligned}$$

1.3. FONCTION EMPIRIQUE

Pour un jour donné, la fonction de répartition évaluée à a nous donne la probabilité qu'il n'y ait pas d'évènement d'intensité plus grande ou égale à a : $F(a) = P(X \leq a) = 1 - p_a$, où X est la variable aléatoire représentant la quantité de pluie reçue durant un jour. Pour obtenir la période de retour, nous n'avons donc qu'à calculer $\frac{1}{1-F(a)} = 1/p_a$. Ici, nous utiliserons la fonction de répartition empirique modifiée

$$F_n(a) = \frac{(\#\{x : x \leq a\} + 1)}{(n+2)},$$

où $\#\{x : x \leq a\}$ représente le nombre de valeurs plus petites ou égales à a. Comme dans Angers et MacGibbon (2013), nous utiliserons cette version de la fonction de répartition empirique pour représenter le fait que l'évènement maximal n'est pas le dernier évènement puisque tous les évènements ne sont pas observés et parce que cette version est celle qui minimise l'erreur quadratique moyenne intégrée (IMSE). Aussi, pour obtenir une période de retour en année, nous avons 365 au dénominateur. Nous obtenons à partir de la fonction précédente la période de retour empirique

$$PdR_n(a) = \frac{1}{365(1-F_n(a))} = \frac{n+2}{365(n+1-\#\{x : x \leq a\})}.$$

Bien sûr, la fonction de répartition empirique est par définition monotone. Comme la dérivée de la période de retour par rapport à la fonction de répartition est strictement positive, celle-ci est aussi une fonction monotone. Nous avons donc par la règle de la dérivée en chaîne que la période de retour est une fonction monotone croissante de a.

Plus un évènement est rare, plus sa période de retour est grande. Regardons le lien inverse, c'est-à-dire, l'effet d'un changement de la période de retour sur la fonction de répartition. Nous avons que

$$F_n(a) = 1 - \frac{1}{365 \cdot PdR_n(a)}.$$

Nous voyons qu'un évènement deux fois plus rare qu'un autre, donc avec une période de retour deux fois plus grande, aura une valeur de fonction de répartition deux fois plus près de 1. Pour interpréter la fonction de répartition en lien avec la période de retour, il faut regarder la différence entre 1 et la valeur de la fonction de répartition. Cela sera utile quand nous comparerons ces deux fonctions.

Nous allons donc avoir en tout temps deux choix : soit estimer directement $PdR(a)$ ou estimer $F(a)$ et ensuite calculer l'estimateur de $PdR(a)$ à partir de l'estimateur de $F(a)$. La deuxième approche est considérée, car elle peut être nécessaire lorsque l'on considère des fonctions qui ne s'estiment pas directement. Par exemple, le taux de panne, $\frac{f(x)}{1-F(x)}$, est une autre fonction de $F(x)$, mais ne s'estime pas directement, car la fonction de densité empirique n'est pas un bon estimateur de $f(x)$.

De plus, comme ces fonctions sont monotones, il serait préférable d'avoir des estimateurs qui sont aussi monotones. Nous ferons pour l'instant abstraction de ce problème et nous y reviendrons au chapitre 3.

Finalement, comme mentionné au début de ce chapitre, nous avons un très grand ensemble de données. Cela est très précieux pour les scientifiques qui

veulent tirer des conclusions des données elles-mêmes. De notre côté, les données sont une excuse pour étudier la simulation d'intervalles de confiance. Donc, pour faciliter les calculs numériques nous ne travaillerons pas directement avec l'ensemble des données. Si nous gardions l'ensemble du jeu de données (plus de 50000 valeurs), nous rencontrerions des problèmes lorsque viendrait le temps de faire certains calculs car la simulation et la régression avec splines prendrait beaucoup de temps. Aussi, comme le but est d'obtenir et visualiser des intervalles de confiance, les intervalles que l'on obtiendrait en utilisant l'ensemble du jeu de données serait très mince et de ce fait serait difficile à visualiser.

Nous allons en premier lieu estimer la fonction de répartition empirique avec toutes les données et échantillonner sur cette fonction pour obtenir un jeu de données avec lequel il est plus facile de travailler. Pour obtenir une période de retour qui ressemble à la période de retour sur l'ensemble des données, nous allons inclure les évènements extrêmes dans notre échantillon et échantillonner sur le reste. Cela est nécessaire, car comme les évènements extrêmes sont très rares, ils ne feraient pas partie d'un échantillon réduit de l'ensemble de données.

La figure 1.2 montre les fonctions de répartition empirique pour les données originales et notre échantillon de 1000 journées incluant les évènements les plus extrêmes. Plus exactement, on présente les couples $\{x_i, F_n(x_i)\}_{i=1}^n$ pour chaque échantillon. Nous remarquons que le grand nombre de jours sans précipitations (plus de la moitié) se traduit par une fonction de répartition qui commence à plus de 0.5. Aussi, la fonction de répartition croît vite vers 1 puisque la plupart des jours ont peu de précipitations.

Autant dans la figure 1.2 pour la fonction de répartition que dans la figure 1.3 pour la période de retour, nous remarquons que les courbes sont comparables et que les intervalles de confiance en résultant le seront donc. Notez que

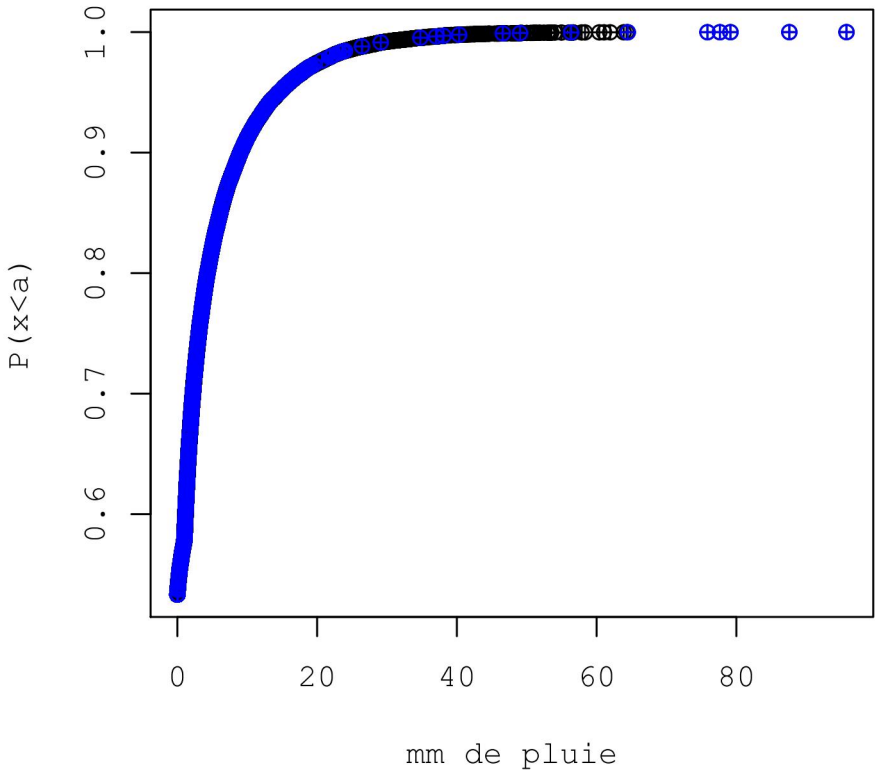


FIGURE 1.2. Fonction de répartition empirique pour l'ensemble des données (noir) et pour un échantillon de 1000 journées (bleu).

dans les deux figures, une partie de la courbe pour l'ensemble des données est cachée par la courbe de l'échantillon puisque les deux courbes se superposent parfaitement. Bien entendu, tous les intervalles présentés dans ce mémoire seront basés sur l'échantillon et seront donc plus larges que si nous avions considéré l'ensemble des données.

1.4. MODÈLE GÉNÉRAL

Le but des sections suivantes sera de développer un modèle qui nous permet de calculer un intervalle de confiance simultané sur la fonction de répartition de notre jeu de données. La fonction de répartition empirique est un bon estimateur non paramétrique de la fonction de répartition. Par contre,

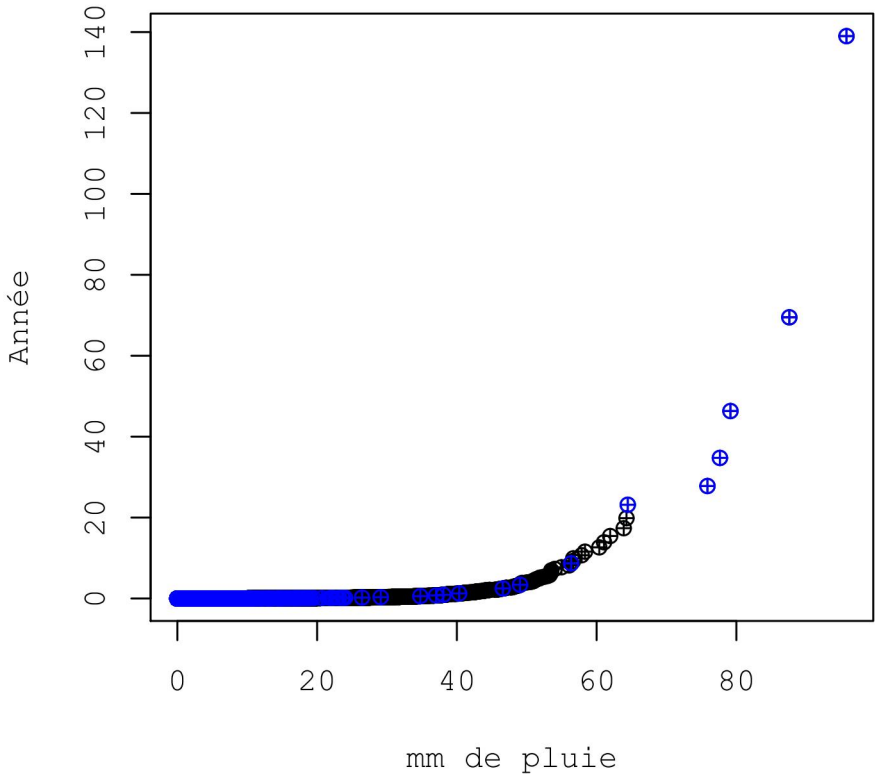


FIGURE 1.3. Période de retour empirique pour l'ensemble des données (noir) et pour un échantillon de 1000 journées (bleu).

comme cette dernière est continue et que la fonction de répartition empirique ne l'est pas, nous allons utiliser différentes méthodes pour obtenir un estimateur continu.

Nous nous placerons dans un contexte où la variable indépendante sera X, la quantité de pluie tombée durant une journée, et la variable dépendante sera la fonction de répartition empirique calculée à partir d'un échantillon de n variables X indépendantes et identiquement distribuées. On fera donc une régression sur les couples $\{x_i, F_n(x_i)\}_{i=1}^n$. Plus généralement, on aura le modèle suivant :

$$F_n(x_i) = C(x_i) + \epsilon_i, i = 1, ..., n,$$

où les erreurs seront centrées, indépendantes et identiquement distribuées et où $C(x)$ est une fonction que l'on voudra monotone et bornée entre 0 et 1, mais afin de développer notre approche à partir d'un modèle simple, nous commencerons notre étude sans imposer de contraintes sur $C(x)$. On a, par le théorème central limite, que

$$E[F_n(x)] = E\left[\frac{(\#\{x : x \leq a\} + 1)}{n+2}\right] = \frac{nE[F_n^*(x)] + 1}{n+2} = \frac{nF(x) + 1}{n+2} \underset{n\to\infty}{\longrightarrow} F(x)$$

où $F_n^*(x) = 1/n \sum_{i=1}^{n} \mathbb{I}_{x_i \leq x}$ est la fonction de répartition empirique non modifiée. De plus, puisque que notre modèle suppose que $E[F_n(x)] = C(x)$, il faut que $F(x)$ soit dans la famille des fonctions pouvant être estimées par $C(x)$. La base qui sera utilisée au dernier chapitre a une très grande flexibilité et permet d'estimer une très grande variété de fonctions. On pourra donc supposer que l'hypothèse que $F(x)$ fait partie de la famille des courbes de $C(x)$ est respectée.

L'hypothèse sur les erreurs permet d'obtenir plusieurs propriétés intéressantes sur les distributions des différents estimateurs obtenues par régression. Par contre, ici comme la variable dépendante est la fonction de répartition empirique, le fait que les erreurs sont indépendantes n'est clairement pas respectée et le fait qu'elles soient identiquement distribuées n'est pas respectée non plus. Toutefois, tout comme dans Merleau et al. (2007), nous supposerons quand même l'indépendance et l'homoscédasticité des erreurs. Aussi, pour que nos données se rapprochent plus de nos hypothèses, nous introduirons au chapitre 3 la transformation angulaire qui permettra de réduire la covariance entre les erreurs. Des simulations aux derniers chapitres viendront confirmer que notre approche est acceptable.

Nous utiliserons d'abord la régression linéaire polynomiale pour approximer la fonction de répartition empirique. Par la suite, nous étudierons différentes façons d'obtenir des intervalles de confiance simultanés sur une courbe de régression. Nous ferons cela d'abord dans un contexte classique, c'est-à-dire

en considérant les coefficients de la courbe de régression comme des valeurs fixes inconnues, puis dans un contexte bayésien où les coefficients de la courbe de régression seront considérés comme des variables aléatoires.

Nous étudierons dans la prochaine section comment estimer nos fonctions et obtenir des intervalles de confiance sous une approche classique. Cela se fera à l'aide de la régression linéaire et d'une base polynomiale.

1.5. RÉGRESSION LINÉAIRE

La régression linéaire multiple est basée sur l'hypothèse que la variable dépendante (y) est une fonction linéaire des variables indépendantes (x) à laquelle une erreur aléatoire a été ajoutée. Pour n observations avec p variables indépendantes on peut l'écrire de cette façon :

$$y_i = B_0 + B_1 x_{i,1} + \ldots + B_p x_{i,p} + \epsilon_i, \epsilon_i \sim N(0, \sigma^2), i = 1, ..., n,$$

où les ϵ_i sont indépendants les uns des autres. Ici, on utilisera plutôt la notation vectorielle où $x_i = (x_{i,1}, ..., x_{i,p}), X = [x_1, ..., x_n]^T, Y = (y_1, ..., y_n)^T, B = (B_0, ..., B_p)^T$ et $\epsilon = (\epsilon_1, ..., \epsilon_n)^T$, le modèle devient donc

$$Y = XB + \epsilon.$$

Les coefficients seront estimés par leur estimateur des moindres carrés, c'est-à-dire, par les $\hat{B}$ qui minimisent $\epsilon^T\epsilon = \sum_{i=1}^{n}(Bx_i - y_i)^2 = Q(B)$. Les $\hat{B}$ peuvent être trouvés facilement en dérivant la fonction précédente :

$$\frac{\partial Q(B)}{\partial B} = 0 \Leftrightarrow X^T(Y - XB) = 0.$$

Donc nous devons avoir que $X^TY = X^TXB$ d'où l'on obtient que $\hat{B} = (X^TX)^{-1}X^TY$. Remarquons que X^TX doit être inversible pour obtenir l'estimateur de B et cela implique que l'on doit avoir des variables linéairement indépendantes. Si ce n'est pas le cas, une des variables est inutile pour faire la régression et nous

pouvons l'enlever. En pratique, il peut aussi y avoir des problèmes pour calculer l'inverse lorsque les variables sont presque colinéaires. Il existe différents critères numériques permettant de calculer si c'est le cas. Notons que $\hat{B}$ est sans biais et que c'est l'estimateur de variance minimale parmi les estimateurs linéaires de B par le théorème de Gauss-Markov.

Si les variables sont presque colinéaires, ce qui est souvent le cas lorsque l'on a un très grand nombre de variables, il est courant d'utiliser des méthodes minimisant la somme des carrés des résidus avec une pénalisation sur les coefficients. Notons entre autres la méthode LASSO (Tibshirani, 1996) et SCAD (Fan et Li, 2001). Ces méthodes ont l'avantage de sélectionner les variables en mettant à 0 les coefficients des variables non significatives et de trouver les coefficients même si la matrice X^TX est difficilement inversible.

Comme nous l'avons mentionné précédemment, l'hypothèse que les erreurs sont indépendantes et identiquement distribuées suivant une $N(0, \sigma^2)$ est utile pour dériver la distribution de Y et $\hat{B}$. Nous dérivons classiquement que $Y \sim N_n(XB, \sigma^2 I)$ et que $\hat{B} \sim N_p(B, \sigma^2(X^TX)^{-1})$. De plus, nous aurons l'estimateur de Y,

$$\hat{Y} = X\hat{B} \sim N_n(XB, \sigma^2 X(X^TX)^{-1}X^T).$$

L'erreur sera estimée par le résidu $\hat{\epsilon}$, où $\hat{\epsilon}_i = y_i - \hat{y}_i$. Notons que

$$\hat{\epsilon} \sim N_n(0, \sigma^2(I - X(X^TX)^{-1}X^T)).$$

Notons que dans les deux distributions précédentes, les matrices de covariances sont singulières puisqu'elles ne sont pas de rang complet. Nous pouvons, à partir des résidus, trouver l'estimateur de σ^2 : $\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} \hat{\epsilon}_i^2}{n-p}$. Cet estimateur est sans biais puisque

$$E(\hat{\sigma}^2) = \sum_{i=1}^{n} \frac{E(\hat{\epsilon}_i^2)}{n-p} = \sum_{i=1}^{n} \frac{Var(\hat{\epsilon}_i)}{n-p} = \frac{trace(Var(\hat{\epsilon}))}{n-p}$$

$$
\begin{aligned}
&= \sigma^2 \frac{n - trace(X(X^TX)^{-1}X^T)}{n-p} = \sigma^2 \frac{n - trace((X^TX)^{-1}X^TX)}{n-p} \\
&= \sigma^2 \frac{n - trace(I_p)}{n-p} = \sigma^2.
\end{aligned}
$$

De plus, $\frac{n-p}{\sigma^2}\hat{\sigma}^2 \sim \chi^2_{n-p}$.

