Gilles STUPFLER

Docteur en Mathématiques Appliquées
Allocataire de recherche et moniteur
IRMA, Université de Strasbourg

Etat civil

Date et lieu de naissance :	15 mai 1988 à Strasbourg (24 ans)
Nationalité :	Française
Situation familiale :	Célibataire
Adresse professionnelle :	UFR Mathématique-Informatique 7, rue René Descartes, 67084 Strasbourg Cedex Tél. : 03.68.85.02.13
Adresse électronique :	`stupfler@math.unistra.fr`
Page personnelle :	`www-irma.u-strasbg.fr/php/home.php?qui=stupfler`

Diplômes et formation

2009-2011 **Doctorat de Mathématiques Appliquées, spécialité Statistique**
Titre : Un modèle de Markov caché en assurance
et Estimation de frontière et de point terminal
Date et lieu de soutenance : 10 novembre 2011, Université de Strasbourg
Directeurs de thèse : Stéphane Girard et Armelle Guillou
Rapporteurs : Hansjoerg Albrecher et Irène Gijbels

2006-2009 **Magistère de Mathématiques**
Mention Très Bien, rang 1, Université de Strasbourg

2008-2009 **Agrégation de mathématiques**
Rang 21

2008-2009 **M2 Mathématiques fondamentales et appliquées**
Mention Très Bien, rang 1, Université de Strasbourg

2007-2008 **M1 Mathématiques fondamentales et appliquées**
Mention Très Bien, rang 1, Université de Strasbourg

2006-2007 **L3 Mathématiques fondamentales et appliquées**
Mention Très Bien, rang 1, Université de Strasbourg

2004-2006 **Classes préparatoires MPSI puis MP***
Lycée Kléber, Strasbourg

Parcours professionnel

2009-2012 **Allocataire de recherche et moniteur**
Equipe “Statistique” de l’IRMA, Université de Strasbourg

Février-juin 2009 **Stage de M2 à l’Université de Strasbourg**
Sujet : Méthodes matricielles pour les probabilités de ruine
Directeur du stage : Karl-Theodor Eisele

Enseignement

Allocataire de recherche et moniteur, Université de Strasbourg (2009-2012)

- **TD de Probabilité-Statistique, L2 Mathématiques (septembre-décembre 2009) et Mathématiques-Economie (septembre-décembre 2010 et 2011)**
  Contenu : notions de combinatoire, probabilités conditionnelles, lois discrètes, fonction génératrice, lois continues, fonction caractéristique, introduction aux théorèmes limites. Total horaire : 118 heures de TD.
- **TD de Statistique Mathématique, L3 Mathématiques (janvier-mai 2010), puis responsable du suivi pédagogique (septembre-décembre 2010 et 2011)**
  Contenu : espérance conditionnelle, vecteurs gaussiens, modèle linéaire, introduction à la statistique inférentielle. Total horaire : 66 heures de TD.
- **TD de Probabilité-Statistique, L3 Mathématiques (novembre-décembre 2011)**
  Contenu : introduction à la statistique inférentielle, méthode du maximum de vraisemblance, loi des grands nombres et théorème central limite. Total horaire : 9 heures de TD.

Le total de mes activités d'enseignement sur cette période s'élève à 193 heures de travaux dirigés sur 3 ans. J'ai également participé à la conception et à la surveillance des examens pour ces modules, ainsi qu'aux jurys des diplômes concernés.

Activités de recherche

Pour le moment, mes activités de recherche se divisent en deux parties distinctes : l'application des modèles de Markov caché à l'actuariat et la théorie des valeurs extrêmes avec application à l'estimation de point terminal et de frontière.

Modèles de Markov caché en actuariat

Référence : [4].

Une partie importante des recherches en mathématiques appliquées à l'actuariat consiste à étudier les processus de pertes en assurance, afin notamment d'aider à la compréhension du processus de réserve. Dans l'article [4], je me suis intéressé à l'étude d'un nouveau modèle pour un tel processus. Le processus considéré est un triplet $(J,\, N,\, S)$, tel que $(J,\, N)$ est un processus de Poisson à modulation markovienne (MMPP) : autrement dit, J est un processus de Markov à temps continu sur un espace d'états fini $\{1,\, \ldots,\, r\}$, $r \geq 1$ et N est un processus à temps continu tel que, conditionnellement à J, N est un processus de Poisson inhomogène d'intensité instantanée $t \mapsto \lambda_{J(t)}$. Le processus J n'est pas observé, d'où la terminologie de modèle de Markov caché. D'autre part, $S = (S_1,\, \ldots,\, S_n)$ est un processus de pertes (éventuellement multivarié), c'est-à-dire que les S_k sont des processus constants par morceaux dont les sauts sont positifs. Le comportement de S est supposé être contrôlé par $(J,\, N)$ de la façon suivante : les S_k ne peuvent sauter qu'à un instant de saut de N et si N saute lorsque J est dans l'état i, alors un saut simultané des processus $S_{k_1},\, \ldots,\, S_{k_p}$ se produit avec probabilité $p(i,\, e)$, où $e = \{k_1,\, \ldots,\, k_p\}$. Dans ce cas, la valeur X du saut a pour loi $\mathbb{P}_{\theta(i,\, e)}$, où $(\mathbb{P}_\theta)_{\theta\in\Theta}$ est un modèle statistique paramétrique. Enfin, les sauts de S sont indépendants sachant $(J,\, N)$. Les paramètres de ce modèle sont :

