Inference for some stochastic processes : with application on thunderstorm data

Van-Cuong Do

THÈSE DE DOCTORAT DE

L'UNIVERSITE BRETAGNE SUD
COMUE UNIVERSITÉ BRETAGNE LOIRE

ECOLE DOCTORALE N° 601
Mathématiques et Sciences et Technologies
de l'Information et de la Communication
Spécialité : *Mathématiques et leurs Interactions*

Par **Van Cuong DO**

Analyse statistique de processus stochastiques Application sur des données d'orages

Thèse présentée et soutenue à Vannes, le 19 avril 2019
Unité de recherche : CNRS UMR 6205
Thèse N° :

Rapporteurs avant soutenance :

Mitra Fouladirad, Professeure, Université de Technologie de Troyes

Jean-Yves Dauxois, Professeur, Université de Toulouse

Composition du Jury :

Président : Prénom Nom Fonction et établissement d'exercice (9) *(à préciser après la soutenance)*

Examinateurs : **Sophie Mercier**, Professeure, Université de Pau et du pays de l'Adour
Jean Vaillant, Professeur, Université des Antilles
Gilles Durieu, Professeur, Université Bretagne Sud

Dir. de thèse : **Evans Gouno,** MCF HDR, Université Bretagne Sud

UNIVERSITE BRETAGNE LOIRE / MATHSTIC

Université Bretagne Sud

Titre : **Analyse statistique de processus stochastiques : application sur des données d'orages**

Mots clés : Processus power-law, processus d'intensité exponentielle, processus auto-excité, covariable, maximum de vraisemblance, estimateur de Bayes.

Résumé : Les travaux présentés dans cette thèse concernent l'analyse statistique de cas particuliers du processus de Cox.

Dans une première partie, nous proposons une synthèse des résultats existants sur le processus power-law (processus d'intensité puissance), synthèse qui ne peut être exhaustive étant donné la popularité de ce processus. Nous considérons une approche bayésienne pour l'inférence des paramètres de ce processus qui nous conduit à introduire et à étudier en détails une distribution que nous appelons *loi H-B*. Cette loi est une loi conjuguée.

Nous proposons des stratégies d'élicitation des hyperparamètres et étudions le comportement des estimateurs de Bayes par des simulations.

Dans un deuxième temps, nous étendons ces travaux au cas du processus d'intensité exponentielle (exponential-law process). Nous considérons le maximum de vraisemblance techniques. Pour l'analyse bayésianne, de la même façon, nous définissons et étudions une loi conjuguée pour l'analyse bayésienne de ce dernier : la loi Gumbel-Modifié et la loi Gamma-Gumbel-Modifié.

Dans la dernière partie de la thèse, nous considérons un processus auto-excité qui intègre une covariable. Ce travail est motivé, à l'origine, par un problème de fiabilité qui concerne des données de défaillances de matériels exposés à des environnements sévères. Les résultats sont illustrés par des applications sur des données d'activités orageuses collectées dans deux départements français.

Enfin, nous donnons quelques directions de travail et perspectives de futurs développements de l'ensemble de nos travaux.

Title: Inference for some stochastic processes with application on thunderstorm data

Keywords: Power-law process, exponential-law process, self-exciting point process, power-law covariate model, maximum likelihood estimation, Bayes estimation.

Abstract: The work presented in this PhD dissertation concerns the statistical analysis of some particular cases of the Cox process.

Firstly, we study the power-law process (PLP). Since the literature for the PLP is abundant, we suggest a state-of-art for the process. For classical approach of the maximum likelihood estimation, we recall some important properties of the MLE of the PLP. For Bayesian approach, we begin with non-informative priors and then try different parametrizations to employ conjugate priors that can integrate different scenarios of prior guesses. That leads us to define a family of distributions that we name H-B distribution as the natural conjugate priors for the PLP. Bayesian analysis with the conjugate priors are conducted via a simulation study and an application on real data.

Secondly, we study the exponential-law process (ELP). We review the maximum likelihood techniques. For Bayesian analysis of the ELP, we define conjugate priors: The Modified-Gumbel distribution and the Gamma-Modified-Gumbel distribution. We conduct a simulation study to compare maximum likelihood estimates and Bayesian estimates.

Thirdly, we investigate self-exciting point processes and we introduce a power-law covariate model to this process. A maximum likelihood procedure for the model is proposed and the Bayesian approach is suggested.

Lastly, we employ an application on thunderstorm data collected in two French regions. We consider a strategy to define a thunderstorm as a temporal process associated with the charges in a particular location. Some selected thunderstorms are analyzed. We propose a reduced maximum likelihood procedure to estimate the parameters of the Hawkes process. We then fit some thunderstorms to the power-law covariate self-exciting point process taking into account the associated charges.

Inference for Some Stochastic Point Processes with Application to Thunderstorm Data

Van-Cuong DO

Laboratoire de Mathématiques de Bretagne Atlantique
Université de Bretagne Sud

Remerciements

Mon long voyage de mathématiques et de découverte de la culture française touche à sa fin. En quatre ans et trois mois, j'ai profité de mon séjour ici à Vannes qui est une petite ville merveilleuse située à proximité de la mer du golfe du Morbihan. La France est un pays magnifique. C'est une chance d'avoir pu rencontrer toutes ces personnes à la fois dans le laboratoire de mathématiques (LMBA : Laboratoire de Mathématiques de Bretagne Atlantique) mais aussi partout en France. Je les remercie sincèrement pour leur gentillesse. Je ne peux pas toutes les nommer ici, mais elles sont gravées dans mon coeur.