Comme σ^2 n'est pas connu, les distributions précédentes de $\hat{Y}$ et de $\hat{B}$ sont inconnues. Les distributions de $\hat{Y}$ et $\hat{B}$ basées sur l'estimateur $\hat{\sigma}^2$ vont donc correspondre à des Student multivariées de $n-p$ degrés de liberté. Plus précisément :

$$
\frac{\hat{Y} - XB}{\hat{\sigma}} \sim T_n(0, X(X^TX)^{-1}X^T, n-p) \text{ et } \frac{\hat{B} - B}{\hat{\sigma}} \sim T_p(0, (X^TX)^{-1}, n-p),
$$

où, $T_n(0, X(X^TX)^{-1}X^T, n-p)$ représente par exemple la distribution d'une Student de dimension n, centrée en 0, avec $X(X^TX)^{-1}X^T$ comme covariance et $n-p$ degrés de liberté (voir annexe 5 de Robert, 2001 et Taboga, 2010 p.452 pour plus de détails). Dans notre cas, comme le nombre de degrés de liberté est très grand ($\gg 100$), les densités de Student seront très proches de lois normales et l'on pourra donc directement approcher les distributions par des lois normales. Aussi, nous aurons besoin du fait que

$$
\begin{aligned}
\frac{(\hat{B} - B)^T(X^TX)^{-1}(\hat{B} - B)}{p\hat{\sigma}^2} &= \frac{\frac{(\hat{B}-B)^T(X^TX)^{-1}(\hat{B}-B)}{p\sigma^2}}{\frac{(n-p)\hat{\sigma}^2}{(n-p)\sigma^2}} \\
&= \frac{(\hat{B} - B)^T(\sigma^2X^TX)^{-1}(\hat{B} - B)/p}{\frac{(n-p)\hat{\sigma}^2}{\sigma^2}/(n-p)} \\
&\sim \frac{\chi^2_p/p}{\chi^2_{n-p}/(n-p)} \sim Fisher(p, n-p). \quad (1.5.1)
\end{aligned}
$$

Finalement, pour une nouvelle observation, $y^* = x^*B + \epsilon^*$, si on suppose que ϵ^* suit aussi une $N(0, \sigma^2)$ et est indépendant des autres observations. On obtient que

$$
\hat{y}^* - y^* \sim N\left(0, \sigma^2[x^{*T}(X^TX)^{-1}x^* + 1]\right).
$$

1.6. INTERVALLE DE CONFIANCE

À partir de ces distributions on peut dériver des intervalles de confiance. Tout d'abord, pour chaque x, on peut obtenir un intervalle de confiance ponctuel de la moyenne $E(y|x)$ à 95% : $\hat{y} \pm t^{0,025}_{n-p}\hat{\sigma}(x^T(X^TX)^{-1}x)^{1/2}$. L'intervalle de prédiction est très semblable : $\hat{y}^* \pm t^{0,025}_{n-p}\hat{\sigma}(x^T(X^TX)^{-1}x + 1)^{1/2}$. Ces intervalles sont ponctuels, c'est-à-dire qu'ils couvrent 95% pour chaque valeur de x individuellement sans considérer les autres valeurs. Nous utiliserons la technique de Scheffé (Seber, 1977, p.134) pour obtenir un intervalle sur la courbe (appelé aussi intervalle de confiance simultané, car on veut considérer l'ensemble des valeurs simultanément).

La méthode de Scheffé repose sur le fait que $\frac{(\hat{B}-B)^TX^TX(\hat{B}-B)}{p\hat{\sigma}^2} \sim F_{p,n-p}$. À partir de là, nous pouvons montrer (Seber, 1977, p.128) que

$$1 - \alpha = Pr\left[\frac{|x^T\hat{B} - x^TB|}{\hat{\sigma}(x^T(X^TX)^{-1}x)^{1/2}} \leq (pF^{\alpha}_{p,n-p})^{1/2}, \forall x\right].$$

On peut donc obtenir un intervalle de confiance simultané à 95% qui est :

$$x^T\hat{B} \pm (pF^{0,05}_{p,n-p})^{1/2}\hat{\sigma}(x^T(X^TX)^{-1}x)^{1/2}\ \forall x = (1, x_1, ..., x_{p-1}). \tag{1.6.1}$$

De la même façon que les intervalles de confiance ponctuels peuvent être transformés pour s'appliquer à des valeurs prédites, les intervalles de confiance simultanés de Scheffé peuvent l'être aussi. Les intervalles de prévision simultanés deviennent alors :

$$x^T\hat{B} \pm (pF^{0.05}_{p,n-p})^{1/2}\hat{\sigma}(x^T(X^TX)^{-1}x + 1)^{1/2}\ \forall x = (1, x_1, ..., x_{p-1}).$$

Nous considérerons aussi une approche non paramétrique basée sur le test de Kolmogorov-Smirnov (Massey, 1951). Ce test sert à tester si un échantillon provient d'une certaine loi de probabilité. Pour se faire, il faut calculer la statistique de Kolmogorov-Smirnov qui est l'écart le plus grand entre la fonction

de répartition empirique découlant de nos n données avec la fonction de répartition théorique ($F_0(x)$) :

$$D_n = \sup_x |F_n^*(x) - F_0(x)|,$$

où $F_n^*(x)$ est la fonction de répartition empirique non modifiée. La variable aléatoire $\sqrt{n}D_n$ converge vers la distribution de Kolmogorov sous l'hypothèse que l'échantillon provient de la loi de $F_0(x)$. Donc,

$$P\left(D_n \leq \frac{K_\alpha}{\sqrt{n}}\right) = 1 - \alpha,$$

et nous pouvons construire un intervalle de confiance conservateur pour la F_0 :

$$F_0 \in F_n(x) \pm \frac{K_\alpha}{\sqrt{n}}.$$

Cet intervalle est conservateur car il considère que l'écart maximal entre la fonction de répartition empirique et la fonction de répartition théorique a lieu sur l'ensemble du domaine. Notons que la valeur critique utilisée ici sera $K_{0,05} = 1,358$.

On peut raffiner cet intervalle de confiance puisque l'on sait que la fonction de répartition est bornée entre 0 et 1. Nous utiliserons donc la même approche que Wang et al. (2013) et les bornes inférieures et supérieures seront respectivement $max(F_n(x) - \frac{K_\alpha}{\sqrt{n}}, 0)$ et $min(F_n(x) + \frac{K_\alpha}{\sqrt{n}}, 1)$.

Pour obtenir un intervalle de confiance sur la période de retour, nous n'avons qu'à transformer l'intervalle de confiance de la fonction de répartition. En effet, comme la période de retour est une fonction bijective, la probabilité que $x^T B$ se trouve dans un intervalle est la même que la probabilité que $PdR(x^T B)$ se trouve dans l'intervalle transformé :

$$\begin{aligned} 1 - \alpha &= Pr\left[|x^T B - x^T \hat{B}| \leq F_\alpha, \forall x = (1, x_1, ..., x_{p-1})\right] \\ &= Pr\left[x^T \hat{B} - F_\alpha \leq x^T B \leq x^T \hat{B} + F_\alpha, \forall x = (1, x_1, ..., x_{p-1})\right] \end{aligned}$$

$$= \Pr\left[PdR(x^T\hat{B} - F_\alpha) \leq PdR(x^TB) \leq PdR(x^T\hat{B} + F_\alpha), \forall x = (1, x_1, ..., x_{p-1})\right],$$

où $F_\alpha = (pF^\alpha_{p,n-p})^{1/2}\hat{\sigma}(x^T(X^TX)^{-1}x)^{1/2}$. Donc, l'intervalle de confiance pour la période de retour est directement la transformation de l'intervalle de confiance de la fonction de répartition.

La méthode delta n'est donc pas nécessaire pour obtenir l'intervalle de confiance sur la période de retour puisque celle-ci est bijective. Par contre, si la fonction d'intérêt était une fonction non bijective, mais deux fois dérivable, nous pourrions utiliser la méthode delta pour obtenir une approximation de l'intervalle de confiance.

1.6.1. Intervalle de crédibilité

Nous nous plaçons maintenant dans un contexte bayésien en considérant les B non plus seulement comme des valeurs fixes inconnues, mais plutôt comme des variables aléatoires pour introduire une nouvelle méthode de calcul d'un intervalle de confiance simultané pour la courbe. Dans ce contexte, nous ne parlerons plus d'intervalle de confiance mais plutôt d'intervalle de crédibilité.

De la section précédente, nous avons que

$$\hat{B} \sim N_p(B, \sigma^2(X^TX)^{-1}).$$

Si nous on suppose maintenant que B est aléatoire et que les données sont fixes on obtient à partir de la distribution de $\hat{B}$:

$$\begin{aligned}
\pi(\hat{B}) &= \frac{|X^TX|^{\frac{1}{2}}}{(2\pi)^{\frac{p}{2}}(\sigma^2)^{\frac{p}{2}}} exp\left(-\frac{1}{2\sigma^2}(\hat{B} - B)^TX^TX(\hat{B} - B)\right) \\
&= \frac{|X^TX|^{\frac{1}{2}}}{(2\pi)^{\frac{p}{2}}(\sigma^2)^{\frac{p}{2}}} exp\left(-\frac{1}{2\sigma^2}(B - \hat{B})^TX^TX(B - \hat{B})\right) \\
&= \pi(B|\mathcal{D})
\end{aligned}$$

qui correspond à la densité de B sachant les données comme nous le verrons au chapitre 2. On obtient donc que $B|\mathcal{D} \sim N_p(\hat{B}, \sigma^2(X^TX)^{-1})$.

L'idée générale de notre méthode pour obtenir un intervalle de crédibilité simulé est de tirer des B de leur distribution afin d'obtenir un ensemble de courbes probables et de prendre l'enveloppe de ces courbes comme bornes.

En fait, on cherche un ensemble $\mathcal{B}$ de valeurs de B tel que $P(B \in \mathcal{B}|\mathcal{D}) = 1 - \alpha$ de telle façon qu'on aura que $P(x^TB \in x^T\mathcal{B}\,\forall x|\mathcal{D}) \geq 1 - \alpha$. On aura alors que $x^T\mathcal{B}$ est un ensemble de crédibilité d'au moins $1 - \alpha$. Les bornes de cet ensemble formeront notre intervalle de crédibilité.

À partir de (1.5.1), on a que

$$\frac{(\hat{B} - B)^TX^TX(\hat{B} - B)}{\hat{\sigma}^2} < \chi^2_p(0,05). \tag{1.6.2}$$

Le côté gauche de (1.6.2) suit une χ^2_p plutôt qu'une $F_{p,n-p}$ puisque $\hat{\sigma}^2$ est considéré comme fixe et non pas comme aléatoire puisque l'on conditionne par rapport aux données.

L'ensemble des B qui respectent (1.6.2) est notre ensemble de crédibilité $\mathcal{B}$. Pour avoir un intervalle à 95%, nous allons donc garder les $B \in \mathcal{B}$. De cette façon, en tirant beaucoup de B (par exemple 1000), on va obtenir un ensemble de crédibilité simulé pour la courbe à partir de l'équation suivante :

$$\hat{y_i} = x_i^TB. \tag{1.6.3}$$

On peut représenter l'intervalle de crédibilité simulé de la courbe en prenant l'enveloppe des courbes simulées. Pour chaque x_i, on obtiendra la valeur maximale et la valeur minimale et nous calculerons une courbe lissée de ces valeurs pour la borne supérieure (que l'on notera $SupF_n(x)$) et la borne inférieure (que l'on notera $InfF_n(x)$). Ici, nous utiliserons la méthode loess() de R pour obtenir ces bornes qui seront les bornes de notre intervalle de crédibilité simulé.

Notons que la méthode par simulation n'est pour l'instant pas nécessaire car la méthode de Scheffé est disponible. Par contre, quand on utilisera des modèles *a priori* complexes sur les hyperparamètres au chapitre 2, la méthode de Scheffé ne sera plus valide et nous devrons utiliser cette méthode de simulation.

1.7. BASE POLYNOMIALE

Notre but est donc d'estimer une fonction $h(x)$ où x est unidimensionnelle et $h(x)$ a une forme complexe non linéaire. On ne peut donc pas se baser sur le modèle simple de régression linéaire, $y = B_0 + B_1 x + \epsilon$, car ce modèle ne peut qu'estimer des fonctions linéaires de x. Nous devons donc utiliser une base pour que notre modèle puisse capturer une fonction non linéaire en x. Le plus simple est d'utiliser une base polynomiale où $x^T = (x^0, x^1, x^2, ..., x^d)$, d étant le degré du polynôme. Notre modèle sera donc : $Y = X^T B + \epsilon$.

Le degré du polynôme nécessaire est directement lié à la complexité de la fonction à estimer. Avec $d \geq n - 1$, il est possible d'obtenir une régression parfaite sur les x, mais la fonction obtenue sera alors très chaotique. On parle alors de surajustement : le modèle prédit parfaitement pour les points de notre ensemble de données, mais donne des prédictions erratiques en dehors de celui-ci. Il faut donc faire un compromis entre une fonction qui estime parfaitement nos observations et une fonction qui est la plus lisse possible.

Un bon moyen de contrer le surajustement est d'utiliser la validation croisée. Pour chaque modèle considéré, on ajuste le modèle sur toutes les observations sauf une et l'on prédit l'observation manquante à l'aide de ce modèle. On fait ensuite la somme des erreurs obtenues au carré qui est notre mesure de qualité du modèle. On calcule ainsi la statistique PRESS (predicted residuals sum of squares) (Allen, 1974 et Weisberg, 1985). Un bon modèle aura une faible valeur de PRESS.

Notez que nous utilisons une version simple de la validation croisée, plusieurs variantes existent et elles sont très utilisées particulièrement dans le domaine des algorithmes d'apprentissage où l'on utilise des modèles qui peuvent souvent s'ajuster parfaitement aux données.

TABLEAU 1.1. PRESS pour différents degrés de la base polynomiale.

d	SSE	PRESS
3	1,535	2,129
4	0,558	1,211
5	0,174	0,516
6	0,087	0,400
7	0,066	0,299
8	0,064	0,211
9	0,063	8,961
10	0,057	43,468

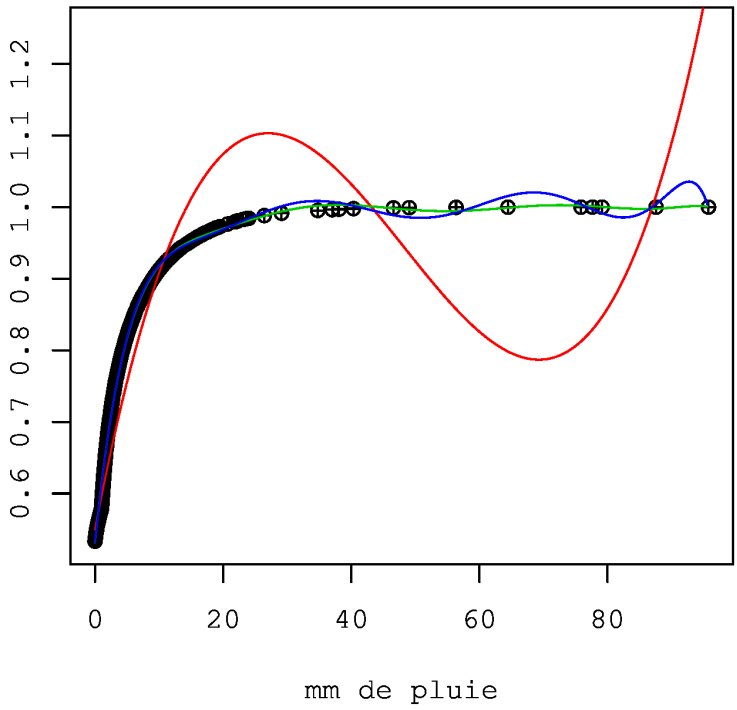


FIGURE 1.4. Fonction de répartition empirique ajustée par un polynôme de degré 3 (rouge), degré 8 (vert) et degré 9 (bleu).

Le tableau 1.1 et la figure 1.4 montrent clairement le phénomène de surajustement : bien que la somme des résidus (SSE) diminue avec l'augmentation du degré, nous voyons clairement qu'à partir du degré 9, il y a surajustement. Nous avons donc choisi un polynôme de degré 8. Remarquons aussi que la fonction estimée ainsi que les intervalles de confiance dépassent 1 ce qui est problématique puisqu'il est impossible que la fonction de répartition dépasse 1.

1.8. RÉSULTATS

La figure 1.5 présente les couples $\{x_i, F_n(x_i)\}_{i=1}^n$ avec la fonction de répartition obtenue par régression polynomiale et les intervalles de confiance ponctuel, simultané (Scheffé) et de Kolmogorov qui y sont liés. On remarque que les intervalles simultanés sont plus larges que les intervalles ponctuels. Aussi, l'intervalle de Kolmogorov a l'avantage d'être plus lisse que l'intervalle de Scheffé puisqu'il ne dépend pas de la base polynomiale.

Dans la figure 1.6, nous montrons la période de retour estimée à partir de la régression polynomiale sur la fonction de répartition empirique :

$$PDR(x_i) = \frac{1}{365(1 - C(x_i))},$$

où $C(x_i)$ est ici la courbe obtenue par régression polynomiale. Les intervalles de confiance ne sont pas calculés car ils n'ont pas de sens pour les raisons qui suivent. Nous voyons qu'il y a des problèmes dès que l'on dépasse 30 mm de pluie. Cela est dû au fait que la fonction de répartition estimée n'est pas bornée à 1, donc quand nous transformons vers la période de retour, il y a divisions par des valeurs très près de 0, négatives ou positives. L'estimation de la période de retour ainsi que ses intervalles de confiance n'ont donc presque aucune utilité.

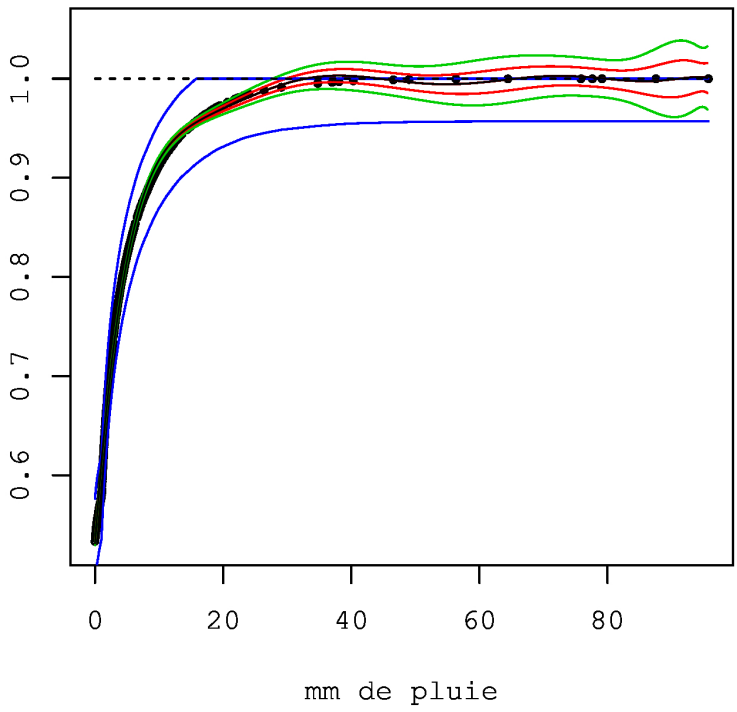


FIGURE 1.5. Intervalles de confiance ponctuel (rouge), simultané (vert) et de Kolmogorov borné (bleu) pour la fonction de répartition.