– les éléments ℓ_{ij} du générateur L de J ;
– les intensités de saut λ_i de N ;
– les probabilités $p(i,\, e)$;

– les paramètres $\theta(i,\, e)$.

En supposant disposer des données suivantes :

- le nombre d'états r de J ;
- la connaissance des processus N et S jusqu'à un certain temps T, qu'on suppose être un instant de saut de N,

et en adaptant une preuve de T. Rydén (1994), on prouve la consistance de l'estimateur du maximum de vraisemblance (EMV) du paramètre $\Phi = (l_{ij},\, \lambda_i,\, p(i,\, e),\, \theta(i,\, e))$. On donne un algorithme EM pour calculer l'EMV, ainsi qu'une procédure d'estimation initiale. On utilise ce procédé sur des jeux de données réelles en assurance dans le cas univarié. Enfin, on réalise une étude sur simulations dans le cas d'un processus S multivarié.

Perspectives – A court terme, il serait intéressant d'implémenter une méthode d'estimation permettant de traiter le cas où les λ_j sont des fonctions du temps, comme par exemple pour des risques de type catastrophe naturelle pour lesquels la variation des λ_j est saisonnière. On peut pour cela utiliser une vraisemblance "splittée", dont on peut démontrer les propriétés asymptotiques en utilisant les propriétés d'ergodicité du modèle ; ceci fait l'objet d'un travail en cours (avec A. Guillou et S. Loisel). Il serait également souhaitable de décrire un algorithme d'estimation initiale des paramètres plus précis, afin de réduire le temps de calcul nécessaire à l'estimation. J'aimerais aussi appliquer le procédé mentionné ci-dessus à des données réelles dans le cadre multivarié. Enfin, remarquons que la méthode décrite ci-dessus requiert la connaissance du nombre d'états r de J, ce qui n'est pas toujours réaliste en pratique. A plus long terme, il serait donc intéressant de développer une méthode permettant de déterminer le nombre d'états du processus d'environnement J et d'étudier ses performances, en s'appuyant sur les travaux existants de T. Rydén (1995).

Théorie des valeurs extrêmes : estimation de point terminal

Références : [1], [3].

Mes recherches en théorie des valeurs extrêmes se sont concentrées jusqu'à présent sur le problème d'estimation d'un point terminal. Le contexte est le suivant : soient $X_1,\, \ldots,\, X_n$ des copies indépendantes d'une variable aléatoire X, telle que le point terminal

$$\theta = \sup\{x > 0 \,|\, \mathbb{P}(X \leq x) < 1\}$$

de la distribution de X est fini. Le problème est d'estimer θ.

J'ai choisi de m'intéresser à la résolution de ce problème en utilisant une méthode des moments d'ordre élevé. Son principe est le suivant : dans le cas où X est positive, on démontre une relation entre le point terminal et les moments d'ordre élevé de la variable aléatoire X. Pour définir ensuite un estimateur de θ, on remplace chaque moment par son estimateur empirique classique. Dans l'article [1], on a supposé que X est positive et considéré l'estimateur $\widehat{\theta}_n$ défini par

$$\frac{1}{\widehat{\theta}_n} = \frac{1}{ap_n}\left\{((a+1)p_n+1)\frac{\widehat{\mu}_{(a+1)p_n}}{\widehat{\mu}_{(a+1)p_n+1}} - (p_n+1)\frac{\widehat{\mu}_{p_n}}{\widehat{\mu}_{p_n+1}}\right\},$$

où $p_n \to \infty$, $a > 0$ et $\widehat{\mu}_{p_n} = \frac{1}{n}\sum_{k=1}^{n} X_k^{p_n}$. Sous une condition sur la vitesse de divergence de la suite (p_n), on démontre la consistance de l'estimateur $\widehat{\theta}_n$ dans un cadre non paramétrique. En supposant que la fonction de répartition F de X appartient au domaine d'attraction de Weibull, on prouve que l'estimateur $\widehat{\theta}_n$ est asymptotiquement normal sous une condition de second ordre classique sur F. Les résultats sont illustrés sur simulations.