Début 2015, j'ai commencé mon programme de doctorat sous la direction de Monsieur Evans Gouno que je souhaite remercier tout particulièrement et que je respecte énormément. Il s'agit d'un projet sur la modélisation statistique pour les données d'orage. Ce projet nécessite beaucoup de connaissances, mais aussi de techniques statistiques pour analyser pour ensuite déduire des données réelles. Il m'a proposé des processus stochastiques comme des processus ponctuels auto-excités, ainsi qu'un processus de loi de puissance sur lequel travailler, que je n'aurais pas étudié s'il n'avait pas été là. Initialement, je ne connaissais pas beaucoup de méthodes et de techniques statistiques; en effet mon travail concernait les probabilités. Le fait est qu'il y a un fossé entre la pratique et la théorie, c'est pourquoi j'ai traversé une période difficile.

Je souhaite remercier les rapporteurs, Prof. Mitra Fouladirah et Prof. Jean-Yves Dauxois, pour leur temps de lire ce manuscrit et pour leurs commentaires constructifs. Je mercie également les autres membres de jury Prof. Jean Vaillant, Prof. Sophie Mercier et Prof. Gilles Durrieu d'avoir accepté de prendre le temps d'évaluer mon travail.

Au cours de mon doctorat, j'ai acquis beaucoup de connaissances et d'expérience en statistique grâce à l'aide de mon directeur de thèse. En effet, Monsieur Evans Gouno et son encadrement furent d'une grande efficacité. Je n'oublierai jamais le temps qu'il m'a consacré lorsque des idées me venaient à l'esprit, le temps passé à préparer des sujets conférences, mais aussi à réviser mes documents. Ses conseils m'ont permis de me remotiver et d'acquérir de la confiance en moi.

Un nombre "incalculable"de personnes m'ont aidé lorsque je suis arrivé en France, je ne vais pas toutes les citer (il faudrait abattre un arbre sinon). Je n'avais aucune notion en français ce qui m'a posé des problèmes à priori infranchissables pour communiquer avec les autres. Heureusement, les membres du laboratoire furent présents à mes côtés pour m'aider à surmonter les difficultés de la vie quotidienne. Mais j'évoquerai aussi toutes les personnes extérieures qui m'ont tant apporté.

Je souhaite particulièrement remercier mes proches collègues Ronan, Jonathan et Eric qui m'ont aidé à apprendre les bases du français et m'ont soutenu dans mes démarches administratives (de vrais google traduction). Grâce à eux, le barrage de la langue a pu être surmonté rapidement.

Ronan, par sa diligence et son opiniâtreté dans les journées de travail m'a inculqué l'organisation. Grâce à eux, j'ai acquis des réflexes professionnels (une vraie machine). Ses nombreuses compétences m'ont impressionné et j'ai cru que je n'arriverai jamais à son niveau (cela dit il est toujours aussi fort).

Jonathan a été une personne d'une gentillesse et d'une patience inégalables. Eric sans qui je n'aurais pu apprendre la persévérance (Grâce à lui j'ai retrouvé un vélo).

Kévin qui m'a montré comment jouer au tennis de table et au jeu Go chinois (du délire).

Tarik et son courage, il m'a démontré que malgré les embûches, il était possible de terminer sa thèse.

Sans oublier Hui qui a passé du temps avec moi les week-ends à discuter sur de nombreux sujets, ce qui m'a aussi permis de pratiquer le français.

Rabih, Erwan, Hélène, Anne-Charlotte, Jamila, Pathé et Hien qui sont aussi de superbes connaissances.

Je pense à mes collègues du laboratoire LMBA, Monsieur Bertrand Patureau, Monsieur Sylvain Barré, Monsieur Salim Lardjane, Monsieur Quansheng Liu, Monsieur Giles Durrieu, Monsieur Emmanuel Frénod, Monsieur Ion Grama, Monsieur Gaël Meigniez, Monsieur Jacques Froment. Ils ont tous été d'un accueil irréprochable. Je remercie aussi très sincèrement Monsieur Liu, directeur du laboratoire, pour nous avoir constamment écoutés et avoir compris les besoins des doctorants étrangers. Il a brillamment introduit les professeurs de français et a organisé des cours. Sans son aide mon niveau de français n'aurait pas évolué. Françoise, une bénévole qui est venue au laboratoire, sa pédagogie et son approche furent d'un soutien considérable. Martine et ses cours de français du soir depuis maintenant quatre ans. Mais encore Jessica et ses fameux "cours de français débutants".

Je remercie sincèrement le personnel administratif du laboratoire LMBA d'avoir fourni tous les équipements nécessaires à mes travaux de recherche: Madame Véronique Vellet et Madame Sandrine de Olivera.

Au cours de mon séjour à Vannes, des français m'ont accueilli en dehors du laboratoire et ils m'ont beaucoup appris sur la culture française, mais aussi bretonne (vive la Bretagne !). Depuis le début, Dominique et sa famille m'ont intégré dans un monde nouveau. Merci mille fois, Jean-François, San-

drine et Marie pour votre hospitalité et votre générosité envers moi. Je ne savais pas qu'il existait des personnes aussi gentilles que vous sur terre. Merci Lucien et ta famille d'avoir invité mes parents chez toi et d'avoir fait un succulent déjeuner. André et Thérèse pour m'avoir hébergé avec mes parents dans votre si belle maison. Les bons souvenirs du voyage que l'on a fait ensemble en France et au Vietnam me reviennent. Alain et ta famille pour m'avoir invité chez vous avec mes parents. Cette journée où nous sommes allés pêcher à pied sur la plage de Sarzeau est inoubliable. Rexie et ton ouverture d'esprit, chaque soir nos repas dans la cuisine et ta fameuse cuisine des Philippines. Merci Chandru, Shareen et Raphael de m'avoir invité chez vous, avec mes parents pour un délicieux repas indien.