Il serait possible de mettre une contrainte sur la fonction estimée pour qu'elle ne dépasse pas 1. Cela entraînerait des contraintes sur les coefficients, mais ensuite résoudre ce problème d'optimisation serait complexe. Nous résoudrons ce problème à l'aide d'une base non linéaire au chapitre 3.

Pour pouvoir mieux visualiser les résultats, nous regarderons seulement la fonction de répartition à partir de maintenant et nous reviendrons à la période de retour au dernier chapitre.

Nous comparons maintenant les intervalles de crédibilité obtenus par simulation à partir de l'équation (1.6.3) avec les intervalles de confiance de Scheffé. Comme nous le voyons dans la figure 1.7, les intervalles de crédibilité simulés sont équivalents aux intervalles de confiance simultanés.

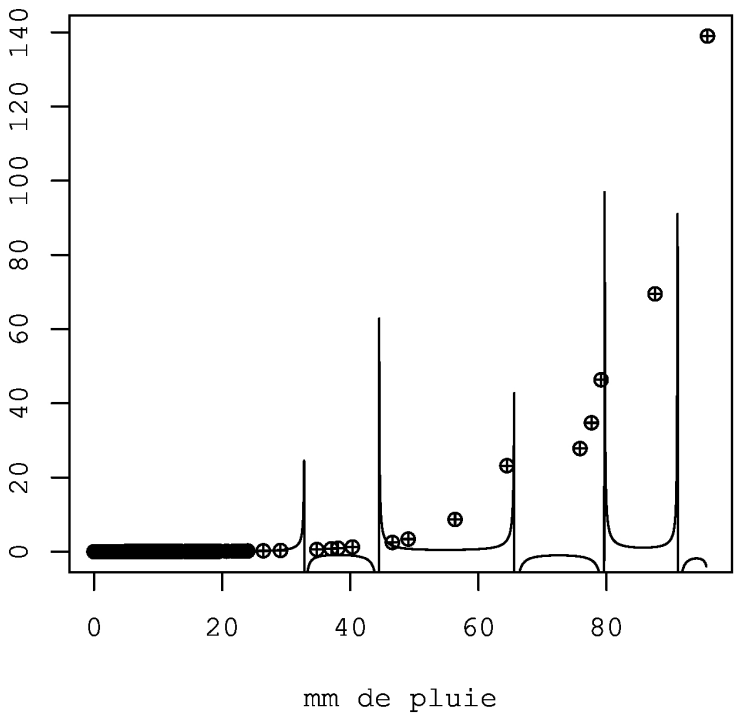


FIGURE 1.6. Période de retour calculée à partir de la fonction de répartition estimée.

Les intervalles de crédibilité simulés sont justifiés dans le cas où il est impossible de dériver des intervalles de confiance simultanés théoriques ou dans le cas où l'on dispose déjà de courbes simulées comme ce sera le cas pour certains modèles bayésiens que nous verrons plus loin.

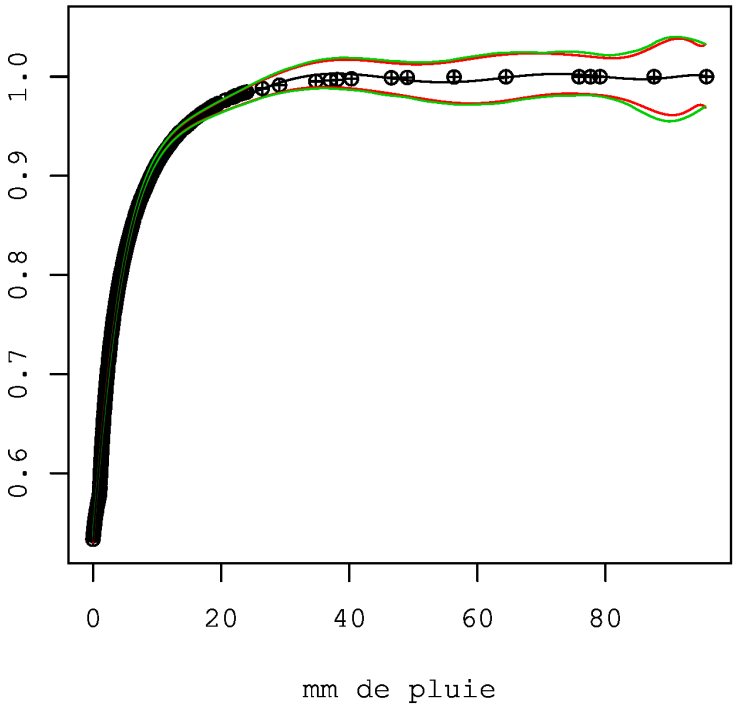


FIGURE 1.7. Intervalles de confiance simultanés de Scheffé (rouge) et intervalle de crédibilité simulés (vert) pour la fonction de répartition.

Chapitre 2

APPROCHE BAYÉSIENNE

2.1. INTRODUCTION

Jusqu'à maintenant (sauf pour la fin de la section 1.6), nous avons supposé que les coefficients des courbes de régression suivaient des lois normales avec certains paramètres. La statistique classique suppose qu'une vraie valeur pour ces paramètres existe et est fixe. En statistique bayésienne, on suppose que les paramètres des distributions suivent eux aussi des distributions de la même façon que les paramètres d'une courbe suivent une distribution. Les distributions sur les paramètres sont appelées distributions *a priori* et les distributions découlant de celles-ci qui incluent l'information des données sont appelées distribution *a posteriori*. Soit un échantillon $x_1, ..., x_n$ indépendant et identiquement distribué provenant d'une distribution supposée connue $f(x|\theta)$ que l'on verra plutôt comme la fonction de vraisemblance de θ, $l(\theta|x)$, nous noterons $\pi(\theta)$ la distribution *a priori* et avec le théorème de Bayes, nous obtiendrons la distribution *a posteriori*

$$\pi(\theta|x) = \frac{l(\theta|x)\pi(\theta)}{\int l(\theta|x)\pi(\theta)d\theta}. \tag{2.1.1}$$

Il existe plusieurs façons de poser les distributions *a priori* que l'on veuille inclure de l'information dans ces distributions ou bien plutôt que l'on veuille les laisser le plus vague possible afin d'induire le moins de contraintes sur la

distribution finale des coefficients de la courbe. Aussi, il faut prendre en considération les difficultés techniques pouvant ressortir du calcul de la distribution *a posteriori*. En effet, il est parfois (souvent) impossible d'obtenir une forme analytique. Certains de ces cas peuvent être contournés avec des méthodes de simulation, mais on préférera souvent choisir un *a priori* qui facilitera les calculs.

Les distributions *a priori* conjuguées sont une famille de distributions *a priori* qui font en sorte que $\pi(\theta|x)$ sera de la même forme que $\pi(\theta)$, on dira alors que $\pi(\theta)$ est une distribution *a priori* conjuguée pour la fonction de vraisemblance $l(\theta|x)$. En particulier, nous savons que pour toutes les distributions de la famille exponentielle, il existe toujours une distribution *a priori* conjuguée. Ici, nous noterons plus précisément qu'une normale est sa propre famille conjuguée : soit une fonction de vraisemblance normale, si nous posons une distribution *a priori* normale sur le paramètre de position, nous obtiendrons une distribution *a posteriori* normale pour celui-ci.

2.2. RÉGRESSION LINÉAIRE BAYÉSIENNE

Nous nous intéressons ici à apporter de l'information sur la distribution des coefficients dans la régression linéaire $Y = XB + e$. Nous étudierons différents modèles où nous supposerons des distributions *a priori* sur différents paramètres de la régression. Débutons avec un modèle où l'*a priori* prend une forme générale et calculons la loi *a posteriori* qui en découle. Nous nous baserons ensuite sur ce résultat pour obtenir les lois *a posteriori* des autres modèles.

Tout d'abord, dans le modèle linéaire la densité est

$$Y|B, \sigma^2 \sim N_n(XB, \sigma^2 I)$$

et comme nous savons que l'on obtiendra une densité *a posteriori* normale en posant une densité *a priori* normale, nous poserons la loi *a priori* suivante :

$$B|B_0, \sigma^2, \Sigma_0 \sim N_p(B_0, \sigma^2\Sigma_0),$$

où les hyperparamètres B_0, σ^2 et Σ_0 sont pour l'instant supposés connus. Nous reviendrons plus loin dans ce chapitre sur la façon de fixer ces hyperparamètres. Nous aurons alors que

$$\pi(B|B_0, \sigma^2, \Sigma_0) = \frac{1}{(2\pi)^{\frac{p}{2}}(\sigma^2)^{\frac{p}{2}}|\Sigma_0|^{\frac{1}{2}}} exp\left(-\frac{1}{2\sigma^2}(B - B_0)^T\Sigma_0^{-1}(B - B_0)\right),$$
$$f(Y|B, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} exp\left(-\frac{1}{2\sigma^2}(Y - XB)^T(Y - XB)\right).$$

Nous pouvons décomposer la forme quadratique dans la distribution de Y de la façon suivante :

$$\begin{aligned}(Y - XB)^T(Y - XB) &= (Y - X\hat{B})^T(Y - X\hat{B}) + (B - \hat{B})^TX^TX(B - \hat{B}) \\ &= SSE + (B - \hat{B})^TX^TX(B - \hat{B}).\end{aligned}$$

Calculons maintenant la densité *a posteriori* :

$$\begin{aligned}\pi(B|\mathcal{D}, B_0, \sigma^2, \Sigma_0) &\propto \pi(B|B_0, \sigma^2, \Sigma_0)l(Y|B, \sigma^2) \\ &\propto exp\left(\frac{-1}{2\sigma^2}(B - B_0)^T\Sigma_0^{-1}(B - B_0)\right) \\ &\quad \times exp\left(\frac{-1}{2\sigma^2}\left(SSE + (B - \hat{B})^TX^TX(B - \hat{B})\right)\right) \\ &\propto exp\left(\frac{-1}{2\sigma^2}\left((B - B_0)^T\Sigma_0^{-1}(B - B_0) + (B - \hat{B})^TX^TX(B - \hat{B})\right)\right) \\ &\propto exp\left(\frac{-1}{2\sigma^2}(B - B_n)^T\Sigma_n^{-1}(B - B_n)\right),\end{aligned}$$

où $B_n = \Sigma_n\left(X^TX\hat{B} + \Sigma_0^{-1}B_0\right)$ et $\Sigma_n = \left(X^TX + \Sigma_0^{-1}\right)^{-1}$.

La loi *a posteriori* est donc

$$B|\mathcal{D}, B_0, \sigma^2, \Sigma_0 \sim N_p(B_n, \sigma^2\Sigma_n).$$

2.2.1. *A priori* de type G

Il existe une méthode pour avoir une densité *a posteriori* simple : la méthode des densités *a priori* de type G (Tiao et Zellner, 1964 et Geinitz, 2009). En choisissant de façon judicieuse les paramètres de la loi normale *a priori*, la distribution *a posteriori* sera particulièrement intéressante. Soit la loi *a priori*

$$B|B_0, \sigma^2, X \sim N_p(B_0, g\sigma^2(X^TX)^{-1}),$$

nous posons donc $\Sigma_0 = g(X^TX)^{-1}$ et nous aurons alors que la loi *a posteriori* résultante sera

$$\begin{aligned} B|\mathcal{D}, B_0, \sigma^2 &\sim N_p\left((X^TX + \frac{1}{g}X^TX)^{-1}(X^TX\hat{B} + \frac{1}{g}X^TXB_0), \sigma^2(X^TX + \frac{1}{g}X^TX)^{-1}\right) \\ &\sim N_p\left(\frac{1}{g+1}(B_0 + g\hat{B}), \frac{g\sigma^2}{g+1}(X^TX)^{-1}\right). \end{aligned}$$

Cette forme a l'avantage de nous donner une façon facile de pondérer l'information *a priori* (B_0). Si nous posons $g = 1$, l'information provenant des données aura un poids égal à l'information provenant de l'*a priori* et lorsque g tend vers l'infini, l'information *a priori* prend une importance qui tend vers 0 et nous revenons au modèle issu de la statistique classique. Nous pouvons donc facilement comparer les intervalles de confiance du modèle avec *a priori* de type G et les intervalles de confiance classiques : en choisissant un g très grand, les intervalles devraient être les mêmes.

2.2.2. Matrice de covariance quelconque

Les densités *a priori* de type G forcent les coefficients de la courbe de régression à avoir une matrice de covariance *a priori* qui a une forme fixe et il

serait intéressant de pouvoir considérer un modèle où cette matrice est libre. Nous nous replaçons donc dans le contexte du premier modèle étudié. Soit $B|B_0, \sigma^2, \Sigma_0 \sim N_p(B_0, \sigma^2\Sigma_0)$, nous obtenons la loi *a posteriori*

$$B|\mathcal{D}, B_0, \sigma^2, \Sigma_0 \sim N_p\left((X^TX + \Sigma_0^{-1})^{-1}(X^TX\hat{B} + \Sigma_0^{-1}B_0), \sigma^2(X^TX + \Sigma_0^{-1})^{-1}\right).$$

2.2.3. Inverse-gamma

Également, comme nous supposons une distribution sur les B, il est logique d'aussi supposer une distribution sur σ^2. Il est connu qu'une distribution inverse-gamma est une distribution *a priori* conjuguée pour le paramètre d'échelle d'une normale (Merleau et al., 2007). Nous poserons donc

$$\sigma^2|a_0, b_0 \sim \Gamma^{-1}(a_0, b_0)$$
$$\text{et } B|B_0, \sigma^2, \Sigma_0 \sim N_p(B_0, \sigma^2\Sigma_0),$$

où la densité correspondante à la distribution inverse-gamma est

$$\pi(\sigma^2|a_0, b_0) = \frac{b_0^{a_0}}{\Gamma(a_0)}(\sigma^2)^{-(a_0+1)}\exp\left(-\frac{b_0}{\sigma^2}\right)$$

et où a_0 et b_0 sont supposés connus pour le moment.

Avec les définitions de SSE, B_n et Σ_n données précédemment nous obtenons que

$$B, \sigma^2|\mathcal{D}, B_0, \Sigma_0, a_0, b_0 \sim N_p\left(B_n, \sigma^2\Sigma_n\right)\Gamma^{-1}\left(a_n, b_n\right)$$

comme distribution *a posterori* conjointe de B et σ^2 où

$$a_n = a_0 + \frac{n}{2}$$
$$\text{et } b_n = \frac{2b_0 + SSE + (\hat{B} - B_0)^T((X^TX)^{-1} + \Sigma_0)^{-1}(\hat{B} - B_0)}{2}.$$

Par contre, comme nous voulons avoir des intervalles de confiance sur les B, il faut marginaliser la distribution précédente. En intégrant par rapport à σ^2

nous obtenons que

$$\begin{aligned}
\pi(B|\mathcal{D}, B_0, \Sigma_0, a_0, b_0) &= \int_{\mathcal{S}} \pi(B, \sigma^2|\mathcal{D}, B_0, \Sigma_0, a_0, b_0)\, d\sigma^2 \\
&= \int_{\mathcal{S}} \frac{1}{(2\pi)^{\frac{p}{2}}(\sigma^2)^{\frac{p}{2}}|\Sigma_n|^{\frac{1}{2}}} exp\left(\frac{-1}{2\sigma^2}(B - B_n)^\top \Sigma_n^{-1}(B - B_n)\right) \\
&\qquad \times \frac{b_n^{a_n}}{\Gamma(a_n)}(\sigma^2)^{-(a_n+1)} \exp\left(-\frac{b_n}{\sigma^2}\right)\, d\sigma^2 \\
&= (2a_n\pi)^{-\frac{p}{2}} \frac{\Gamma(\frac{2a_n+p}{2})}{\Gamma(a_n)} \left|\frac{b_n}{a_n}\Sigma_n\right|^{-\frac{1}{2}} \\
&\qquad \times \left(1 + \frac{1}{2b_n}(B - B_n)^\top \Sigma_n^{-1}(B - B_n)\right)^{-\frac{2a_n+p}{2}}
\end{aligned}$$

qui est une densité de Student multidimensionnelle à $2a_n$ degrés de liberté, centrée B_n et avec $\frac{b_n}{a_n}\Sigma_n$ comme covariance (voir Taboga, 2010, p. 452). Donc, nous avons que

$$B|\mathcal{D}, B_0, \Sigma_0, a_0, b_0 \sim T_p\left(B_n, \frac{b_n}{a_n}\Sigma_n, 2a_n\right).$$

Pour plus de détails sur les calculs effectués ici, voir Murphy (2007).

2.2.4. Inverse-gamma et inverse-Wishart

La dernière complexification du modèle est l'ajout d'une densité *a priori* sur Σ_0. Ici, il est naturel de poser comme loi *a priori* une inverse-Wishart. Cette distribution est basée sur la distribution de Wishart qui est une généralisation en plusieurs dimensions de la distribution du khi-deux. Nous aurons donc les distributions *a priori* suivantes :

$$\Sigma_0|\nu_0, \Lambda_0 \sim W^{-1}(\nu_0, \Lambda_0),$$

$$\sigma^2|a_0, b_0 \sim \Gamma^{-1}(a_0, b_0)$$

$$\text{et } B|B_0, \sigma^2, \Sigma_0 \sim N_p(B_0, \sigma^2\Sigma_0).$$

Notons que la densité de la distribution inverse-Wishart est

$$\pi(\Sigma_0|\nu_0,\Lambda_0) = \frac{|\Lambda_0|^{\frac{\nu_0}{2}}}{2^{\frac{\nu_0 p}{2}}\Gamma_p(\frac{\nu_0}{2})}|\Sigma_0|^{-\frac{\nu_0+p+1}{2}}\exp\left(-\frac{1}{2}tr(\Lambda_0\Sigma_0^{-1})\right),$$

où la $\Gamma_p(\cdot)$ correspond à la fonction gamma multivariée et où ν_0 et Λ_0 sont supposés connus encore une fois (Robert, 2001).