Une généralisation de cette méthode est proposée dans [3], lorsque X n'est plus nécessairement positive : on considère l'estimateur

$$\widetilde{\theta}_n := \frac{1}{a}\left\{\ln\left[\frac{\widetilde{\mu}_{p_n}}{\widetilde{\mu}_{p_n+1}}\right] - \ln\left[\frac{\widetilde{\mu}_{(a+1)p_n}}{\widetilde{\mu}_{(a+1)p_n+a+1}}\right]\right\}$$

où $\widetilde{\mu}_{p_n} = \frac{1}{n}\sum_{k=1}^{n} e^{p_n X_k}$. Les propriétés asymptotiques de l'estimateur sont décrites, en particulier lorsque la fonction de répartition de X appartient au domaine d'attraction de Weibull.

Perspectives – Une première direction à court terme pour poursuivre cet axe de recherche serait la construction et l'étude d'un test permettant de distinguer les distributions à queue légère ayant un point terminal fini, donc appartenant au domaine d'attraction de Gumbel, des distributions appartenant au domaine d'attraction de Weibull. A plus long terme, je souhaiterais notamment proposer une méthode permettant de choisir l'ordre des moments (p_n) et le paramètre a intervenant dans la définition des estimateurs proposés. Ensuite, j'aimerais concevoir et travailler sur un estimateur de l'indice des valeurs extrêmes utilisant des moments d'ordre élevé pour une variable aléatoire dont la fonction de répartition appartient au domaine d'attraction de Fréchet, ce dernier domaine regroupant la majorité des applications pratiques, notamment en hydrologie et en actuariat. Il serait également intéressant d'étudier des estimateurs de quantiles extrêmes en utilisant la méthode des moments d'ordre élevé et d'appliquer le procédé obtenu à des cas pratiques : calcul de niveaux de retour de pluie en hydrologie, estimation du coût d'un sinistre d'ampleur exceptionnelle en actuariat, par exemple.

Théorie des valeurs extrêmes : estimation de frontière

Référence : [2].

Je me suis aussi intéressé à la résolution du problème d'estimation d'une frontière. Ce problème s'écrit comme suit : étant données $(X_1, Y_1), \ldots, (X_n, Y_n)$ n copies indépendantes d'un vecteur aléatoire (X, Y) dont la densité commune a un support $S = \{(x, y) \in \Omega \times \mathbb{R} \mid 0 \leq y \leq g(x)\}$, où Ω est un sous-ensemble compact de $\mathbb{R}^d$, $d \geq 1$, on cherche à estimer la fonction inconnue g, appelée la frontière de S. Dans l'article [2], on suppose que X est à densité f et on utilise une régression par noyau couplée à la méthode des moments d'ordre élevé pour la variable Y : l'estimateur est défini par

$$\frac{1}{\widehat{g}_n(x)} = \frac{1}{ap_n}\left[((a+1)p_n+1)\frac{\widehat{\mu}_{(a+1)p_n}(x)}{\widehat{\mu}_{(a+1)p_n+1}(x)} - (p_n+1)\frac{\widehat{\mu}_{p_n}(x)}{\widehat{\mu}_{p_n+1}(x)}\right]$$

avec $\widehat{\mu}_{p_n}(x) = \frac{1}{n}\sum_{i=1}^{n} Y_i^{p_n} K_{h_n}(x - X_i)$, où $p_n \to \infty$, $h_n \to 0$, $a > 0$, K est une densité de probabilité à support contenu dans la boule unité de $\mathbb{R}^d$, qu'on appelle noyau et $K_h(u) = h^{-d} K(u/h)$. La suite (h_n) est appelée fenêtre. Sous des hypothèses sur la vitesse de divergence de (p_n) ainsi que sur la taille de la fenêtre (h_n) et à l'aide d'une hypothèse non paramétrique sur la fonction de répartition conditionnelle de Y sachant X, on prouve que l'estimateur $\widehat{g}_n(x)$ est ponctuellement consistant. En supposant que la fonction de répartition conditionnelle de Y sachant $\{X = x\}$ appartient au domaine d'attraction de Weibull et possède un développement polynomial au voisinage de $g(x)$ (on dit qu'elle appartient à la classe de Hall) et sous des hypothèses de régularité classiques, on prouve que l'estimateur construit est ponctuellement asymptotiquement normal. Les résultats sont illustrés sur simulations.