Je tiens également à remercier la communauté vietnamienne présente à Vannes pour avoir partagé avec moi le "mal du pays". Phu pour m'avoir transmis de la patience et de la sérénité. Du et son intelligence, son efficacité dans de nombreuses activités. Thong et ses astuces dans la vie de tous les jours, un vrai "office de tourisme". Thuy et son amitié sans arrière-pensée, un bel exemple de sociabilité. Les époux Tra et Tan pour leur intelligence et leur humilité. J'admire beaucoup votre réussite. Les époux Phuc et Hanh pour m'avoir montré un travail acharné et un accomplissement dans la recherche et la collaboration. Quang et ses excellents repas cuisinés (je meurs de faim à l'idée de pouvoir un jour y goûter à nouveau). Je rends également hommage à certaines familles franco-vietnamiennes : la famille de l'oncle Hung, la famille de l'oncle Khanh, la famille du frère Ngo et de la soeur Thu.

Merci à mes proches et collègues du Vietnam, vous êtes magiques par votre soutien et votre affection.

Enfin, mon amour va tout droit vers ma famille, qui m'aime inconditionnellement et me soutient en toutes circonstances.

Fait à Vannes, le 11 avril 2019

DO Van-Cuong

Acknowledgements

My long journey of mathematics and French culture as a whole is coming to an end. Over the course of four years and three months I have been enjoying my stay here in Vannes, a small wonderful city by the sea of the gulf of Morbihan, in the beautiful country of France. I am grateful that I have met many people inside and outside of the mathematical laboratory life. From the bottom of my heart, I would like to express my sincere thanks for their kindness and support. I cannot give all the names here, but I will never forget anyone for the rest of my life.

In the beginning of 2015, I started my PhD program under the supervision of Dr. Evans Gouno with a project on statistical modeling for thunderstorm data. This project requires a lot of knowledge and techniques in statistics to analyze and make the inference on real data. He suggested to work on the self-exciting point processes and the power-law process. At the beginning, I was not familiar to many statistical methods and techniques since I had been working on theoretical probability. Later I realized that it is not a light matter working from theoretical aspects to practical ones.

During my PhD work, I have learned a lot of knowledge in statistics with the help of my supervisor Dr. Evans Gouno. I am really grateful for his extreme but very effective guidance. He treated me like a friend and value my opinions. Because of that, I respect him more. I will never forget the time we spent together discussing the innovative ideas, preparing for conference presentations and implementation of the papers. His guidance makes me feel more comfortable, confident, and motivated.

My full appreciation is devoted to the referees who spent time reading and giving many useful suggestions: Prof. Mitra Fouladirah and Prof. Jean-Yves Dauxois . I also express my sincere thanks to Prof. Jean Vaillant, Prof. Sophie Mercier, Prof. Gilles Durrieu for accepting to sit on the jury of my defense.

When I arrived in France, I was totally null about French language, so I faced a lot of difficulties on communication with French people. Fortunately, the members of the lab are always willing to help me overcome the difficulties day by day. Ronan, Jonathan, and Eric taught me the basic necessary French words and they helped me as interpreters for my administrative procedures. Their help from the very first days of my stay in France is unforgettable.

I will forever cherish all my sweet memories in this wonderful campus with my friends who influenced my way of living. We have not only studied, relaxed and exchanged our ideas together, but also helped each other always.

Ronan, your diligence in your regular and academic work impressed me so much. I have learned a lot about professional attitude from you. You inspire me to become a better version of myself. Jonathan, you are always nice to me that you give time explaining many of my questions with patience. Eric, you drove me to every bike shop looking for my stolen bike until we found it. That was unbelievable! Kévin, you showed me how to play table tennis and Go Chinese game with your great techniques. Tarik, you showed me how to be patient and never give up. Hui, you are always with me at weekends discussing many interesting topics and practicing French. Rabih, Erwan, Hélène, Anne-Charlotte, Jamila, Pathé, and Hien you are always friendly to me.

I shall be cherishing forever my sweet memories in this wonderful campus with my friends who influenced my way of living. We have not only studied, relaxed and exchanged our ideas together, but helped each other, always. Ronan, you are diligence in your regular and academic work impressed me so much. I have learned a lot of professional attitudes from you. You make me a little embarrassed because you have been doing much better job than me. Jonathan, you are always nice to me that you explain many of my questions with patience. Eric, you drove me to every bike shops seeking for my stolen bike and we found it. That was unbelievable! Kévin, you showed me how to play table tennis and Go Chinese game with your great techniques. Tarik, you showed me how to be patient and never give up. Hui, you are always with me at weekends discussing many interesting topics and practicing French. Rabih, Erwan, Hélène, Anne-Charlotte, Jamila, Pathé, Hien, Chandru, you are always friendly to me.

I have the pleasure in thanking all the faculty members of the lab LMBA. The team includes: Dr. Bertrand Patureau, Dr. Sylvain Barré, Dr. Salim Lardjane, Prof. Quansheng Liu, Prof. Giles Durrieu, Prof. Emmanuel Frénod, Prof. Ion Grama, Prof. Gaël Meigniez, and Prof. Jacques Froment. Bertrand is my great friend, is always kind, and willing to help. I still remember many good meals we prepared in his home. I would like to thank Mr. Liu, the director of the lab who always listens to, and understands the needs of foreign PhD scholars. He introduced French teachers and organized French courses for us. Because of his initiative, we have had great chances to learn French. Madame Françoise Laveuve comes to the lab and teaches us French voluntarily. She is a very good teacher and she is so kind to us. Madame Martine Wallerich teaches us French lessons in the evening. Madame Jessica Kologrecki teaches us French lessons for beginners. Thank to them, I have improved my French level a lot.