Ici, les calculs étant plus complexes que précédemment, nous procéderons par étape. Tout d'abord, à la manière de Berger (1985) p.184, on définit les densités marginales de Y de la façon suivante : $m_1(Y|\sigma^2,\Sigma_0)$ est la densité marginale de Y sachant σ^2 et Σ_0, c'est la densité de Y intégrée sur B, $m_2(Y|\Sigma_0)$ est la densité m_1 intégrée sur σ^2 et $m_3(Y)$ est la densité m_2 intégré sur Σ_0. En se servant du fait que

$$\pi(B|\mathcal{D},B_0,\sigma^2,\Sigma_0) = \frac{f(Y|B,\sigma^2,\Sigma_0)\pi(B|B_0,\sigma^2,\Sigma_0)}{m_1(Y|\sigma^2,\Sigma_0)},$$

$$\pi(\sigma^2|\mathcal{D},a_0,b_0,\Sigma_0) = \frac{m_1(Y|\sigma^2,\Sigma_0)\pi(\sigma^2|a_0,b_0,\Sigma_0)}{m_2(Y|\Sigma_0)} \text{ et}$$

$$\pi(\Sigma_0|\mathcal{D},\nu_0,\Lambda_0) = \frac{m_2(Y|\Sigma_0)\pi(\Sigma_0|\nu_0,\Lambda_0)}{m_3(Y)},$$

et du fait que nous pouvons obtenir les distributions marginales à chaque étape, nous dérivons successivement les distributions *a posteriori* une à une. Les détails se trouvent à l'annexe A. Nous trouvons ainsi que

$$B|\mathcal{D},B_0,\sigma^2,\Sigma_0 \sim N_p(B_n,\sigma^2\Sigma_n),$$

$$\sigma^2|\mathcal{D},a_0,b_0,\Sigma_0 \sim \Gamma^{-1}(a_n,b_n) \text{ et}$$

$$\begin{aligned}\pi(\Sigma_0|\mathcal{D},\nu_0,\Lambda_0) &\propto (b_n)^{-a_n}|(X^TX+\Sigma_0^{-1})^{-1}|^{\frac{1}{2}}|\Sigma_0|^{-\frac{\nu_0+p+2}{2}} \\ &\quad\times\exp\left(-\frac{1}{2}tr(\Lambda_0\Sigma_0^{-1})\right) \\ &= h(\Sigma_0). \end{aligned} \tag{2.2.1}$$

Il est impossible de ramener la distribution *a posteriori* de Σ_0 à une distribution connue et il sera donc impossible d'obtenir une forme analytique de $\pi(B|\mathcal{D})$ puisqu'il ne sera pas possible de faire l'intégrale de la densité conjointe.

2.3. INTERVALLE DE CRÉDIBILITÉ SIMULÉ

Rappelons que l'idée générale de notre méthode pour obtenir un intervalle de crédibilité simulé est de tirer des B de leur distribution *a posteriori* afin d'obtenir des courbes probables et de prendre l'enveloppe de ces courbes comme bornes de notre intervalle.

Plus précisément, le but est de trouver un ensemble $\mathcal{B}$ de valeurs de B tel que $P(B \in \mathcal{B}|\mathcal{D}) = 1 - \alpha$ de telle façon qu'on aura que $P(x^T B \in x^T \mathcal{B}\, \forall x|\mathcal{D}) \geq 1-\alpha$. On aura alors que $x^T\mathcal{B}$ est un ensemble de crédibilité d'au moins $1-\alpha$. Les bornes de cet ensemble formeront notre intervalle de crédibilité. Comme on ne peut pas obtenir directement les B qui correspondent aux bornes de notre intervalle, nous devrons en simuler une grande quantité à partir de $\pi(B|\mathcal{D})$ (que nous noterons $B_{(i)}$), pour obtenir une approximation de l'ensemble $\mathcal{B}$ et ensuite prendre l'enveloppe des courbes $x^T B_{(i)}$ obtenues à partir des $B_{(i)}$ simulés pour obtenir notre intervalle.

L'enveloppe des courbes simulées sera fait de la même façon qu'au chapitre 1. Pour chaque x_j de notre échantillon, on obtiendra la valeur maximale et la valeur minimale de $x_j^T B_{(i)}$ et nous calculerons une courbe lissée de ces valeurs pour la borne supérieure (que l'on notera $\mathrm{Sup}F_n(x)$) et la borne inférieure (que l'on notera $\mathrm{Inf}F_n(x)$). Ici, nous utiliserons la méthode loess() de R pour obtenir ces bornes qui seront les bornes de notre intervalle de crédibilité simulé.

Pour approximer $\mathcal{B}$, nous garderons 95% des $B_{(i)}$ simulés en rejetant ceux qui ont la plus faible probabilité *a posteriori* ($\pi(B|\mathcal{D})$).

Dans le modèle le plus complexe présenté à la section 2.2.4, il était impossible d'obtenir une forme analytique de $\pi(B|\mathcal{D})$. Nous allons donc tirer des B

non pas de $\pi(B|\mathcal{D})$ directement, mais plutôt tirer un Σ_0 et un σ^2 et ensuite tirer un B de $\pi(B|\mathcal{D}, \sigma^2, \Sigma_0)$. Pour ce faire, nous devrons faire appel à la technique de simulation Monte-Carlo comme décrit dans la section suivante. De plus, nous devrons approximer $\pi(B|\mathcal{D})$ pour choisir 95% des $B_{(i)}$ simulés ayant la plus grande probabilité *a posteriori* (voir section 2.6.2).

2.4. Simulation Monte-Carlo

Une simulation Monte-Carlo sert à estimer des intégrales de forme

$$\int_\Theta g(\theta) f(x|\theta) \pi(\theta) d\theta. \tag{2.4.1}$$

Elle se base sur le fait que le $\pi(\theta)$ est en fait une distribution de probabilité de θ. On sait que si l'on peut générer M valeurs de θ_i indépendants à partir de $\pi(\theta)$, alors

$$\frac{1}{M} \sum_{i=1}^{M} g(\theta_i) f(x|\theta_i)$$

converge presque sûrement vers (2.4.1) quand M tend vers $+\infty$ en raison de la loi forte des grands nombres.

De la même façon, nous pouvons nous servir de cette technique pour calculer

$$\int_\Theta g(\theta) \pi(\theta|x) d\theta = \frac{\int_\Theta g(\theta) f(x|\theta) \pi(\theta) d\theta}{\int_\Theta f(x|\theta) \pi(\theta) d\theta}$$

qui sera estimé par

$$\frac{\sum_{i=1}^{M} g(\theta_i) f(x|\theta_i)}{\sum_{i=1}^{M} f(x|\theta_i)}, \tag{2.4.2}$$

où les θ_i seront simulés à partir de $\pi(\theta)$. Pour plus de détails concernant la simulation Monte-Carlo, voir la section 6.2.2 de Robert (2001).

Comme il est impossible de calculer $\pi(B|\mathcal{D})$ et que nous ne pouvons simuler directement des Σ_0 de $\pi(\Sigma_0|\mathcal{D})$, en se basant sur l'équation (2.4.2), nous avons

que

$$\begin{aligned}
E_\pi[B|\mathcal{D}] &= \int_{\mathcal{B}} B\pi(B|\mathcal{D})\, dB \\
&= \int_{\mathcal{B}}\int_{\mathcal{S}}\int_{\Sigma} B\pi(B|\sigma^2,\Sigma_0,Y)\pi(\sigma^2|\Sigma_0,Y)\pi(\Sigma_0|\mathcal{D})\, d\Sigma_0 d\sigma^2 dB \\
&= \int_{\mathcal{B}}\int_{\mathcal{S}}\int_{\Sigma} B\pi(B|\sigma^2,\Sigma_0,Y)\pi(\sigma^2|\Sigma_0,Y)\frac{m_2(Y|\Sigma_0)\pi(\Sigma_0)}{\int_\Sigma m_2(Y|\Sigma_0)\pi(\Sigma_0)\, d\Sigma_0}\, d\Sigma_0 d\sigma^2 dB \\
&= \frac{\int_{\mathcal{B}}\int_{\mathcal{S}}\int_{\Sigma} B\pi(B|\sigma^2,\Sigma_0,Y)\pi(\sigma^2|\Sigma_0,Y)m_2(Y|\Sigma_0)\pi(\Sigma_0)\, d\Sigma_0 d\sigma^2 dB}{\int_\Sigma m_2(Y|\Sigma_0)\pi(\Sigma_0)\, d\Sigma_0} \\
&\cong \frac{\sum_{i=1}^{M} B_{(i)} m_2(Y|\Sigma_{0(i)})}{\sum_{i=1}^{M} m_2(Y|\Sigma_{0(i)})},
\end{aligned}$$

où $m_2(Y|\Sigma_0)$ est la densité marginale de Y sachant Σ_0 et où l'on simule dans l'ordre : $\Sigma_{0(i)}$ à partir de $\pi(\Sigma_0)$, σ_i^2 à partir de $\pi(\sigma^2|\mathcal{D},\Sigma_{0(i)})$ et $B_{(i)}$ à partir de $\pi(B|\mathcal{D},\sigma_i^2,\Sigma_{0(i)})$.

De la même façon, nous obtenons que

$$E_\pi[\sigma^2|\mathcal{D}] \cong \frac{\sum_{i=1}^{m} \sigma_i^2 m_2(Y|\Sigma_{0(i)})}{\sum_{i=1}^{m} m_2(Y|\Sigma_{0(i)})}.$$

2.5. HYPERPARAMÈTRES

Le choix des hyperparamètres de notre modèle, soit les paramètres des distributions *a priori* est important pour pouvoir tester notre modèle. Comme nous détenons peu d'information *a priori* sur ceux-ci dans le cas présent, l'idéal serait que ces hyperparamètres soient choisis pour faire en sorte qu'ils soient le moins informatifs possible. Ainsi, nous utiliserons une méthode de Bayes empirique (Casella, 1985) pour spécifier nos hyperparamètres. Plus particulièrement, nous utiliserons la méthode du maximum de vraisemblance de type 2 (Ghosh, Delampady et Samanta, 2006). Cette méthode utilise l'information du jeu de données pour définir les hyperparamètres en sélectionnant ceux qui maximisent la vraisemblance marginale des Y.

Nous poserons pour l'ensemble des modèles que $B_0 \approx \hat{B}$. Plus précisément, nous posons B_0 égal à $\hat{B}$ arrondi à un chiffre significatif. Lorsque σ^2 n'est pas une variable aléatoire, nous poserons qu'il est égal à $\hat{\sigma}^2$. Lorsque nous supposons que σ^2 suit une inverse-gamma, on pose $a_0 = 2$ pour avoir une variance infinie et donc une distribution *a priori* non informative (il faut que $a_0 > 2$ pour qu'une inverse-gamma ait une variance finie puisque $Var(\sigma^2) = \frac{b_0^2}{(a_0-1)^2(a_0-2)}$) et $b_0 = (a_0 - 1)\hat{\sigma}^2$, car $E(\sigma^2) = b_0/(a_0 - 1)$.

Pour Σ_0, lorsqu'il est considéré fixe, nous nous baserons sur l'approche de la densité *a priori* de type G comme dans Merleau et al. (2007) et nous poserons qu'il est égal à $g(X^TX)^{-1}$ où $1/g$ représente l'importance que l'on accorde à l'information *a priori*. On posera donc que g est grand, ici, nous l'avons fixé à 10000. Lorsque Σ_0 suit une inverse-Wishart, plus ν_0 est petit, moins on accorde d'importance à l'information *a priori*, mais celui-ci doit être plus grand que $p-1$ (Nydick, 2012). On pose donc $\nu_0 = p$. Pour Λ_0, on pose $\Lambda_0 = g(X^TX)^{-1}$ encore une fois pour que l'espérance de Σ_0 soit près de $g(X^TX)^{-1}$.

Notons aussi que plus on est haut dans la hiérarchie d'un modèle, moins les hyperparamètres ont d'influence (DeGroot et Goel, 1981) et donc que les choix de ν_0 et Λ_0 sont moins importants.

2.6. RÉSULTATS

Ici, ce que nous voulons calculer est en fait un intervalle de crédibilité sur la courbe de régression. Pour ce faire, nous procéderons par simulation de courbes à partir de la distribution *a posteriori* des B. Nous utiliserons la technique de construction des intervalles de crédibilité simulés décrite à la section 2.3.

Dans toutes nos simulations, le nombre de $B_{(i)}$ simulés sera de 1000 afin que l'intervalle de crédibilité soit fiable. Nous nous sommes assurés que l'intervalle ne changeait pas de façon significative en augmentant la taille de la simulation

à plus de 1000. De plus, rappelons que nous gardons 95% des $B_{(i)}$ les plus probables *a posteriori*.

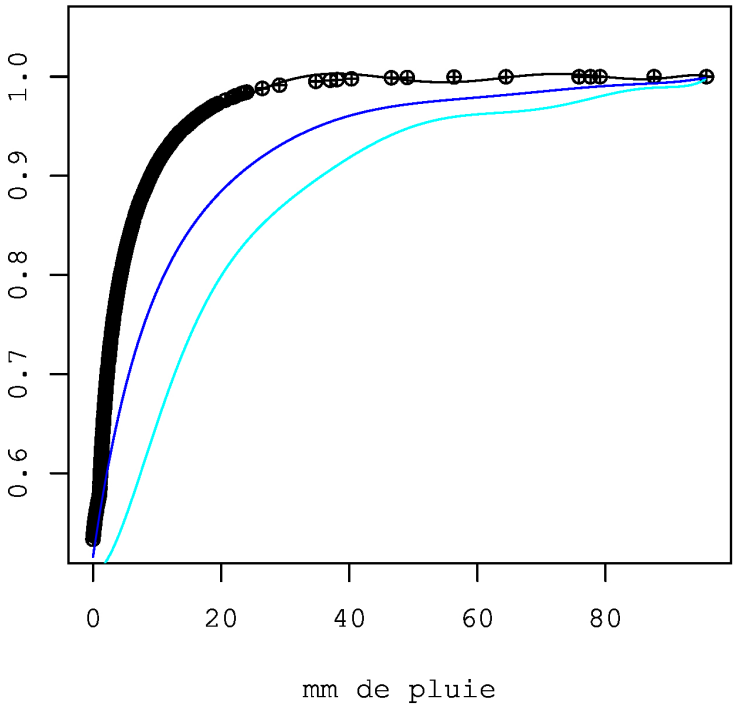
Nous utiliserons aussi le concept de fonction de répartition *a priori* et *a posteriori*. La fonction de répartition *a priori* est en fait la courbe correspondant à $X^T E_\pi[B]$ et celle *a posteriori* correspond à $X^T E_\pi[B|\mathcal{D}]$.

Nous présentons les intervalles simulés obtenus pour deux modèles : le modèle avec *a priori* de type G et le modèle avec inverse-gamma et inverse-Wishart. Dans le premier modèle, nous pourrons choisir 95% des $B_{(i)}$ les plus probables *a posteriori* directement puisque l'on connaît explicitement $\pi(B|\mathcal{D})$. De plus, nous pourrons comparer les intervalles simulés avec des intervalles de crédibilité simultanés obtenus en se basant sur les intervalles simultanés de Scheffé. Dans le deuxième modèle, nous ne connaissons pas explicitement $\pi(B|\mathcal{D})$, il faudra donc avoir recours à une modification de la méthode.

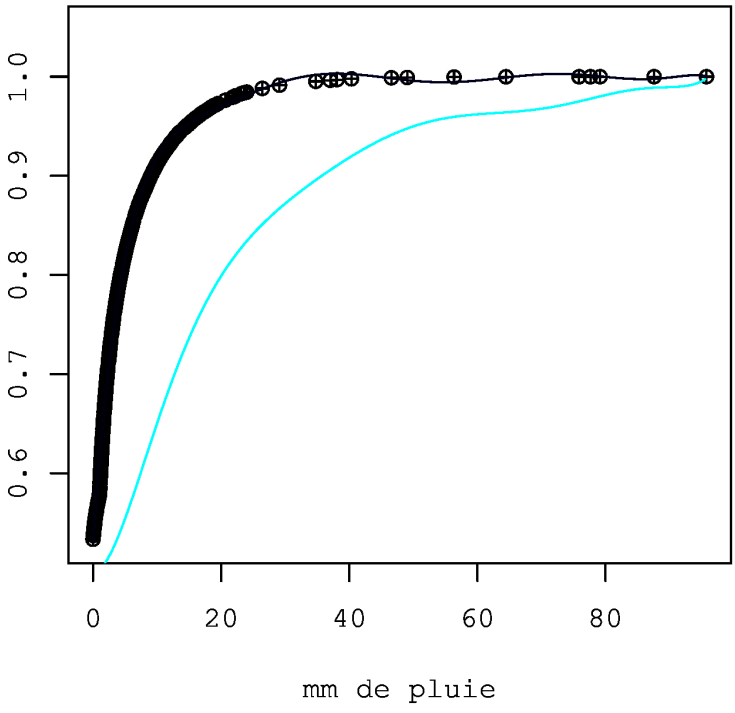
2.6.1. *A priori* de type G

Nous montrerons d'abord l'influence de la distribution *a priori* en choisissant un B_0 éloigné de $\hat{B}$. Pour cela, nous nous servirons du modèle avec un *a priori* de type G en posant deux valeurs différentes du paramètre g. Lorsque g prend la valeur 1, l'information *a priori* a exactement le même poids dans la distribution *a posteriori* que l'information provenant des données. Lorsque g est grand (ici $g = 10^4$), l'information *a priori* prend un poids presque nul dans la distribution *a posteriori*.

Dans la figure 2.1, nous voyons que lorsque $g = 1$, la fonction de répartition *a posteriori* est effectivement la moyenne entre la fonction de répartition de la régression et celle *a priori*. Lorsque g est grand, la fonction de répartition *a posteriori* correspond exactement à la fonction de répartition de la régression. Notons que dans les deux cas la fonction de répartition *a priori* est la même car g n'influence pas l'espérance de la distribution *a priori*.


0.6
0.7
0.8
0.9
1.0
0
20
40
60
80
mm de pluie

(a) $g = 1$.


0.6
0.7
0.8
0.9
1.0
0
20
40
60
80
mm de pluie

(b) $g = 10^4$.

Nous allons maintenant se baser sur les intervalles de confiance simultanés de Scheffé pour obtenir un intervalle de crédibilité simultané.