Perspectives – Un travail est en cours pour démontrer la consistance uniforme presque sûre de l'estimateur dans un cadre non-paramétrique (en collaboration avec S. Girard et A. Guillou) et pour calculer sa vitesse de convergence presque sûre lorsque la loi conditionnelle de Y sachant X appartient à la classe de Hall. A plus long terme, je souhaiterais obtenir la loi de l'écart uniforme entre l'estimateur et la frontière, afin de pouvoir construire des bandes de confiance asymptotiques pour l'estimateur $\widehat{g}_n$. J'aimerais aussi concevoir une procédure permettant de choisir la fenêtre (h_n), l'ordre des moments (p_n) et le paramètre a intervenant dans la définition de l'estimateur. Des recherches sont également en cours pour appliquer cette méthode à des données réelles, notamment en hydrologie et en économétrie : dans ce dernier cadre, il est fréquent que les jeux de données réelles présentent des valeurs aberrantes (outliers), il serait donc intéressant de tester la sensibilité de notre estimateur à ces valeurs et, le cas échéant, de proposer une correction pour le rendre moins sensible à ces valeurs.

Publications, communications et séminaires

En cas de convocation à l'audition, j'adresserai les articles [1] à [4].

Article accepté

[1] S. Girard, A. Guillou, and G. Stupfler. Estimating an endpoint with high order moments. *Test*, à paraître, 2012.

Article en révision

[2] S. Girard, A. Guillou, and G. Stupfler. Frontier estimation with kernel regression on high order moments, en révision.

Articles soumis

[3] S. Girard, A. Guillou, and G. Stupfler. Estimating an endpoint with high order moments in the Weibull domain of attraction, soumis.

[4] A. Guillou, S. Loisel, and G. Stupfler. Estimation of the parameters of a Markov-modulated loss process in insurance, soumis.

Conférence internationale

[5] G. Stupfler. Estimating an endpoint using high order moments. In *7th Conference on Extreme Value Analysis, Probabilistic and Statistical Models and their Applications*, Lyon, France, 2011.

Conférence nationale

[6] G. Stupfler. Estimation de point terminal dans le domaine d'attraction de Weibull par une méthode des moments d'ordre élevé. In *44èmes Journées de Statistique*, Bruxelles, Belgique, 2012.

Invitation à des séminaires

[7] G. Stupfler. Estimation de point terminal dans le domaine d'attraction de Weibull par une méthode des moments d'ordre élevé. Séminaire de statistique de l'Université de Grenoble, mars 2012.

[8] G. Stupfler. Estimation de frontière par une méthode des moments d'ordre élevé. Séminaire de statistique de l'Université Toulouse III, mars 2012.

[9] G. Stupfler. Estimation des paramètres d'un processus de pertes à modulation markovienne en assurance. Séminaire de statistique de l'Université Lyon I, mars 2012.

[10] G. Stupfler. Estimation de frontière par une méthode des moments d'ordre élevé. Séminaire de statistique de la Toulouse School of Economics, février 2012.

[11] G. Stupfler. Estimation de frontière par une méthode des moments d'ordre élevé. Séminaire de probabilités-statistique de l'Université Montpellier II, février 2012.

[12] G. Stupfler. Estimation des paramètres d'un processus de pertes à modulation markovienne en assurance. Séminaire du laboratoire SAF à l'ISFA, Université Lyon I, décembre 2011.

[13] G. Stupfler. Méthode des moments d'ordre élevé pour un point terminal. Séminaire de statistique de l'Université de Strasbourg, avril 2011.

J'ai également exposé à trois reprises au séminaire doctorants de l'Université de Strasbourg.

Contacts

Recherche

- Hansjoerg Albrecher, professeur à HEC Lausanne.
  E-mail : `Hansjoerg.Albrecher@unil.ch`
- Stéphane Girard, chargé de recherche à l'INRIA Rhône-Alpes.
  E-mail : `stephane.girard@inria.fr`
- Armelle Guillou, professeur à l'Université de Strasbourg.
  E-mail : `guillou@math.unistra.fr`
- Stéphane Loisel, professeur à l'ISFA, Université de Lyon 1.
  E-mail : `stephane.loisel@univ-lyon1.fr`

Enseignement

- Ségolen Geffray, co-responsable du module Statistique Mathématique.
  E-mail : `geffray@math.unistra.fr`
- Raphaële Supper, responsable du suivi pédagogique en Probabilité-Statistique L3.
  E-mail : `supper@math.unistra.fr`

Informations diverses

- Logiciels scientifiques : Scilab, R.
- Utilisation courante d'OpenOffice (Calc, Writer).
- Anglais : bilingue.
- Espagnol : pratique courante.