I render my sincere thanks to the administrative staffs of the LMBA lab for providing all the necessary facilities required for my research work: Madame Véronique Vellet and Madame Sandrine de Olivera.

I am very much thankful to many friendly French people outside the lab, they have taught me a lot about French culture. I have been greatly supported by Dominique and her family from the beginning. She is so helpful to me that without her I would not have that much experience of French villages, French cities and French landscapes. Jean-François and his family are so kind and generous to me. Lucien and his family welcomed me and my parents to his home having lunch together. André and Thérèse hosted me and my parents in their beautiful house and brought us to visit many nice places in France. We have had a lot of nice memories traveling together in France and in Vietnam. Alain and his family welcomed me and my parents having dinner(s) at their home. Still I remember the time when we went for fishing in the beach of Sarzeau and we got some seafoods. Jaqueline invited me to have dinner several times. Rexie introduces his delicious Filipino cuisines and taught me how to open my mind and my spirit. Thank to Raphael and Shareen for the delicious Indian dinner they shared to me and my parents. Thank also to Chandru, the amazing and interesting Indian guy for his tasty Indian recipes prepared with pride.

I would like to convey thanks to all the Vietnamese friends in Vannes for creating a home away from home by the following: Phú "the real leader" with his great patience and calmness in every situation; Du "the competitive" with his intelligence and his effectiveness in many works; Thông "the reliable guy" with his broadened knowledge in many aspects of life in France; Thuy with her innocence(?) and kindness; Tân and his wife Trà with their smartness and humbleness, I admired them a lot; Phúc and his wife Hạnh with their hard work and dedication in doing research and collaboration; and Quang "the master chef" with his great skill in cooking. I also pay my acknowledgment to some of French-Vietnamese families: family of uncle Hùng, family of uncle Khánh, and family of brother Ngô and sister Thu.

I express my appreciation to my close relatives, my close friends, and my colleagues. Despite of being far from me, they are always giving me the constant support and consolation.

Last, but not the least, I wish to express my deep and heartfelt gratitude to my beloved parents, my family, especially to my beloved son. They have loved me unconditionally, have supported me continuously in any circumstances and have been a huge inspiration to my success. Viva la France, Ciao!

Vannes, April 11, 2019

DO Van-Cuong

Lời cảm ơn

Chuyến hành trình dài với Toán học và Văn hóa của tôi trên đất Pháp đã đi đến chặng cuối cùng. Tôi xin mượn hai câu thơ của nhà thơ Chế Lan Viên trong bài thơ **"Tiếng hát con tàu"** để nói hộ tâm trạng của mình trong những ngày cuối cùng ở nước Pháp:

Khi ta ở chỉ là nơi đất ở,
Khi ta đi đất bỗng hóa tâm hồn.

Trong suốt bốn năm và ba tháng qua, tôi đã tận hưởng một cuộc sống đầy màu sắc tại Vannes, thành phố biển xinh đẹp bên bờ vịnh Morbihan của đất nước Pháp tươi đẹp. Tôi may mắn gặp gỡ rất nhiều người trong hoặc ngoài khoa Toán LMBA (Laboratoire de Mathématiques de Bretagne Atlantiques). Từ tận đáy lòng, tôi muốn nhắn gửi lời cảm tạ chân thành nhất tới tất cả mọi người vì những tình cảm và lòng tử tế của mọi người dành cho tôi. Dù tôi không thể kể hết tên tất cả mọi người ở đây nhưng tôi sẽ không bao giờ quên bất cứ ai.

Vào đầu năm 2015, tôi bắt đầu làm Nghiên cứu sinh dưới sự hướng dẫn của thầy Evans Gouno với một dự án về lập mô hình thống kê cho dữ liệu giông bão. Dự án này đòi hỏi nhiều kiến thức và kỹ thuật để phân tích và suy luận về dữ liệu thực. Thầy đề xuất tôi nghiên cứu một số quá trình ngẫu nhiên như quá trình ngẫu nhiên tự diễn biến (SEPP), quá trình ngẫu nhiên có cường độ luật mũ (PLP). Khi mới bắt đầu, tôi không thành thạo nhiều phương pháp và kỹ thuật thống kê vì trước đó tôi học thạc sỹ chuyên ngành xác suất. Tôi đã nhận ra rằng việc chuyển từ lý thuyết xác suất đến phân tích dữ liệu thực tiễn không phải là một nhiệm vụ dễ dàng.

Trong quá trình làm luận án, tôi đã học được rất nhiều kiến thức và kinh nghiệm về thống kê nhờ vào sự giúp đỡ của Thầy. Tôi thực sự biết ơn Thầy vì sự hướng dẫn tận tình, tỉ mỉ, hiệu quả và đặc biệt là Thầy luôn tôn trọng học trò. Thầy là người giải ảo cho tôi nhiều điều mà tôi đã từng ngộ nhận về xác suất và thống kê. Thầy dạy cho tôi biết khiêm tốn và kiên nhẫn học hỏi những điều cơ bản nhất trước khi muốn làm điều gì đó lớn hơn. Có những điều tưởng như đơn giản nhưng nếu cẩn thận suy xét thì hóa ra chúng không đơn giản chút nào. Sẽ không bao giờ tôi quên thời gian chúng tôi cùng nhau thảo luận về những ý tưởng mới, chuẩn bị các báo cáo hội thảo, viết và sửa lỗi các bài báo. Thầy khiến tôi cảm thấy thoải mái, tiếp thêm cho tôi động lực và sự tự tin trong công việc nghiên cứu đầy khó khăn, thử thách.