Les intervalles de Scheffé sont basés sur le fait que, lorsque l'on considère B comme fixe, $\hat{B} \sim N_p(B, \sigma^2(X^TX)^{-1})$ ce qui fait en sorte que $\frac{(\hat{B}-B)^T X^T X(\hat{B}-B)}{p\hat{\sigma}^2} \sim F_{p,n-p}$. Ici, nous considérons B comme aléatoire et considérons de plus que

$$B|\mathcal{D}, B_0, \sigma^2 \sim N_p\left(\frac{1}{g+1}(B_0 + g\hat{B}), \frac{g\sigma^2}{g+1}(X^TX)^{-1}\right).$$

De plus, de la même façon qu'avec (1.6.2), (1.5.1) devient dans ce contexte

$$\frac{\left(\frac{1}{g+1}(B_0 + g\hat{B}) - B\right)^T X^TX \left(\frac{1}{g+1}(B_0 + g\hat{B}) - B\right)}{\frac{g\hat{\sigma}^2}{g+1}} < \chi^2_p(\alpha).$$

Pour obtenir un intervalle de crédibilité simultané valide dans le modèle avec *a priori* de type G, il faut poser $B_0 = \hat{B}$, on aura que

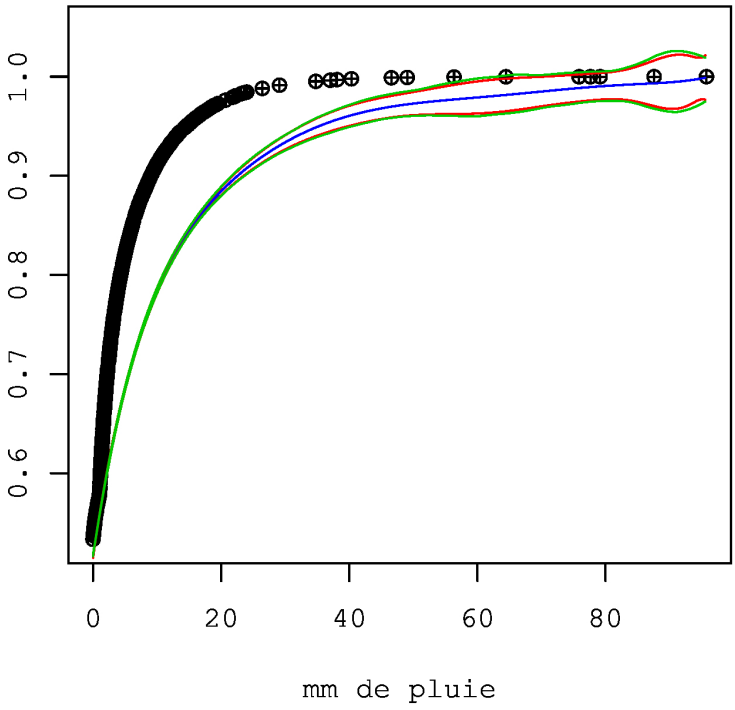
$$\hat{\sigma}^2 = \frac{1}{n-p}\sum_{i=1}^{n}\left(e_i - \frac{x_i^T(B_0 - \hat{B})}{g+1}\right)^2 = \frac{1}{n-p}\sum_{i=1}^{n} e_i^2,$$

où $e_i = y_i - x_i^T\hat{B}$. Cela nous permet d'utiliser une inégalité semblable à celle utilisée dans la méthode de Scheffé :

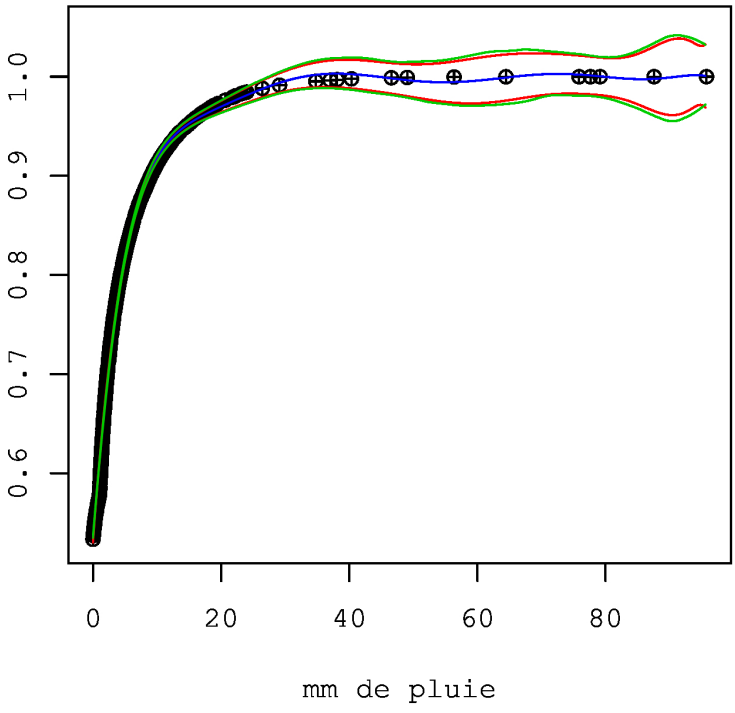
$$\begin{aligned}
1-\alpha &= Pr\left[\frac{|x^T\frac{1}{g+1}(B_0 + g\hat{B}) - x^TB|}{\hat{\sigma}(\frac{g}{g+1}x^T(X^TX)^{-1}x)^{1/2}} \leq \sqrt{\chi^2_p(\alpha)}, \forall x|\mathcal{D}\right] \\
&= Pr\left[\frac{|x^T\frac{1}{g+1}(\hat{B} + g\hat{B}) - x^TB|}{\hat{\sigma}(\frac{g}{g+1}x^T(X^TX)^{-1}x)^{1/2}} \leq \sqrt{\chi^2_p(\alpha)}, \forall x|\mathcal{D}\right] \\
&= Pr\left[\frac{|x^T\hat{B} - x^TB|}{\hat{\sigma}(\frac{g}{g+1}x^T(X^TX)^{-1}x)^{1/2}} \leq \sqrt{\chi^2_p(\alpha)}, \forall x|\mathcal{D}\right].
\end{aligned}$$

On peut donc obtenir un intervalle de crédibilité simultané théorique à 95% qui est :

$$x^T\left(\frac{1}{g+1}(B_0 + g\hat{B})\right) \pm \sqrt{\chi^2_p(0,05)}\hat{\sigma}\left(\frac{g}{g+1}x^T(X^TX)^{-1}x\right)^{1/2} \quad \forall x = (1, x_1, ..., x_{p-1}).$$


0.6
0.7
0.8
0.9
1.0
0
20
40
60
80
mm de pluie

(a) g = 1.


0.6
0.7
0.8
0.9
1.0
0
20
40
60
80
mm de pluie

(b) g = 10000.

Dans la figure 2.2, on compare les intervalles de crédibilité simultanés théoriques avec les intervalles de crédibilité simulés. La figure montre que les intervalles simulés correspondent bien aux intervalles théoriques.

2.6.2. Inverse-gamma et inverse-Wishart

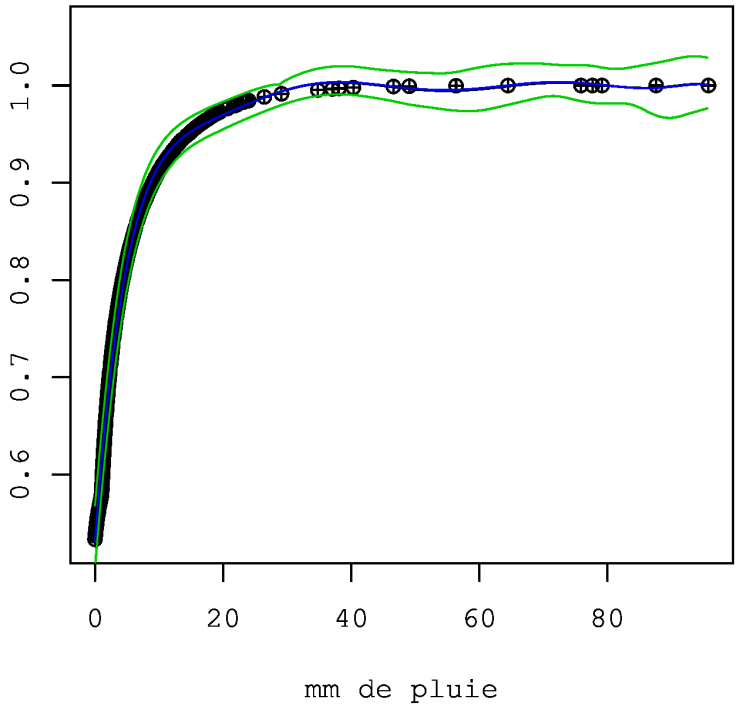
Nous passons maintenant à la simulation d'intervalles de crédibilité pour le modèle avec inverse-gamma et inverse-Wishart. Comme nous l'avons mentionné à la section portant sur les hyperparamètres, pour ces simulations, nous choisirons B_0 près de $\hat{B}$. Ici, comme il n'est plus possible de simuler directement à partir de $\pi(B|\mathcal{D})$, nous simulerons les $B_{(i)}$ par simulation Monte-Carlo comme décrit à la section 2.4. On simulera dans l'ordre : $\Sigma_{0(i)}$ à partir de $\pi(\Sigma_0)$, σ^2_i à partir de $\pi(\sigma^2|\Sigma_{0(i)}, Y)$ et $B_{(i)}$ à partir de $\pi(B|\sigma^2_i, \Sigma_{0(i)}, Y)$.

Pour obtenir un intervalle à 95%, nous pouvons plus ordonner les courbes à l'aide de la probabilité *a posteriori* de chaque $B_{(i)}$ puisque la distributon $\pi(B|\mathcal{D})$ n'est pas connue explicitement. Par contre nous connaissons à une constante près la distribution conjointe *a posteriori* de B, σ^2 et Σ_0 qui est

$$\begin{aligned}\pi(B, \sigma^2, \Sigma_0|\mathcal{D}) &= \pi(\Sigma_0|\mathcal{D})\pi(\sigma^2|\mathcal{D}, \Sigma_0)\pi(B|\mathcal{D}, \sigma^2, \Sigma_0)\\ &\propto h(\Sigma_0)\pi(\sigma^2|\mathcal{D}, \Sigma_0)\pi(B|\mathcal{D}, \sigma^2, \Sigma_0),\end{aligned}$$

où $h(\Sigma_0)$ est défini dans l'équation (2.2.1). Nous ordonnerons les courbes selon $\pi(B, \sigma^2, \Sigma_0|\mathcal{D})$ en retranchant 5% des courbes les moins probables. Nous obtiendrons ensuite l'enveloppe des courbes simulées qui correspondra à la borne supérieure et à la borne inférieure comme décrit à la section 2.3.

Aussi, notons que le fait que $\pi(B|\mathcal{D})$ ne soit pas connue explicitement fait en sorte que l'on ne peut pas se baser sur la méthode de Scheffé pour obtenir un intervalle de crédibilité théorique comme on l'a fait avec le modèle précédent. Nous ne pouvons donc pas comparer l'intervalle de crédibilité simulé obtenu avec un intervalle de crédibilité théorique.



(a) Intervalle de crédibilité simulé.

FIGURE 2.3. Intervalles de crédibilité simulé (vert) pour la fonction de répartition, fonction de répartition calculée par Monte-Carlo (bleu) et fonction de répartition de la régression (noir).

Nous pouvons voir dans la figure 2.3 l'intervalle de crédibilité simulé. Notons que la courbe calculée à l'aide de simulations Monte-Carlo est une approximation de la fonction de répartition *a posteriori* $X^T E_\pi[B|\mathcal{D}]$. Cette approximation est obtenue grâce à l'approximation de $E_\pi[B|\mathcal{D}]$, tel qu'expliqué à la fin de la section 2.4. Cette approximation se superpose parfaitement à la courbe de régression ce qui confirme que notre simulation est valide.

Remarquons toutefois que les intervalles de crédibilité pour la fonction de répartition dépassent encore 1 et cela entraîne qu'il est impossible d'obtenir des intervalles de crédibilité pour la période de retour. Cela est dû à l'absence de contrainte sur les coefficients de la régression comme au chapitre 1. Nous

concluons que la base polynomiale utilisée jusqu'à maintenant manque de flexibilité pour s'ajuster correctement au jeu de données. De plus, elle n'assure pas que la fonction simulée soit bornée entre 0 et 1 et monotone. Le prochain chapitre vise à trouver une base résolvant ces problèmes.

Chapitre 3

BASE NON LINÉAIRE

3.1. TRANSFORMATION ANGULAIRE

Pour nous assurer que les courbes simulées soient bornées entre 0 et 1 et pour stabiliser la variance, nous effectuerons une transformation angulaire de la fonction de répartition empirique. Soit a une valeur réelle positive, la transformation angulaire (Anscombe, 1948), ou transformation arc sinus racine carrée, est définie comme suit : $angular(a) = \arcsin\sqrt{a}$.

Cette transformation vise à stabiliser la variance et à réduire la covariance. En effet, rappelons que dans le chapitre 1 et 2, nous avons fait l'hypothèse que nos Y (variables dépendantes de la régression) étaient indépendants et identiquement distribués. Ceci n'est pas tout-à-fait exacte, comme nous en discuterons plus bas. Pour prendre en compte que les Y ne sont pas indépendants et identiquement distribués, nous utiliserons la transformation angulaire.

Le modèle général que nous avons adopté est le suivant :

$$F_n(x_i) = C(x_i) + e_i, i = 1, ..., n,$$

où $C(x_i)$ est une fonction quelconque et où on suppose que $F(x)$ fait partie des fonctions que peut estimer $C(x)$. Si on prend la fonction de répartition empirique non modifiée $F_n^*(x)$, on a par le théorème centrale limite que e_i converge en distribution vers une normale centrée en 0 et de variance $F(x_i)(1 - F(x_i))/n$

lorsque n tend vers l'infini (Doob, 1949). Donc avec $F_n(x)$, on a que e_i converge en distribution vers une normale centrée en 0 et de variance

$$\frac{n}{(n+2)^2}F(x_i)(1-F(x_i))$$

lorsque n tend vers l'infini. On sait aussi par le théorème de Donsker (Donsker, 1952) que

$$Cov(e_i, e_j) = \frac{n}{(n+2)^2}(min(F(x_i), F(x_j)) - F(x_i)F(x_j)).$$

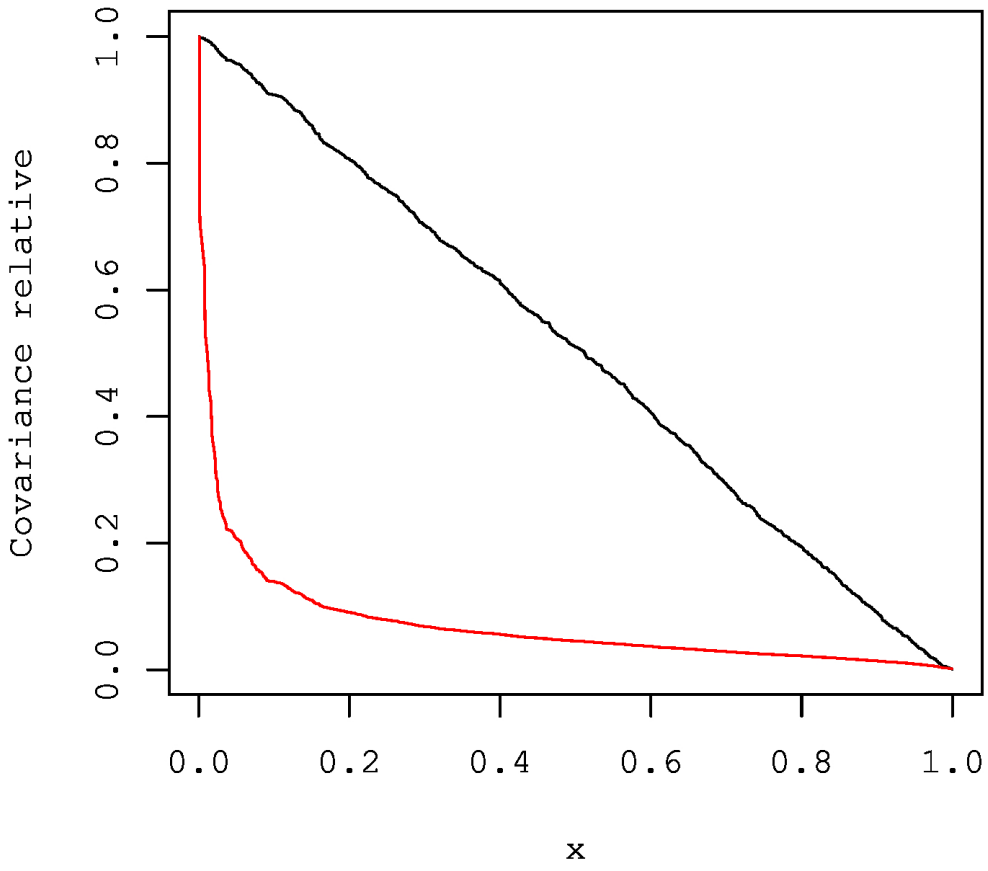
Donc, clairement les hypothèses d'indépendance et d'homoscédasticité ne sont pas respectées. Par contre, comme Anscombe (1948) le mentionne, la transformation angulaire est particulièrement appropriée pour stabiliser la variance d'une proportion calculée à partir d'une variable binomiale comme c'est ici le cas pour $F_n(x_i)$. En effet, on a que si $X \sim B(n, p)$, la variable $Y = \arcsin\sqrt{X/n}$ sera à peu près normal avec une variance égale à $1/4 + O(n^{-2})$. Donc, la variance sera effectivement stabilisée puisqu'elle ne sera plus dépendante de la proportion.

De plus, nous pouvons observer que la covariance est grandement diminuée par rapport à la variance. En effet, on peut approximer la covariance de la transformation angulaire de $F_n(x)$ avec la méthode delta (voir Klein, 1953) p.258 et l'annexe B pour les détails). On obtient que

$$Cov(angular(F_i), angular(F_j)) = \frac{n}{(n+2)^2}\frac{1}{4\sqrt{F_i(1-F_i)F_j(1-F_j)}}(min(F_i, F_j) - F_iF_j),$$

où $F_i = F_n(x_i)$. Dans la figure 3.1, on présente la première ligne de la matrice de covariance proportionnellement à la variance du premier élément de la matrice avant et après la transformation pour une fonction de répartition empirique d'un échantillon de taille 1000 provenant d'une uniforme $[0, 1]$. En d'autres mots, on présente la covariance relative de la première observation

avec les autres observations ordonnées en ordre croissant avant la transformation angulaire ($Cov(F_1, F_i)/Var(F_1)$) et cette même covariance relative après la transformation ($Cov(angular(F_1), angular(F_i))/Var(angular(F_1))$). On remarque que la covariance diminue rapidement vers 0 lorsque l'on s'éloigne de la diagonale et donc que la transformation angulaire nous aide à se rapprocher de l'hypothèse d'indépendance.



(a)

FIGURE 3.1. La rapport entre la covariance et la variance avant la transformation en noir et après la transformation en rouge de la fonction de répartition empirique d'une uniforme $[0, 1]$.

Pour étudier les effets de passer par une transformation angulaire lorsque l'on fait la régression, il faut étudier le comportement de la fonction inverse de notre transformation. En effet, nous voudrons faire la régression sur $angular(F_n(x))$, on aura donc :

$$angular(F_n(x_i)) = C(x_i) + e_i, i = 1, ..., n.$$

Pour avoir une courbe de régression dans l'espace de la fonction empirique, il faudra donc calculer $\sin^2(C(x))$. Il est clair que si $C(x)$ est bornée entre 0 et $\pi/2$, on a que l'estimation de $F_n(x)$ est bornée entre 0 et 1. Nous justifierons dans la section suivante le fait que $C(x)$ est bornée entre 0 et $\pi/2$.