Với tấm lòng chân thành, tôi xin gửi lời cảm ơn tới hai giáo sư làm phản biện cho luận án của tôi: giáo sư Mitra Fouladirah và giáo sư Jean-Yves Dauxois. Tôi cũng bày tỏ lòng cảm ơn đến các thành viên khác trong hội đồng chấm luận án cho tôi: giáo sư Jean Vaillant, giáo sư Sophie Mercier và giáo sư Gilles Durrieu.

Khi mới đến Pháp, tôi hoàn toàn không biết tiếng Pháp nên tôi gặp rất nhiều khó khăn khi giao tiếp với người Pháp. May mắn thay, các thành viên của khoa Toán LMBA đã luôn sẵn sàng giúp tôi vượt qua những khó khăn trong cuộc sống hàng ngày. Ronan, Jonathan, Eric đã dạy tôi những từ tiếng Pháp đầu tiên và giúp tôi làm các thủ tục hành chính bằng tiếng Pháp. Sự giúp đỡ của các bạn trong những ngày đầu tiên tôi ở Pháp là điều thật đáng trân trọng và tôi không thể nào quên và tôi biết ơn các bạn về sự giúp đỡ quý báu đó.

Ronan "từ điển", tôi tự hỏi liệu có tồn tại điều gì mà cậu không thể trả lời được? Sự siêng năng và sự đều đặn của cậu trong công việc hàng ngày đã gây ấn tượng mạnh với tôi. Tôi đã học được nhiều từ thái độ chuyên nghiệp từ cậu. Cậu đã truyền cảm hứng cho tôi để tôi trở thành một phiên bản tốt hơn chính mình! Jonathan, cậu luôn tốt với tôi và kiên nhẫn giải thích cho tôi nhiều câu hỏi. Eric, cậu đã lái xe chở tôi đi lùng sục tất cả các cửa hàng đồ cũ để tìm chiếc xe đạp bị mất cắp của tôi. Lúc trên đường đến cửa hàng cuối cùng, cậu bảo hãy tập trung tư tưởng và tin là sẽ tìm thấy nó và điều đó quả thực đã xảy ra như một phép màu! Kévin, cậu đã dạy cho tôi cách chơi bóng bàn đúng kỹ thuật và dạy tôi chơi môn cờ vây Trung Quốc. Môn thể thao nào cậu cũng chơi rất cừ! Tarik, cậu đã cho tôi thấy sự kiên nhẫn đến cùng và không bao giờ bỏ cuộc trong khi làm luận án. Hui, chúng ta đã tận hưởng những ngày nghỉ cuối tuần cùng đến trường để thảo luận về nhiều chủ đề thú vị và thực hành tiếng Pháp. Rabih, Erwan, Hélène, Anne-Charlotte, Jamila, Pathé, và Hiền, các bạn luôn thân thiện với tôi.

Tôi muốn nói lời cảm ơn đến tất cả thành viên cơ hữu của khoa Toán LMBA: tiến sỹ Bertrand Patureau, tiến sỹ Sylvain Barré, tiến sỹ Salim Lardjane, giáo sư Quansheng Liu, giáo sư Gilles Durrieu, giáo sư Emmanuel Frénod, giáo sư Ion Grama, giáo sư Gaël Meigniez và giáo sư Jacques Froment, vì lòng hiếu khách với nghiên cứu sinh nước ngoài. Bertrand thực sự là một người bạn lớn vì anh luôn thân thiện và sẵn lòng giúp đỡ tôi. Anh còn mời chúng tôi đến nhà một vài lần để cùng nấu ăn với gia đình anh. Giáo sư Liu, giám đốc Khoa, là người luôn lắng nghe và thấu hiểu nhu cầu của các Nghiên cứu sinh nước ngoài. Ông đã đề xuất tổ chức các khóa học tiếng Pháp cho chúng tôi, nhờ đó chúng tôi đã cải thiện nhiều trình độ tiếng Pháp.

Tôi chân thành cảm ơn cô Françoise đã tình nguyện đến khoa Toán dạy chúng tôi tiếng Pháp thứ Sáu hàng tuần. Cô là một giáo viên tuyệt vời và luôn tử tế với học trò của mình. Tôi cũng sẽ nhớ cô Martine, người dạy tôi ở lớp tiếng Pháp buổi tối trong suốt bốn năm. Cô luôn khuyến khính tinh thần học tập của tôi làm tôi yêu thích tiếng Pháp hơn. Và chắc chắn tôi không thể quên cô Jessica, người đã dạy nhóm Nghiên cứu sinh chúng tôi những bài học tiếng Pháp cho người mới bắt đầu.

Trong thời gian sống ở Vannes, tôi cũng đã may mắn gặp được nhiều người Pháp tốt bụng bên ngoài khoa Toán. Chính họ đã dạy tôi rất nhiều về văn hóa Pháp. Nhờ vậy mà sau bốn năm tôi đã tiếp thụ được không ít nền văn hóa vĩ đại từ đất nước của những Molière, Hugo và Balzac.

Dominique và gia đình bà ấy đã giúp tôi rất nhiều từ trước tới giờ. Không có bà, tôi đã không thể có nhiều trải nghiệm như thế về những ngôi làng, những thành phố Pháp và những cảnh đẹp của vùng Bretagne nước Pháp. Tôi mang ơn bà rất nhiều. Gia đình ông Jean-François, bà Sandrine và cô Marie luôn hiếu khách và hào phóng giúp đỡ tôi. Lòng tử tế của họ làm tôi thực sự cảm động. Ông bà Lucien đã nhiệt tình mời tôi và bố mẹ tôi đến nhà họ ăn trưa. Ông André và bà Thérèse đã đón tiếp trọng thị tôi và bố mẹ tôi trong ngôi nhà xinh đẹp của họ. Chúng tôi đã có rất nhiều kỷ niệm đẹp khi đi du lịch cùng nhau ở Pháp và ở Việt Nam. Gia đình ông Alain và bà Annick đã mời tôi và bố mẹ tôi dùng bữa tối tại nhà họ. Tôi vẫn còn nhớ lần được đi cùng ông và gia đình tìm bắt hải sản trên bãi biển Sarzeau và chúng ta đã thu được một mớ kha khá! Bà Danielle đã nhiều lần đến đón tôi đi xem phim. Bà còn tốt bụng mời tôi và bố mẹ đến nhà uống trà và tặng bố mẹ tôi một món quà nhỏ đem về Việt Nam. Tôi chắc chắn sẽ gửi cho bà một tấm bưu thiếp từ Hà Nội như đã hứa. Bà hàng xóm tốt bụng và thân thiện Jaqueline luôn đón chào tôi nồng nhiệt và đã thết đãi tôi những bữa tối Pháp ngon tuyệt.

Cũng đến từ bên ngoài nước Pháp, tôi đã gặp gỡ rất nhiều bạn bè từ khắp năm châu. Điều đó giúp tôi hiểu thêm nhiều về một thế giới đa dạng và đầy màu sắc. Cô Widya đến từ Indonesia, cậu Nikita đến từ Đức, cô Vasiliki đến từ Hy Lạp, ngoài ra còn rất nhiều người bạn khác đến từ châu Mỹ, châu Phi.

Chandru, Shareen và Raphael, những người bạn Ấn Độ đã giới thiệu cho tôi và bố mẹ tôi những đặc sản đến từ quê hương của Đức Phật.

Rexie, cậu đã nấu cùng tôi nhiều bữa tối với các món Philippines rất giống món ăn Việt nhưng cách chế biến rất khác biệt. Cậu đã dạy tôi rất nhiều điều mới, đã giới thiệu với tôi Sách Mặc Môn và về LDS. Tôi sẽ nhớ mãi tấm lòng rộng mở và chân thành của các tín hữu LDS khác dành cho tôi.

Cộng đồng người Việt Nam ở Vannes đã chia sẻ với tôi cuộc sống xa nhà, tôi muôn vàn cảm cảm kích những tình cảm đó. Anh Phú đã cho tôi thấy sự kiên nhẫn và điềm tĩnh tuyệt vời trong mọi tình huống. Cậu bạn Du luôn cho thấy sự thông minh và hiệu quả tuyệt vời trong công việc. Cậu bạn Thông luôn thể hiện là một người hiểu biết rất nhiều điều hữu ích về mọi khía cạnh của cuộc sống ở Pháp. Cô bạn Thủy luôn cho thấy sự hồn nhiên không bao giờ cạn và lòng tốt vô hạn với tất cả mọi người. Vợ chồng Tân-Trà là một cặp trai tài gái sắc nhưng lại cực kỳ thông minh và khiêm tốn. Các em thật đáng để anh ngưỡng mộ. Vợ chồng Phúc-Hạnh là một cặp đôi hoàn hảo với anh chồng giỏi giang và cô vợ đảm đang. Anh rất khâm phục những thành tích của Phúc trong nghiên cứu khoa học và hợp tác quốc tế. Anh bạn Quang là một vua đầu bếp thực sự với những hiểu biết không bao giờ làm tôi hết ngạc nhiên về các món ăn và nghệ thuật nấu ăn. Cảm ơn Quang về những món ăn ngon mà cậu đã nấu cho tôi.

Tôi gửi lời cảm ơn chân thành tới một số gia đình người Pháp gốc Việt đã chào đón tôi nồng nhiệt: gia đình bác Hùng, gia đình bác Khánh, gia đình anh Ngô và chị Thu.

Tôi biết ơn và trân trọng những tình cảm từ người thân, bạn bè và đồng nghiệp của tôi. Họ ở cách xa tôi nhưng luôn dành cho tôi rất nhiều sự ủng hộ và niềm an ủi lớn lao.

Cuối cùng, tôi dành tất cả tình yêu và lòng biết ơn cho bố mẹ và gia đình tôi. Chính nhờ tình yêu vô điều kiện trong mọi hoàn cảnh của họ mà tôi mới có được ngày hôm nay.

Xin gửi lời chào tạm biệt nước Pháp và hẹn ngày trở lại. Con trai tôi sẽ có ngày đến nước Pháp học tập theo bước chân bố từng đi qua.

Vannes, ngày 11 tháng Tư năm 2019

Đỗ Văn Cường

Contents

Analyse statistique de processus stochastiques : application à des données d'orages

Résumé

Les travaux présentés dans cette thèse concernent l'analyse statistique de cas particuliers du processus de Cox.

Dans une première partie, nous proposons une synthèse des résultats existants sur le processus power-law (processus d'intensité puissance), synthèse qui ne peut être exhaustive étant donné la popularité de ce processus. Nous considérons une approche bayésienne pour l'inférence des paramètres de ce processus qui nous conduit à introduire et à étudier en détails une distribution que nous appelons *loi H-B*. Cette loi est une loi conjuguée. Nous proposons des stratégies d'élicitation des hyperparamètres et étudions le comportement des estimateurs de Bayes par des simulations.