De plus, si $C(x)$ est monotone, on aura que l'estimation de $F_n(x)$ l'est aussi puisque la fonction $\sin^2(x)$ est monotone entre 0 et $\pi/2$. Montrons simplement que la dérivée de $\sin^2(x)$ est positive entre 0 et $\pi/2$:

$$\frac{d\sin^2(x)}{dx} = 2\sin(x)\cos(x) \geq 0, \text{ si } x \in [0, \pi/2].$$

Il faut donc que $C(x)$ soit à la fois monotone et bornée entre 0 et $\pi/2$ pour que l'estimation de la fonction empirique ait les propriétés recherchées. Nous pourrions essayer d'utiliser une base polynomiale, mais celle-ci ne serait pas forcément monotone. En fait, les contraintes requises pour assurer la monotonicité seraient trop fortes et réduiraient encore la flexibilité de la base polynomiale. Nous utiliserons donc une base plus adaptée à notre problème : une base de splines monotones. Dans les prochaines sections, nous regarderons le problème d'estimer $angular(F(x))$ avec des splines.

3.2. SPLINES

Les principaux avantages de la base polynomiale sont certainement sa simplicité et le fait qu'elle engendre des fonctions continues. Par contre, un défaut important est qu'elle ne peut s'ajuster localement sans faire de changement

global. Comme la fonction est définie pour l'ensemble du domaine de la même façon, un changement dans une partie du domaine amène un changement sur l'ensemble du domaine. Pour avoir une base qui s'ajuste aux changements locaux, nous pouvons tout simplement séparer le domaine en morceaux et avoir un polynôme pour chaque morceau. Nous obtenons une fonction polynomiale définie par morceaux appelée spline.

Notons que le domaine des splines est très vaste et a de nombreuses applications qui ne seront pas traitées ici. Wegman et Wright (1983) font un bon survol de l'utilisation des splines en statistique. Pour notre besoin, nous traiterons seulement des splines polynomiales de régression telles que définies dans Ramsay (1988). En particulier, ce dernier expose dans son article la simplicité et la grande flexibilité des splines monotones et c'est cette approche que nous utiliserons.

Nous définissons une fonction spline sur un intervalle fermé $[L, U]$. Le domaine est divisé en sous-intervalles $[\xi_i, \xi_{i+1}[$ avec $L = \xi_1 < \xi_2 < ... < \xi_m = U$. La fonction est un polynôme $P_i : [\xi_i, \xi_{i+1}[\rightarrow \mathbb{R}$ d'un certain degré prédéfini $k-1$ dans chaque sous-intervalle. La continuité de la spline est assurée en posant des restrictions sur les dérivées des polynômes aux points $\{\xi_2, ..., \xi_{m-1}\}$. Pour chacun de ces points, nous posons $P_i^{(j)}(\xi_i) = P_{i+1}^{(j)}(\xi_i)$ pour $j = 0, ..., k-2$. Ce choix fait en sorte que la spline est continue, car pour tout i, au minimum $P_i(\xi_i) = P_{i+1}(\xi_i)$ et que partout ailleurs la spline est un polynôme. Aussi, c'est la contrainte de continuité la plus sévère que l'on puisse poser sans forcer que la spline ne soit définie que par un seul polynôme. En effet, si toutes les dérivées de deux polynômes de degrés $k-1$ sont les mêmes jusqu'aux dérivées d'ordre $k-1$, les polynômes sont égaux.

Pour la définition des algorithmes de calcul des splines, il est pratique de définir une suite de noeuds

$t = \{t_1, ..., t_{k+m+k}\}$ où

$$L = \xi_1 = t_1 = ... = t_k < t_{k+1} = \xi_2 < ... < t_{k+m} = \xi_{m-1} < t_{k+m+1} = ...$$
$$= t_{k+m+k} = \xi_m = U.$$

Donc les k premiers noeuds sont égaux à L, les k derniers sont égaux à U et ceux au centre correspondent aux points où la spline change de polynôme.

Nous voulons maintenant définir une nouvelle base de splines $M_j(\cdot|k, t)$ où $j = 1, ..., m + k$ de sorte que toutes splines f d'ordre k avec la suite de noeuds t puissent être représentées comme une combinaison linéaire de la base :

$$f = \sum_{j=1}^{m+k} a_j M_j.$$

La base que nous utiliserons est celle décrite par Curry et Schoenberg (1966) et aussi celle utilisée par Ramsay (1988) : la base M-spline. Cette base est définie de sorte que M_j est positive sur $[t_j, t_{j+k}[$ et nulle partout ailleurs avec la propriété qu'elle intègre à 1. Notons aussi qu'elle est équivalente à la base B-spline (voir Boor, 2001) qui est très utilisée. La meilleure façon de définir cette base est de façon récursive :

$$M_j(x|1, t) = \begin{cases} \dfrac{1}{t_{j+1} - t_j} & \text{si } t_j \leq x < t_{j+1} \\ 0 & \text{sinon} \end{cases}$$

$$M_j(x|k, t) = \begin{cases} \dfrac{k\left((x - t_j)M_j(x|k-1, t) + (t_{j+k} - x)M_{j+1}(x|k-1, t)\right)}{(k-1)(t_{j+k} - t_j)} & \\ & \text{si } t_{j+k} - t_j \neq 0 \\ 0 & \text{sinon.} \end{cases}$$

Comme chaque M_j est non nul seulement sur un certain intervalle, un changement d'un coefficient dans la combinaison linéaire n'affectera f que sur cet intervalle. La spline a donc la capacité recherchée de s'adapter localement. Aussi,

pour modéliser une fonction non négative, comme les M_j sont non négatifs nous n'avons qu'à poser la contrainte que tous les coefficients soient positifs. Finalement, pour avoir une fonction qui intègre à 1, la contrainte $\sum a_j = 1$ est suffisante.

3.3. SPLINES MONOTONES

Notre objectif étant d'obtenir une base monotone, nous pouvons donc utiliser la base M-spline. L'astuce pour obtenir une base qui engendrera des fonctions monotones est d'intégrer les M_j. Nous définissons la base I-spline, pour spline intégrée, de la façon suivante :

$$I_j(x|k,t) = \int_L^x M_j(u|k,t)du.$$

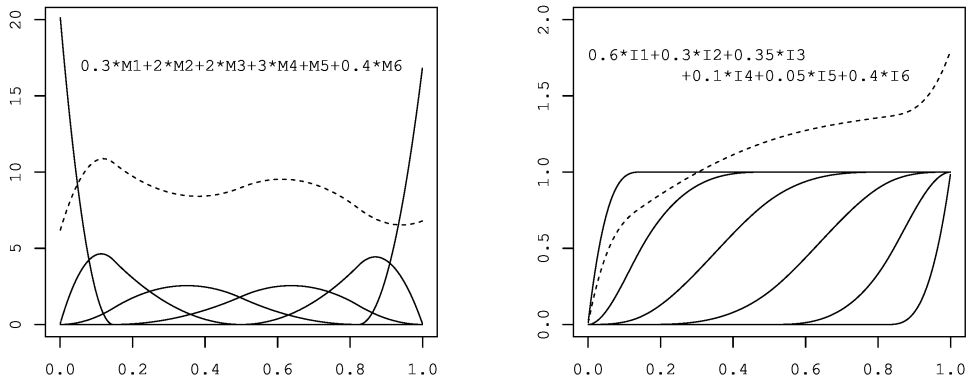
En effet, la positivité des M_j combinée avec la contrainte de non-négativité sur les coefficients fait en sorte qu'une combinaison linéaire des I_j sera bien monotone.

Ramsay (1988) nous donne un algorithme récursif plus pratique pour calculer les I_j :

$\forall x \in [t_i, t_{i+1}[$,

$$I_j(x|k,t) = \begin{cases} 0 & \text{si } j > i \\ 1 & \text{si } i-k > j \\ \sum_{m=j}^{i} (t_{m+k+1} - t_m) M_m(x|k+1,t)/(k+1) & \text{sinon.} \end{cases}$$

Dans la figure 3.2, les bases M-spline et I-spline d'ordre 3 avec trois noeuds intérieurs sont exposées. Nous voyons les $m+k=6$ composantes de chacune des bases de gauche à droite et une combinaison linéaire pour chacune. Les propriétés dont nous avons parlées précédemment (positivité, intégration à 1 et monotonicité) sont bien respectées.



(a) $M-spline$ (b) $I-spline$

FIGURE 3.2. La base M-spline et I-spline d'ordre 3 avec trois noeuds intérieurs ($k = 3, m = 3$) ainsi qu'une combinaison linéaire pour chacune des bases.

La condition que $C(x)$ doit être bornée entre 0 et $\pi/2$ n'est pas forcément respectée. Par contre, le fait que l'on simule des courbes qui sont monotones combiné avec le fait que le maximum des valeurs que les courbes estiment est inférieur à $\pi/2$ fait en sorte que nos courbes seront presque assurément entre 0 et $\pi/2$. De plus, on peut imposer une contrainte supplémentaire sur le domaine des $B_{(i)}$ simulés en rejetant les $B_{(i)}$ qui font en sorte que notre courbe dépasse $\pi/2$. En pratique ceci n'est pas nécessaire et les simulations viendront confirmer que nos courbes sont bien entre 0 et $\pi/2$. En fait, dans les simulations plus bas, nous avons vérifié si les courbes simulées dépassaient $\pi/2$ et cela n'est jamais arrivé.

Pour la suite du chapitre, la base I-spline sera utilisée.

3.4. CHOIX DU DEGRÉ DE LA BASE, DE L'EMPLACEMENT ET DU NOMBRE DES NOEUDS

Pour avoir une spline qui s'ajuste bien à nos données, trois facteurs rentrent en ligne de compte : le degré de la spline, l'emplacement des noeuds intérieurs

et leur nombre. Deux points sont à considérer pour la sélection des noeuds. Premièrement, plus il y a de noeuds, plus la spline sera flexible. Deuxièmement, plus il y a d'observations entre chaque noeud, plus la spline sera bien définie.

Dans notre application, nous avons un grand nombre de données réparties de façon non uniforme sur l'ensemble du domaine. Tel que le mentionne Ramsay (1988), bien qu'à l'endroit d'un noeud la spline ait plus de liberté pour s'ajuster aux données, la forme de la spline n'est pas très sensible aux choix de l'emplacement des noeuds. Par contre, ici nous nous intéressons particulièrement aux données extrêmes puisque le but ultime est de construire un intervalle de confiance pour la période de retour qui est intéressante surtout pour ces dernières. Pour s'assurer que les extrêmes sont bien ajustés par la spline, nous placerons des noeuds au troisième point le plus petit et au troisième point le plus grand. Ensuite, comme nous voulons avoir une spline qui s'ajuste bien sur l'ensemble du domaine, nous choisirons donc d'espacer uniformément les noeuds entre ces derniers. Hastie et Tibshirani (1990) recommandent aussi cette approche.

Pour ce qui est du nombre de noeuds, la plupart des applications n'en requièrent pas une grande quantité. En effet, Ramsay (1988) note qu'à moins d'avoir besoin d'une très grande flexibilité, quelques noeuds suffisent. Par contre, après quelques simulations, nous avons remarqué que comme de légers changements dans les extrêmes influençaient fortement la forme de la période de retour, il était primordial que l'ajustement soit presque parfait et donc qu'il y ait suffisamment de noeuds. Nous utilisons donc une approche similiaire à celle de He et Shi (1998) où l'on prend le nombre de noeuds qui minimise l'AIC (critère d'information d'Akaike) pour la régression sur les données

d'intérêt. Le nombre de noeuds variera donc en fonction des données utilisées pour la régression. Dans notre application et nos simulations, nous avons trouvé qu'entre 6 et 12 noeuds étaient généralement suffisants.

Il nous reste donc à déterminer le degré de la base qui sera utilisé. Pour ce faire, nous regarderons le R^2 ajusté de la spline obtenue pour estimer $angular(F(x))$ pour des bases de petits degrés ($k = 3, 4, 5$), car nous n'avons pas rencontré dans la littérature d'utilisation de splines de plus haut degré. Les calculs sont faits sur l'échantillon des données de précipitations utilisé au chapitre 1 et 2.

TABLEAU 3.1. R^2 ajusté pour différents degrés de la base I-spline.

k	3	4	5
R^2 ajusté	0.9983	0.9986	0.9987

Dans le tableau 3.1, nous remarquons que l'augmentation du degré résulte en un gain très faible du R^2 ajusté et que la base d'ordre 3 ($k = 3$) est donc satisfaisante. Notons aussi qu'il est généralement reconnu que l'oeil humain ne peut pas distinguer des discontinuités d'ordre plus grande que 2. L'intérêt d'une base d'ordre plus grande que 3 semble donc limité puisque celle-ci offre déjà des dérivées première et seconde continues (Friedman, Hastie et Tibshirani, 2009).

Rappelons que X est la matrice des données de départ. Pour la suite, nous noterons la transformation de cette matrice vers la base des splines monotones d'ordre 3 S_X, c'est-à-dire qu'au lieu d'avoir $Y = XB$ comme régression, nous aurons $Y = S_X B$.

3.5. RÉGRESSION AVEC CONTRAINTES

Comme nous devons avoir des coefficients positifs avec la base des I-splines pour avoir une fonction monotone, nous ne pouvons utiliser la technique de régression des moindres carrés habituelle. Le package R mgcv

contient la fonction pcls (penalized constrained least-squares fitting) qui implémente l'algorithme décrit dans Gill, Murray et Wright (1981). Cette fonction résout des problèmes de moindres carrés avec des contraintes d'inégalité linéaire (et de pénalité si nécessaire) en utilisant une méthode de résolution provenant de l'optimisation quadratique. Nos contraintes ici sont bien linéaires et le problème des moindres carrés peut se voir comme un problème de minimisation d'une fonction quadratique qui est exactement le but de l'optimisation quadratique.

Nous allons donc pouvoir faire la régression suivante avec la fonction pcls :

$$angular(\hat{F}(x)) = \sum_{j=1}^{m+k} B_j I_j(x|k,t) = BS_X, \text{ avec } B_j \geq 0,$$

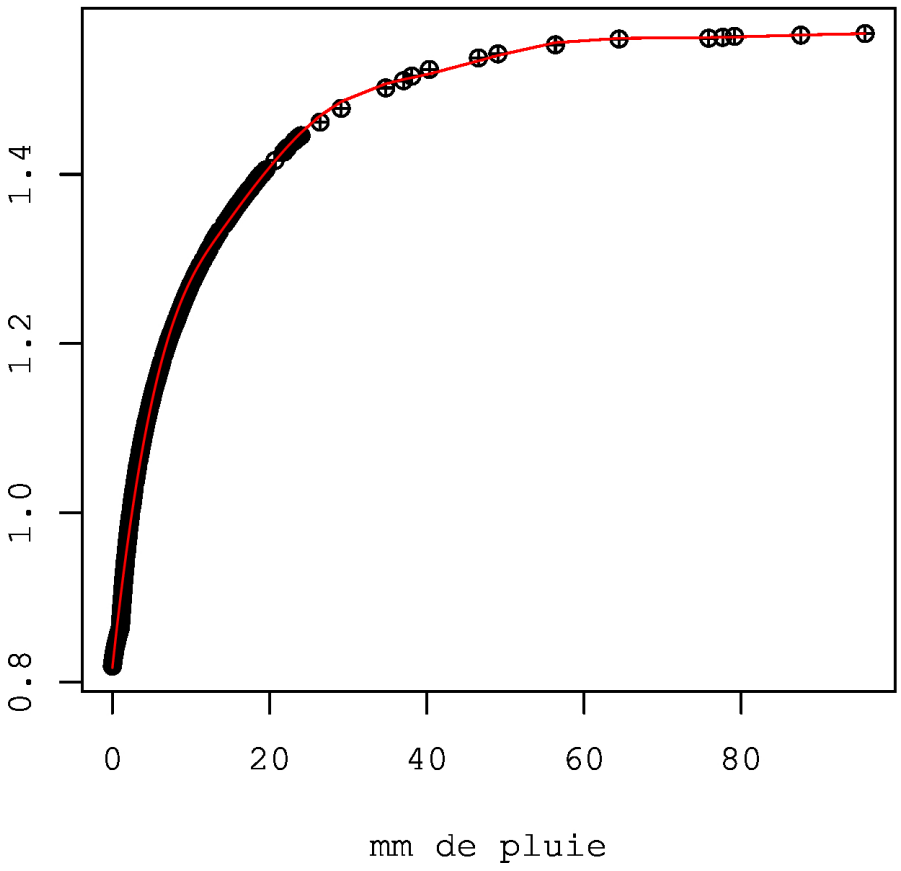
et obtenir une fonction monotone estimant $F(x)$ avec précision en appliquant l'inverse de la transformation angulaire. Dans la figure 3.3, nous voyons que l'ajustement aux données est satisfaisant et que la courbe obtenue est bien monotone.

La contrainte de positivité sur les coefficients fait en sorte que nous ne pouvons plus utiliser une simple normale multivariée comme *a priori*, car celle-ci supposerait que les coefficients peuvent être négatifs. Une approche consiste à utiliser une normale multivariée tronquée. La densité d'une normale multivariée tronquée a la forme suivante :

$$f_{X_1,...,X_p}(x_1,...,x_p) \propto exp\left(-0,5(x-\mu)^T\Sigma^{-1}(x-\mu)\right) I(x \in P),$$

où P est l'espace des x positifs dans notre cas. Nous noterons une variable qui suit une telle distribution $X \sim NP_p(\mu, \Sigma)$.

Nous utiliserons une méthode de rejet due à von Neumann (1951) pour simuler la normale (voir Devroye, 1986). Cette méthode est utilisée lorsqu'on ne peut pas directement simuler à partir d'une densité, mais qu'il est possible de simuler à partir d'une densité similaire. En fait, ici, elle consiste simplement



(a) $angular(\hat{F}(x))$

FIGURE 3.3. La courbe de régression obtenue avec la base de splines monotones.

à simuler à partir d'une normale multivariée et de rejeter l'échantillon simulé tant qu'il se trouve dans la zone tronquée. Cette méthode fonctionne bien si le taux d'acceptation est assez élevé. Dans notre cas, la variance est assez faible sur les coefficients de sorte que le taux d'acceptation est très élevé et cette méthode est donc suffisante. D'autres techniques ont été développées dans Robert (1995) et Damien et Walker (2001) lorsque le taux d'acceptation est trop faible, mais elles n'ont pas été nécessaires ici.

Le lecteur intéressé à simuler des normales tronquées notera qu'il existe un package R qui implémente des algorithmes de simulation très efficaces : tmvtnorm (Wilhelm et Manjunath, 2010).