Dans un deuxième temps, nous étendons ces travaux au cas du processus d'intensité exponentielle (exponential-law process). De la même façon, nous définissons et étudions une loi conjuguée pour l'analyse bayésienne de ce dernier.

Dans la dernière partie de la thèse, nous considérons un processus auto-excité qui intègre une covariable. Ce travail est motivé, à l'origine, par un problème de fiabilité qui concerne des données de défaillances de matériels exposés à des environnements sévères. Les résultats sont illustrés par des applications sur des données d'activités orageuses collectées dans deux départements français. Enfin, nous donnons quelques directions de travail et perspectives pour de futurs développements de l'ensemble de nos travaux.

Mots Clés : Processus power-law, processus d'intensité exponentielle, processus auto-excité, covariable, maximum de vraisemblance, estimateur de Bayes.

Inference for some stochastic processes with aplication on Thunderstorm data

Abstract

The work presented in this PhD dissertation concerns the statistical analysis of some particular cases of the Cox process.

In the first part, we introduce some important statistical concepts and notions as well as techniques for stochastic point processes.

In the second part, we study the power-law process (PLP). Since the literature for the PLP is abundant, we suggest a state-of-art for this process. We consider the classical approach and recall some important properties of the maximum likelihood estimators for the PLP's parameters. Then we investigate the Bayesian approach with non-informative priors. We construct conjugate priors and define a family of distributions that we name H-B distribution as the natural conjugate priors for the PLP. Bayesian computations with H-B prior are conducted via a simulation study and an application on real data.

In the third part, we study the exponential-law process (ELP). We review the maximum likelihood techniques. For Bayesian analysis of the ELP, we introduce the modified-Gumbel distributions and Gamma-modified-Gumbel distributions as conjugate priors. We conduct a simulation study to compare maximum likelihood estimates and Bayesian estimates.

In the fourth part, we investigate self-exciting point processes. We integrate a power-law covariate model to one of these processes. A maximum likelihood procedure for the model is proposed and the Bayesian approach is suggested.

The last part of the thesis is devoted to an application on thunderstorm data collected in two French regions. We consider a strategy to define a thunderstorm as a temporal process associated with the electrical charges in a given location. Some selected thunderstorms are analyzed. We propose a reduced maximum likelihood procedure to estimate the parameters of the Hawkes process. We then fit some thunderstorms to the power-law covariate self-exciting point process taking into account the associated electrical charges.

In the conclusion, we give some perspectives for further work.

Keywords: Power-law process, exponential-law process, self-exciting point process, power-law covariate model, maximum likelihood estimation, Bayes estimation.

Résumé

Les travaux présentés dans cette thèse concernent l'analyse statistique de processus stochastiques. Nous nous intéressons à l'estimation des paramètres de l'intensité du processus power-law (*processus d'intensité puissance*), du processus exponential-law (*processus d'intensité exponentielle*) et d'un processus auto-excité à covariables.

La motivation première du travail est l'analyse de données d'orages qui ont été collectées dans deux départements français : l'Ardèche et la Drôme.

L'objectif est d'obtenir une modélisation de l'activité orageuse ; cette modélisation ayant pour finalité une meilleure connaissance de l'effet de variables liées à l'environnement sur la propension à la défaillance de matériels. Il s'agit donc d'un problème de fiabilité.

La nature des données dont nous disposons, nous a conduit à nous orienter vers une modélisation *processus de comptage*. La particularité de notre travail est la prise en compte de variables caractérisant l'environnement. Nous avons proposé un modèle où l'expression de l'intensité dépend de variables exogènes. Ce modèle s'apparente au modèle de Cox mais ici l'expression de l'intensité ne dépend pas seulement du temps mais également du nombre de sauts survenu par le passé. Nous envisageons l'analyse de ce processus du point de vue classique (maximum de vraisemblance) mais également du point de vue bayésien.

La littérature concernant les processus stochastiques est très abondantes. Les approches sont multiples. On rencontres une myriades de notations, définitions et résultats, déclinés sous de nombreuses formes. Ceci nous a conduit à rédiger un chapitre préliminaire afin de fixer l'approche et les notations que nous adopterons tout au long de la thèse.

Avant de construire le modèle auto-excité à covariables pour les données d'orages, nous avons choisi d'explorer les modèles plus classiques que sont le processus power-law et le processus exponential-law. Pour chacun de ces modèles, nous proposons un état de l'art avant d'apporter une contribution en définissant des lois qui se trouvent être des lois a priori conjuguées pour l'estimation bayésienne des paramètres de ces modèles. Une attention particulière est portée à la question de l'élicitation des hyperparamètres.

Le Processus Power-Law

Le processus power-law (PLP) est un processus de Poisson non-homogène dont l'intensité est de la forme : $\lambda(t) = \dfrac{\beta}{\alpha^\beta} t^{\beta-1}$. Il est bien connu et « intensivement » utilisé dans le domaine de la fiabilité. L'intérêt de ce processus est qu'il permet de couvrir de nombreuses situations relatives à la concentration, à l'accumulation de l'événement d'intérêt que le processus est supposé représenter. Nous en présentons un état de l'art. Un autre intérêt du processus power-law est que les estimateurs du maximum de vraisemblance existent sous forme explicites. Nous rappelons les principaux résultats utiles pour l'inférence.