Pour l'ensemble des résultats et simulations qui suivent, nous utiliserons le modèle le plus complexe développé au chapitre 2, soit le modèle inverse-gamma et inverse-Wishart. Pour les détails sur la distribution *a posteriori* des coefficients, nous ramenons le lecteur à la section 2.2 où nous avons développé la structure des distributions *a priori* que nous posons et les *a posteriori* résultantes ainsi que les hyperparamètres utilisés. Ici, le seul changement est la distribution *a priori* sur les coefficients qui sera

$$B|\sigma^2, \Sigma \sim NP_p(B_0, \sigma^2\Sigma).$$

La distribution *a posteriori* qui découle de cette *a priori* est aussi une normale tronquée. Les calculs sont très similaires à ceux réalisés au chapitre 2 :

$$\begin{aligned}\pi(B|\mathcal{D}, \sigma^2, \Sigma_0) &\propto \pi(B|\sigma^2, \Sigma_0)l(B|\mathcal{D}) \\ &\propto \exp\left(-\frac{1}{2\sigma^2}(B - B_0)^T\Sigma^{-1}(B - B_0)\right) I(B \in P)l(B|\mathcal{D}) \\ &\propto \exp\left(-\frac{1}{2\sigma^2}(B - A)^T(S_X^T S_X + \Sigma^{-1})(B - A)\right) I(B \in P),\end{aligned}$$

où $A = (S_X^T S_X + \Sigma_0^{-1})^{-1}(\Sigma_0^{-1}B_0 + S_X^T Y)$. Donc, la distribution *a posteriori* est

$$B|\mathcal{D}, \sigma^2, \Sigma_0 \sim NP_p(A, \sigma^2(S_X^T S_X + \Sigma_0^{-1})^{-1}).$$

Les autres distributions restent inchangées et nous procéderons de la même façon qu'au chapitre 2 pour simuler les courbes.

3.6. SIMULATIONS

Nous allons maintenant analyser les propriétés fréquentistes de notre intervalle de crédibilité simulé (tel que décrit à la section 2.3) pour le comparer

avec la méthode de Scheffé. Plus précisément, nous allons produire notre intervalle de crédibilité et calculer la probabilité de couverture de cet intervalle si on considère que la courbe est fixe. Ceci viendra valider que notre approche par simulation produit un intervalle simultané valide. Nous procéderons en simulant des données à partir d'une fonction de répartition connue. Nous utiliserons la base décrite au présent chapitre (transformation angulaire avec splines monotones) pour estimer la fonction de répartition empirique. De plus, nous utiliserons le modèle bayésien inverse-gamma et inverse-Wishart.

Nous voulons donc nous assurer que les intervalles de crédibilité obtenus ont bien un niveau de 95%. Pour ce faire, nous simulerons d'abord des données à partir d'une uniforme sur l'intervalle $[0, 1]$. Cette distribution a été choisie car, à partir d'une uniforme, on peut obtenir une grande variété de distributions différentes en utilisant des transformations bijectives (Devroye, 1986). Nous ferons trois simulations où la taille échantillonale sera respectivement, 100, 500 et 1000. Pour chaque taille échantillonnale nous simulerons 500 échantillons et nous calculerons l'intervalle de crédibilité simultané de niveau 95% en posant le nombre d'itérations de la simulation Monte Carlo à 1000. Ici, le calcul pour l'intervalle de crédibilité simulé est le même que celui décrit à la section 2.6.2 sauf que l'on utilise la transformation angulaire et les splines monotones ce qui nous assure que la courbe est bornée entre 0 et 1 et qu'elle est monotone. Nous vérifierons ensuite le nombre de fois où la vraie fonction de répartition se trouve dans l'intervalle de crédibilité, c'est-à-dire que l'on vérifie pour chacun des points de notre échantillon si la fonction de répartition tombe à l'intérieur de l'intervalle, pour ainsi obtenir une estimation de la probabilité de couverture. Nous comparerons la probabilité de couverture des intervalles simulés avec la probabilité de couverture des intervalles de confiance de Kolmogorov borné.

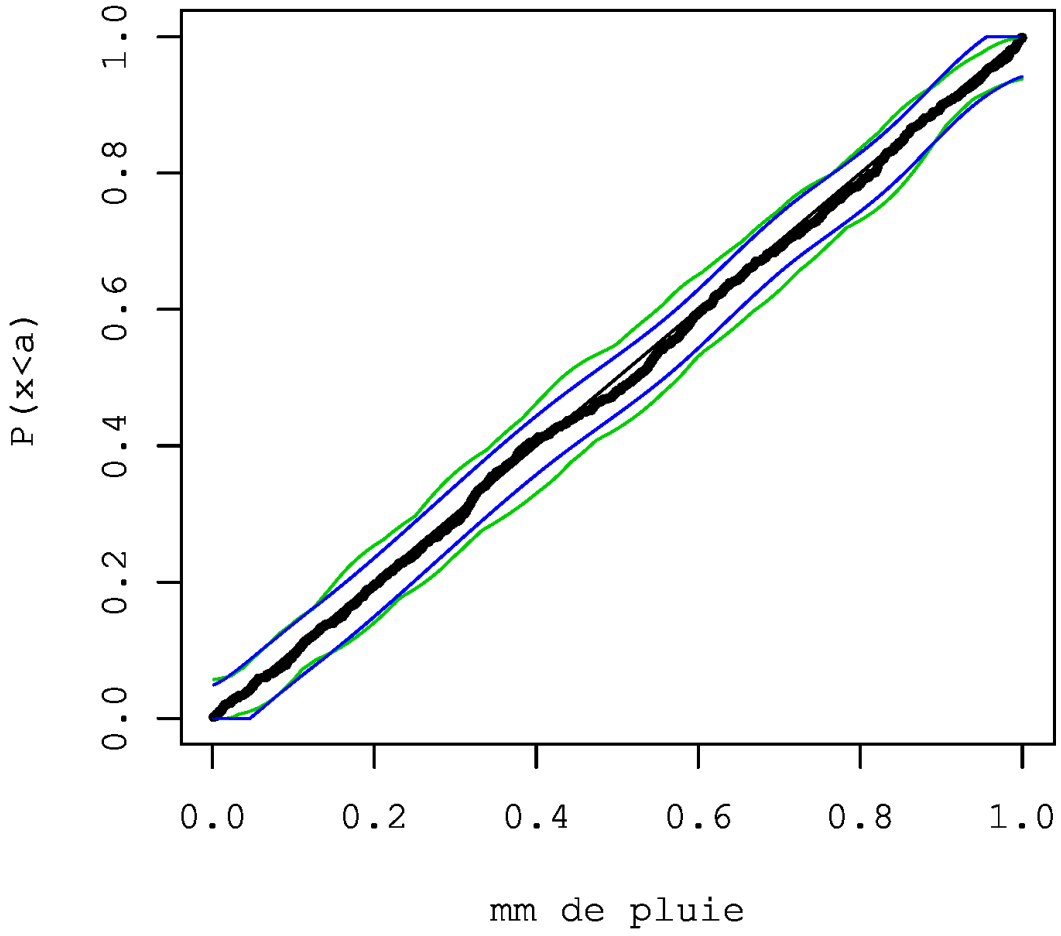


FIGURE 3.4. La fonction de répartition empirique pour un échantillon de taille 1000 provenant d'une uniforme. La vraie fonction de répartition (noir), l'intervalle de crédibilité simulé (vert) et l'intervalle de confiance de Kolmogorov borné (bleu).

Dans la figure 3.4, on présente une des simulations pour un échantillon de taille 1000. On remarque que la vraie fonction de répartition se trouve en effet dans les deux intervalles (ceci devrait se produire en théorie 95 fois sur 100). Aussi, notons que l'intervalle de Kolmogorov est moins conservateur au centre, mais légèrement plus conservateur dans les extrémités que l'intervalle simulé.

TABLEAU 3.2. Probabilité de couverture estimée par simulation

n	$1-\alpha$ simulé	$1-\alpha$ Kolmogorov borné
100	0.994	0.982
500	0.982	0.980
1000	0.966	0.979

Dans le tableau 3.2 sont présentés les résultats des simulations. On peut voir que les intervalles de confiance de Kolmogorov sont trop conservateurs même avec un taille échantillonnale de 1000 ce qui rejoint en partie les résultats de Wang et al. (2013). Par contre, bien que les intervalles simulés soient aussi conservateurs, ils convergent plus rapidement vers 95% et notre approche semble donc appropriée. Nous pouvons donc appliquer notre méthode à nos données de précipitation.

3.7. RÉSULTATS

Dans les figures 3.5 et 3.6, nous présentons les intervalles de crédibilité basés sur les courbes simulées et les intervalles de confiance de Kolmogorov bornés. Ici, les problèmes rencontrés au chapitre 2 ne sont plus présents. Comme les courbes simulées sont monotones et que nous utilisons la transformation angulaire, nous simulons des fonctions de répartition valides (monotone et bornée entre 0 et 1). Il est donc possible d'obtenir un intervalle de crédibilité simulé pour la période de retour. De plus, cet intervalle est conforme à notre intuition : plus un évènement est extrême, moins on est certain de sa période de retour. Notons aussi que même si l'estimation de la fonction de répartition semble être égale à 1 à partir de 40 mm de pluie, c'est en fait les légères différences avec 1 qui permettent d'obtenir une estimation valide de la période de retour. Comme nous l'avons mentionné au chapitre 1, un évènement deux fois plus rare qu'un autre aura une valeur dans la fonction de répartition deux fois plus près de 1. C'est donc de toutes petites différences sur l'estimation de la fonction de répartition qui auront de grands impacts

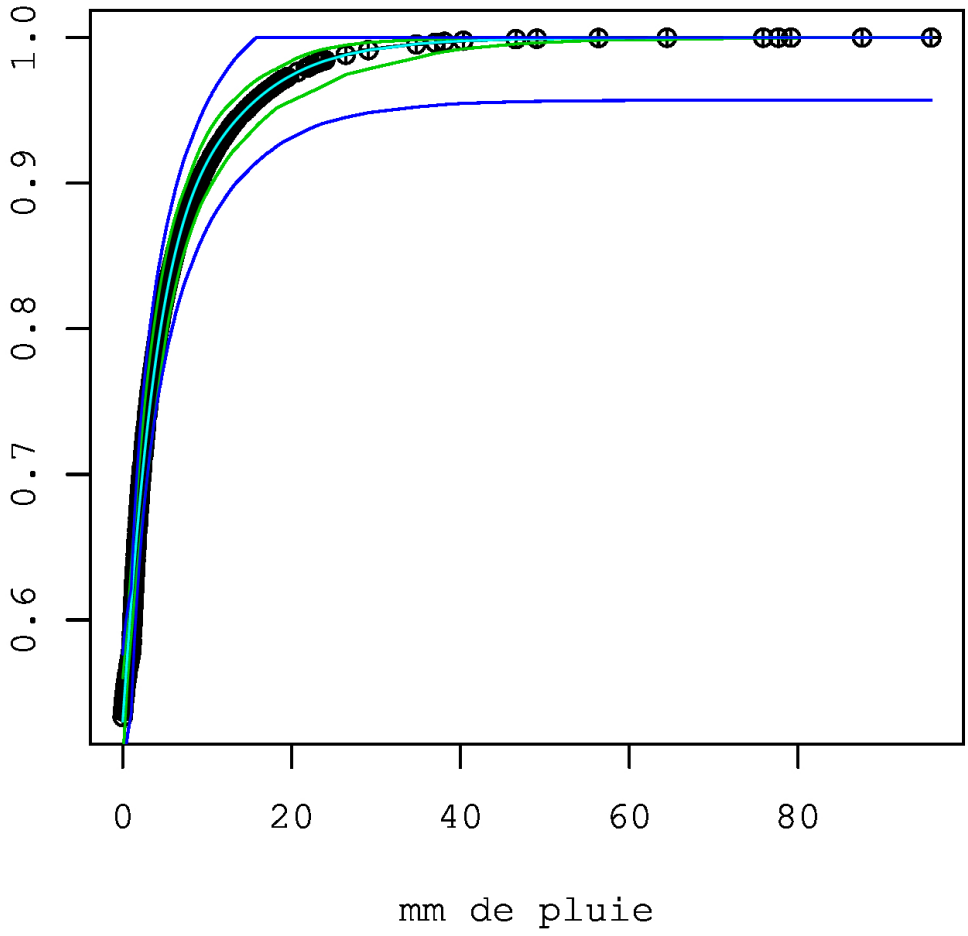
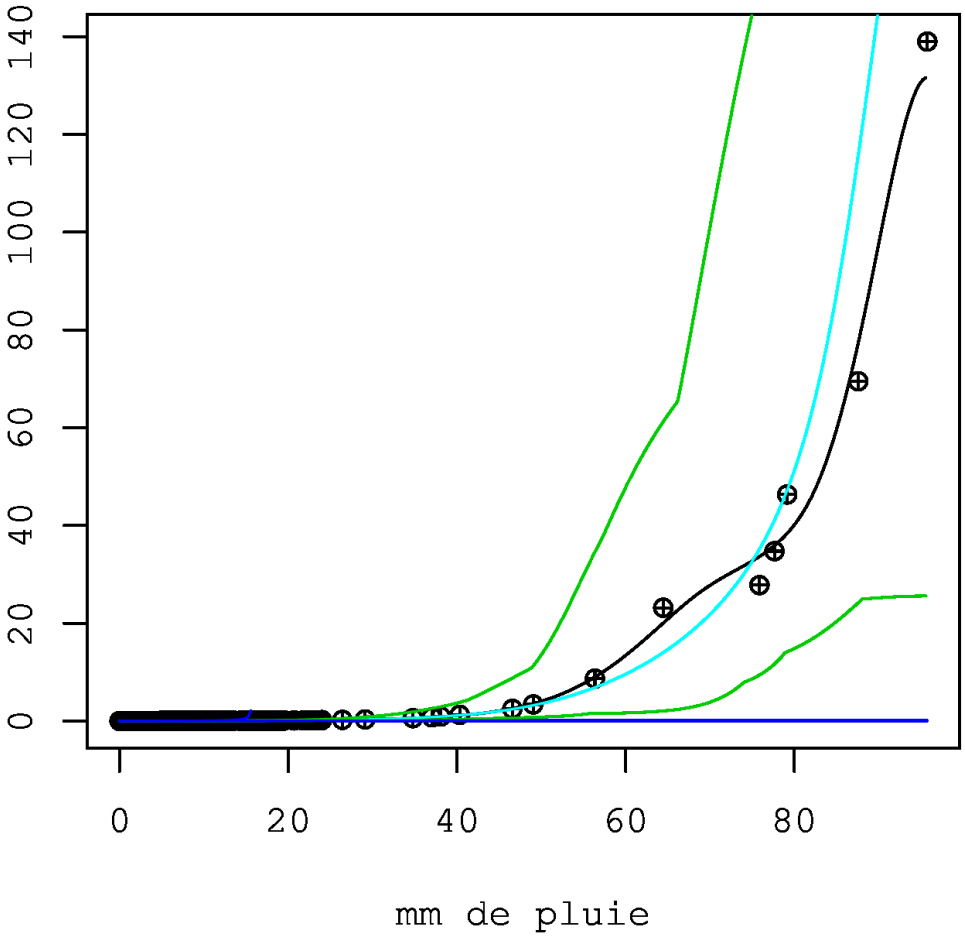


FIGURE 3.5. L'intervalle de crédibilité simulé (vert) et l'intervalle de confiance de Kolmogorov borné (bleu) ainsi que la courbe de régression (noir) et son estimation par Monte-Carlo (bleu pâle) pour la fonction de répartition empirique. La courbe de régression n'est pas visible puisque l'estimation s'y superpose presque parfaitement.

sur la période de retour. Aussi, cela fait en sorte qu'il y a beaucoup d'incertitude pour la période de retour à partir de 40 mm de pluie et cela se traduit par un intervalle de confiance très large. Remarquons aussi que l'intervalle de confiance de Kolmogorov borné est inutile pour la période de retour, car il est beaucoup trop conservateur quand la fonction de répartition empirique



(a) Période de retour

FIGURE 3.6. L'intervalle de crédibilité simulé (vert) et l'intervalle de confiance de Kolmogorov borné (bleu) ainsi que la courbe de régression (noir) et son estimation par Monte-Carlo (bleu pâle) pour la période de retour.

est près de 1. En effet, on peut remarquer qu'à partir de 16 mm de pluie, la borne supérieure est égale à 1 et ne permet donc plus de faire une transformation vers la période de retour puisque cela engendre une division par 0 (rappel : $PdR_n(a) = 1/(365(1 - F_n(a)))$).

De plus, nous n'avons plus les propriétés asymptotiques de B puisque nous n'utilisons plus l'estimateur des moindres carrés, ces propriétés sont complexes à obtenir et sont en partie traitées par Judge et Takayama (1966).

L'avantage de notre approche est que la probabilité de couverture obtenu par simulation devrait être valide même si la zone de troncation est importante, c'est-à-dire, même si les propriétés asymptotiques ne sont pas disponibles pour l'estimateur contrairement à l'approche classique. En effet, dans le cadre où nous devons estimer des fonctions monotones, il n'est plus possible d'utiliser l'estimateur des moindres carrés. Plus l'estimateur classique se trouve près de la zone de troncation, plus l'estimateur obtenu sera loin de l'estimateur classique et donc le niveau de confiance des intervalles calculés à partir des propriétés de l'estimateur classique ne sera plus valide. De plus, comme noté au chapitre 2, lorsque nous n'avons pas la distribution des B *a posteriori*, comme c'est ici le cas, il n'est plus possible de s'inspirer de la méthode de Scheffé comme nous l'avons fait à la section 2.6.1.

De plus, nous observons que l'intervalle de confiance de Kolmogorov borné ne peut être utilisé pour obtenir un intervalle de confiance pour la période de retour alors que notre méthode permet d'en obtenir un.

CONCLUSION

Le but de ce mémoire était de s'intéresser aux intervalles de confiance simultanés dans le cadre bayésien. Nous nous sommes plus particulièrement intéressés à la simulation d'intervalles de crédibilité simultanés pour la période de retour, une fonction non linéaire basée sur des données de précipitations.

Dans le chapitre 1, nous avons d'abord présenté les données utilisées, la période de retour et la fonction de répartition empirique. Nous avons ensuite fait un bref rappel du cadre classique de la régression linéaire et des différentes propriétés des estimateurs utilisés, puis nous avons présenté trois façons d'obtenir des intervalles de confiance simultanés : avec la méthode de Scheffé, la méthode de Kolmogorov et par simulation. En utilisant la base polynomiale pour approximer la fonction de répartition empirique, nous avons simulé des intervalles de confiance et les avons comparés à ceux obtenus avec les deux autres méthodes. Il y avait alors des problèmes importants puisque la base polynomiale ne forçait pas les courbes simulées à respecter les propriétés d'une fonction de répartition, c'est-à-dire que la courbe soit croissante et bornée entre 0 et 1. De plus, les hypothèses d'indépendance et d'homoscédasticité n'étaient pas respectées.

Au chapitre 2, nous avons introduit l'approche bayésienne et vu différents modèles de distributions *a priori* pour les paramètres de la régression linéaire. Nous avons, entre autres, traité des *a priori* de type G et d'un modèle plus complexe basé sur des *a priori* inverse-gamma et inverse-Wishart. Dans ce dernier

cas, il était impossible d'obtenir la distribution *a posteriori* des coefficients de la régression linéaire et nous avons dû faire appel à la méthode de simulation Monte-Carlo pour simuler des intervalles de crédibilité.

Finalement, au chapitre 3, nous avons trouvé une solution aux limitations de la base polynomiale en utilisant une base non linéaire. Cette base était fondée sur la transformation angulaire et les splines monotones que nous avons donc survolées. Aussi, il a été nécessaire d'utiliser la régression avec contraintes et un *a priori* pour coefficients avec contrainte afin de s'assurer que les coefficients des courbes simulées étaient positifs, ce dont nous avions besoin pour que les courbes ne soient pas décroissantes.

Nous rappelons que l'obtention d'intervalles de confiance simultanés par simulation est justifiée lorsqu'il est impossible de dériver des intervalles de confiance simultanés théoriques comme dans le cas d'une fonction non linéaire ou encore, dès que la distribution *a posteriori* est trop complexe. Aussi, notre approche est utile surtout lorsqu'il faut déjà faire appel à la simulation Monte-Carlo pour calculer une courbe de prédiction. Dans ce cas, nous pouvons nous servir des courbes obtenues lors de la simulation pour obtenir directement un intervalle de crédibilité simultané.

Aussi, le travail fait sur la base non linéaire est intéressant puisqu'il permet d'approximer des fonctions de répartition avec une bonne fiabilité. Certaines bases de fonctions monotones autres que les splines pourraient être explorées, comme les ondelettes. Pour un exemple d'ondelettes monotones, voir Angers et MacGibbon (2013).

Dans ce mémoire, nous avons seulement parlé de la construction d'intervalles de confiance. Une fois notre méthode développée, il pourrait être intéressant de l'utiliser pour conduire des tests d'hypothèses. On pourrait, par

exemple, construire un intervalle de confiance pour une période de retour basée sur les données de précipitations avant 1980 et vérifier si la période de retour des précipitations après 2000 coïncide avec cet intervalle.

Nous terminons en rappelant que la validité de notre méthode repose sur certaines hypothèses. Nous avons fait l'hypothèse que les courbes à estimer pouvaient être décrites par un nombre fini de paramètres, et que ces paramètres suivaient certaines distributions de probabilités. Ici, la stabilisation de la variance qu'apporte la transformation angulaire semble être suffisante pour se rapprocher des hypothèses.

Bibliographie

Allen, D. M. (1974). The relationship between variable selection and data augmentation and a method for prediction. *Technometrics*, **16**, 125–127.

Angers, J.-F. et MacGibbon, B. (2013), Hazard function estimation with nonnegative "wavelets", *Statistics & Probability Letters*, **83**, 969–978.

Anscombe, F. J. (1948) The transformation of Poisson, binomial and negative-binomial data. *Biometrika*, **35**, 246–254.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag, NewYork.

Boor, C. (2001). *A Practical Guide to Splines (revised edition)*. Springer-Verlag, NewYork.

Casella, G. (1985). An introduction to empirical Bayes data analysis. *American Statistician*, **39**, 83–87

Curry, H. B. et Schoenberg, I. J. (1966), On Polya frequency functions IV : The fundamental spline functions and their limits. *Journal d'analyse mathématique*, **17**, 71–107.

Damien, P. et Walker, S. G. (2001). Sampling truncated normal, beta, and gamma densities. *Journal of Computational and Graphical Statistics*, **10**, 206–215.

DeGroot, M. H. et Goel, P. K. (1981). Information about hyperparamters in hierarchical models. *Journal of the American Statistical Association*, **76**, 140–147.

Devroye, L. (1986). *Non-uniform Random Variate Generation*. Springer-Verlag, New York.

Donsker, M. D. (1952). Justification and extension of Doob's heuristic approach to the Kolmogorov-Smirnov theorems. *Annals of Mathematical Statistics*, **23**, 277–281.

Doob, J. L. (1949). Heuristic approach to the Kolmogorov-Smirnov theorems. *Annals of Mathematical Statistics*, **20**, 393–403.

Eagleson, P. S. (1978). Climate, soil, and vegetation : 2. The distribution of annual precipitation derived from observed storm sequences, *Water Resources Research*, **14**, 713–721.

Fan, J. et Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, **96**, 1348–1360.

Friedman, J., Hastie, T. et Tibshirani, R. (2009). *The Elements of Statistical Learning : Data Mining, Inference, and Prediction*. 2e édition. Springer-Verlag, New York.

Geinitz, S. (2009). Prior covariance choices and the g prior. *Seminar on Bayesian Linear Model*, University of Zurich.

Gill, P. E., Murray, W. et Wright, M. H. (1981). *Practical Optimization*, Academic Press, New York.

Ghosh, J. K., Delampady, M. et Samanta, T. (2006). *An Introduction to Bayesian Analysis*, Springer, New York.

Hastie, T. et Tibshirani R. (1990). *Generalized Additive Models*. Chapman & Hall/CRC, New York.

He, X. et Shi, P. (1998). Monotone B-spline smoothing. *Journal of the American Statistical Association*, **93**, 643–650.

Judge, G. G. et Takayama, T. (1966), Inequality restrictions in regression analysis. *Journal of the American Statistical Association*, **61**, 166–181.

Klein, L. R. (1953), *A Textbook of Econometrics*. Row, Peterson, New York.

Massey, F. J., Jr. (1951). The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association*, **46**, 68–78.

Merleau, J., Perreault, L., Angers, J.-F. et Favre, A.-C. (2007). Bayesian modeling of hydrographs. *Water Resources Research*, **43**, W10432.

Murphy, K. (2007). *Conjugate Bayesian analysis of the Gaussian distribution*. University of British Colombia. [En ligne]. Disponible : http ://www.cs.ubc.ca/ murphyk/Papers/bayesGauss.pdf [Consulté le 30 mars 2014].

Nydick, S. W. (2012). *The Wishart and inverse Wishart distributions*. University of Minnesota. [En ligne]. Disponible : http ://www.tc.umn.edu/ nydic001/docs/unpubs/Wishart_Distribution.pdf [Consulté le 30 mars 2014].

Ramsay, J. O. (1988). Monotone regression splines in action. *Statistical Science*, **3**, 425–441.

Robert, C. P. (1995). Simulation of truncated normal variables. *Statistics and Computing*. **5**, 121–125.

Robert, C. P. (2001). *The Bayesian Choice*. 2e édition, Springer-Verlag, New York.

Seber, G. A. F. (1977). *Linear Regression Analysis*. Wiley, New York.

Taboga, M. (2010). *Lectures on Probability and Mathematical Statistics*. 2e édition, Amazon CreateSpace.

Tiao, G. C. et Zellner, A. (1964). Bayes's theorem and the use of prior knowledge in regression analysis. *Biometrika*, **51**, 219–230.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* **58**, 267–288.

von Neumann, J. (1951). Various techniques used in connection with random digits. *National Bureau of Standards Applied Math Series*, **12**, 36–38.

Wang, J., Cheng, F. et Yang, L. (2013) Smooth simultaneous confidence bands for cumulative distribution functions. *Journal of Nonparametric Statistics*, **25**, 395–407.

Wegman, E. et Wright I. (1983). Splines in statistics. *Journal of the American Statistical Association*, **78**, 351–365.

Weisberg, S. (1985). *Applied Linear Regression*. Wiley, New York.

Wilhelm, S. et Manjunath, B. G. (2010). tmvtnorm : A package for the truncated multivariate normal distribution. *The R Journal*, **2**, 25–29.

Annexe A

DENSITÉS *A POSTERIORI* DU MODÈLE INVERSE-GAMMA ET INVERSE-WISHART

Nous détaillons ici les calculs des densités *a posteriori* du modèle inverse-gamma et inverse-Wishart. Pour plus de contexte, voir la section 2.2.4. Nous posons les distributions *a priori* suivantes :

$$\Sigma_0|\nu_0, \Lambda_0 \sim W^{-1}(\nu_0, \Lambda_0),$$

$$\sigma^2|a_0, b_0 \sim \Gamma^{-1}(a_0, b_0)$$

$$\text{et } B|B_0, \sigma^2, \Sigma_0 \sim N_p(B_0, \sigma^2\Sigma_0).$$

Rappelons que la densité de la distribution inverse-Wishart est

$$\pi(\Sigma_0|\nu_0, \Lambda_0) = \frac{|\Lambda_0|^{\frac{\nu_0}{2}}}{2^{\frac{\nu_0 p}{2}}\Gamma_p(\frac{\nu_0}{2})}|\Sigma_0|^{-\frac{\nu_0+p+1}{2}} \exp\left(-\frac{1}{2}\text{tr}(\Lambda_0\Sigma_0^{-1})\right),$$

où $\Gamma_p(\cdot)$ correspond à la fonction gamma multivariée et est définie de la façon suivante :

$$\Gamma_p(a) = \pi^{p(p-1)/4}\prod_{j=1}^{p}\Gamma(a + (1-j)/2).$$

Aussi, on définit les densités marginales de Y de la façon suivante : $m_1(Y|\sigma^2, \Sigma_0)$ est la densité marginale de Y sachant σ^2 et Σ_0, c'est la densité de Y intégrée sur B, $m_2(Y|\Sigma_0)$ est la densité m_1 intégrée sur σ^2 et $m_3(Y)$ est la densité m_2 intégrée sur Σ_0.

Nous allons procéder par étape en se servant du fait que

$$\pi(B|\mathcal{D}, B_0, \sigma^2, \Sigma_0) = \frac{f(Y|B, \sigma^2, \Sigma_0)\pi(B|B_0, \sigma^2, \Sigma_0)}{m_1(Y|\sigma^2, \Sigma_0)}, \tag{A.0.1}$$

$$\pi(\sigma^2|\mathcal{D}, a_0, b_0, \Sigma_0) = \frac{m_1(Y|\sigma^2, \Sigma_0)\pi(\sigma^2|a_0, b_0, \Sigma_0)}{m_2(Y|\Sigma_0)} \text{ et} \tag{A.0.2}$$

$$\pi(\Sigma_0|\mathcal{D}, \nu_0, \Lambda_0) = \frac{m_2(Y|\Sigma_0)\pi(\Sigma_0|\nu_0, \Lambda_0)}{m_3(Y)}, \tag{A.0.3}$$

et du fait que nous pouvons obtenir les distributions marginales à chaque étape, nous dérivons successivement les distributions *a posteriori* une à une.

D'abord de (A.0.1) découle que

$$\begin{aligned}
\pi(B|B_0, \sigma^2, \Sigma_0)f(Y|B, \sigma^2, \Sigma_0) &= \frac{1}{(2\pi)^{\frac{p}{2}}(\sigma^2)^{\frac{p}{2}}|\Sigma_0|^{\frac{1}{2}}} exp\left(\frac{-1}{2\sigma^2}(B - B_0)^T\Sigma_0^{-1}(B - B_0)\right) \\
&\quad \times \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} exp\left(\frac{-1}{2\sigma^2}(Y - XB)^T(Y - XB)\right) \\
&= ... \\
&= \frac{1}{(2\pi)^{\frac{p}{2}}(\sigma^2)^{\frac{p}{2}}|\Sigma_n|^{\frac{1}{2}}} exp\left(\frac{-1}{2\sigma^2}(B - B_n)^T\Sigma_n^{-1}(B - B_0)\right) \\
&\quad \times \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}}\frac{|\Sigma_n|^{\frac{1}{2}}}{|\Sigma_0|^{\frac{1}{2}}} exp\left(\frac{-1}{2\sigma^2}(SSE \right. \\
&\qquad \left. + (\hat{B} - B_0)^T((X^TX)^{-1} + \Sigma_0)^{-1}(\hat{B} - B_0)\right) \\
&= \pi(B|\mathcal{D}, B_0, \sigma^2, \Sigma_0)m_1(Y|\sigma^2, \Sigma_0),
\end{aligned}$$

où

$$\begin{aligned}
B_n &= \Sigma_n\left(X^TX\hat{B} + \Sigma_0^{-1}B_0\right) \\
\text{et } \Sigma_n &= \left(X^TX + \Sigma_0^{-1}\right)^{-1}.
\end{aligned}$$

Avec ce résultat, on peut obtenir explicitement m_1 puisque $\pi(B|\mathcal{D}, B_0, \sigma^2, \Sigma_0)$ doit intégrer à 1. À l'aide de (A.0.2), on obtient donc que

$$\pi(\sigma^2|a_0, b_0)m_1(Y|\sigma^2, \Sigma_0)$$

$$
\begin{aligned}
&= \frac{b_0^{a_0}}{\Gamma(a_0)}(\sigma^2)^{-(a_0+1)}\exp\left(-\frac{b_0}{\sigma^2}\right) \times \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}}\frac{|\Sigma_n|^{\frac{1}{2}}}{|\Sigma_0|^{\frac{1}{2}}} \\
&\qquad exp\left(\frac{-1}{2\sigma^2}(SSE + (\hat{B}-B_0)^T((X^TX)^{-1}+\Sigma_0)^{-1}(\hat{B}-B_0)\right) \\
&= \frac{b_0^{a_0}}{\Gamma(a_0)}\frac{(\sigma^2)^{-(a_0+1)}}{(\sigma^2)^{\frac{n}{2}}}\frac{1}{(2\pi)^{\frac{n}{2}}}\frac{|\Sigma_n|^{\frac{1}{2}}}{|\Sigma_0|^{\frac{1}{2}}} \\
&\qquad \times \exp\left(-\frac{2b_0 + SSE + (\hat{B}-B_0)^T((X^TX)^{-1}+\Sigma_0)^{-1}(\hat{B}-B_0)}{2\sigma^2}\right) \\
&= \frac{b_n^{a_n}}{\Gamma(a_n)}(\sigma^2)^{-(a_n+1)}\exp\left(-\frac{b_n}{\sigma^2}\right) \\
&\qquad \times \frac{1}{(2\pi)^{\frac{n}{2}}}b_0^{a_0}\frac{\Gamma(a_n)}{\Gamma(a_0)}b_n^{-a_n}\frac{|\Sigma_n|^{\frac{1}{2}}}{|\Sigma_0|^{\frac{1}{2}}} \\
&= \pi(\sigma^2|\mathcal{D}, a_0, b_0, \Sigma_0)m_2(Y|\Sigma_0),
\end{aligned}
$$

où

$$
\begin{aligned}
a_n &= a_0 + \frac{n}{2} \\
\text{et } b_n &= \frac{2b_0 + SSE + (\hat{B}-B_0)^T((X^TX)^{-1}+\Sigma_0)^{-1}(\hat{B}-B_0)}{2}.
\end{aligned}
$$

On peut encore trouver explicitement m_2 puisque $\pi(\sigma^2|\mathcal{D}, a_0, b_0, \Sigma_0)$ doit intégrer à 1. Finalement, avec (A.0.3), on obtient que

$$
\begin{aligned}
\pi(\Sigma_0|\nu_0, \Lambda_0)m_2(Y|\Sigma_0) &= \frac{|\Lambda_0|^{\frac{\nu_0}{2}}}{2^{\frac{\nu_0 p}{2}}\Gamma_p(\frac{\nu_0}{2})}|\Sigma_0|^{-\frac{\nu_0+p+1}{2}}\exp\left(-\frac{1}{2}tr(\Lambda_0\Sigma_0^{-1})\right) \\
&\qquad \times \frac{1}{(2\pi)^{\frac{n}{2}}}b_0^{a_0}\frac{\Gamma(a_n)}{\Gamma(a_0)}b_n^{-a_n}\frac{|\Sigma_n|^{\frac{1}{2}}}{|\Sigma_0|^{\frac{1}{2}}} \\
&= (b_n)^{-a_n}|\Sigma_n|^{\frac{1}{2}}|\Sigma_0|^{-\frac{\nu_0+p+2}{2}}\frac{|\Lambda_0|^{\frac{\nu_0}{2}}}{2^{\frac{\nu_0 p}{2}}\Gamma_p(\frac{\nu_0}{2})}\exp\left(-\frac{1}{2}tr(\Lambda_0\Sigma_0^{-1})\right) \\
&\qquad \times \frac{1}{(2\pi)^{\frac{n}{2}}}b_0^{a_0}\frac{\Gamma(a_n)}{\Gamma(a_0)} \\
&= \pi(\Sigma_0|\mathcal{D}, \nu_0, \Lambda_0)m_3(Y),
\end{aligned}
$$

Nous avons ainsi les distributions *a posteriori* suivantes :

$$B|\mathcal{D}, B_0, \sigma^2, \Sigma_0 \sim N_p(B_n, \sigma^2\Sigma_n),$$

$$\sigma^2|\mathcal{D}, a_0, b_0, \Sigma_0 \sim \Gamma^{-1}(a_n, b_n)$$

$$\text{et } \pi(\Sigma_0|\mathcal{D}, \nu_0, \Lambda_0) \propto b_n^{-a_n}|\Sigma_n|^{\frac{1}{2}}|\Sigma_0|^{-\frac{\nu_0+p+2}{2}} \exp\left(-\frac{1}{2}\text{tr}(\Lambda_0\Sigma_0^{-1})\right).$$

Notons qu'il est impossible d'obtenir l'expression exacte de $\pi(\Sigma_0|\mathcal{D}, \nu_0, \Lambda_0)$ puisque $m_3(Y)$ est inconnue.

Annexe B

APPROXIMATION DE LA COVARIANCE APRÈS LA TRANSFORMATION ANGULAIRE

On définit d'abord $F_i = F_n(X_i)$, la variable aléatoire correspondant à la valeur de la fonction empirique pour la i-ème observation, puis $\widehat{F}_i = F_n(x_i)$, la valeur de la fonction empirique pour la i-ème observation pour un échantillon particulier. On estime $E[F_i]$ par $\widehat{F}_i$. On définit maintenant $g(F_i) = \arcsin(\sqrt{F_i})$ la transformation angulaire de F_i. On a déjà noté que

$$Cov[F_r, F_s] = \frac{n}{(n+2)^2}(min(F_r, F_s) - F_r F_s).$$

Klein (1953) nous indique que :

$$Cov[g(F_r), g(F_s)] \approx \sum_i \left.\frac{\partial g(F_r)}{\partial F_i}\right|_{\widehat{F}_r} \left.\frac{\partial g(F_s)}{\partial F_i}\right|_{\widehat{F}_s} + \sum_i \sum_{j \neq i} \left.\frac{\partial g(F_r)}{\partial F_i}\right|_{\widehat{F}_r} \left.\frac{\partial g(F_s)}{\partial F_i}\right|_{\widehat{F}_s} Cov[F_i, F_j].$$

Ici, on a que

$$\frac{\partial g(F_r)}{\partial F_i} = \begin{cases} \dfrac{1}{2\sqrt{F_r(1-F_r)}} & \text{si } r = i \\ 0 & \text{sinon} \end{cases}$$

Donc, on a que

$$Cov[g(F_r), g(F_s)] \approx \frac{1}{4\sqrt{\widehat{F}_r(1-\widehat{F}_r)\widehat{F}_s(1-\widehat{F}_s)}} Cov[F_r, F_s].$$