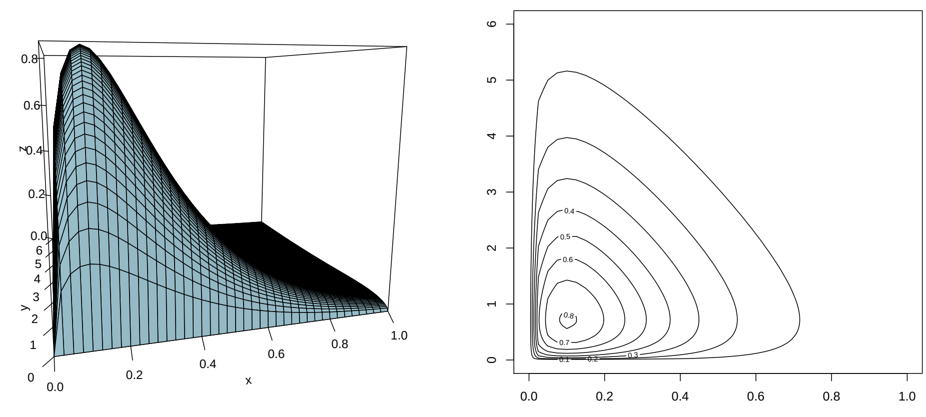
Nous considérons alors une approche bayésienne et définissons la loi H-B. Cette loi caractérisée par 5 paramètres (a,b,c,d,m) tous positifs et tels que $b < c^a$, est une loi bivariée sur $\mathbb{R}^+ \times \mathbb{R}^+$ dont la densité est :

$$f_{X,Y}(x,y) = K\, x^{m-1} y^{a-1} b^y \exp\left\{-dxc^y\right\},$$

où $K = d^m[\log(c^m/b)]^a / \Gamma(m)\Gamma(a)$.

La figure 1 est la représentation de la densité d'une loi H-B de paramètres $(1.5, 5, 0.5, 1, 1.5)$ et de la courbe de niveau associée. Cette loi se trouve être une loi conjuguée pour l'estimation des paramètres de l'intensité d'un processus power-law.

FIGURE 1 : *Densité de la loi de probabilité H-B de paramètres $a = 1.5$, $b = 5$, $c = 0.5$, $d = 1$ et $m = 1.5$*



Une attention particulière est portée à la question de l'élicitation des hyperparamètres et différentes stratégies sont proposées. Ces travaux ont fait l'objet d'une communication et d'une publication dans les proceedings d'une conférence internationale à Ho Chi Ming City en mai 2016 (The 1st Conference on Applied Mathematics in Engineering and Reliability (ICAMER 2016)).

Le Processus exponential-law

Ce que nous avons choisi d'appeler le *processus exponential-law* est une processus de Poisson dont l'intensité est de forme exponentielle et a pour expression $\lambda(t) = \alpha e^{\beta t}$, $\alpha > 0$, $\beta \in \mathbb{R}$. Bien qu'il soit moins fréquemment rencontré que le processus power-law dans la littérature, il permet également de couvrir différents cas de comportement du processus suivant les valeurs des paramètres α et β.

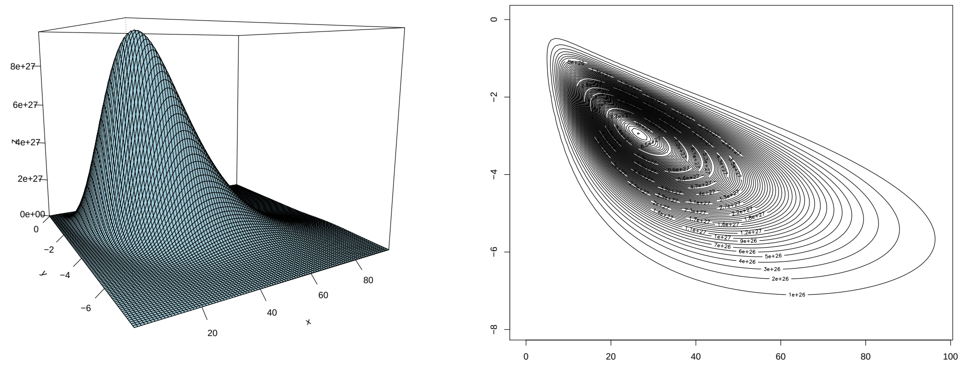
Nous avons comme pour le PLP, mené une étude détaillée de ce processus. Pour procéder à son analyse bayésienne nous avons introduit une distribution particulière que nous avons appelée : *loi Gamma-Gumbel modifiée*. Cette loi bivariée possède 5 paramètres (a, b, c, d, m) tous positifs et tels que : $a \geqslant 2$, $c \leqslant b \leqslant c(a-1)$ et $m \in \mathbb{N}$. Sa densité a pour expression :

$$f_{X,Y}(x,y) = K\ x^{a-1} y^m \exp\left\{by - dx\left(e^{cy}-1\right)/y\right\}.$$

où K est une constante de normalisation que nous avons calculée.

La figure 2 donne la représentation d'une loi Gamma-Gumbel modifiée pour $a = 10$, $b = 10$, $c = 2$, $d = 1$ et $m = 0$ et de la courbe de niveau associée.

FIGURE 2 : *Densité de la loi de probabilité Gamma-Gumbel modifiée de paramètres* $a = 10$, $b = 10$, $c = 2$, $d = 1$ *et* $m = 0$



Notre travail étend les résultats obtenus par Huang & Bier [28]. Comme pour le PLP, nous envisageons le problème de l'elicitation des hyperparamètres. Le comportement des estimateurs de Bayes est étudié à travers des simulations.

Les Processus auto-excités

Nous avons construit un *modèle auto-excité à covariables* pour les données d'orages en nous appuyant sur les techniques utilisées pour modéliser l'activité sismique.

En effet, dans ce domaine, de nombreux modèles ont été proposés, modèles reposant sur des considérations géo-physiques et sur les processus ponctuels auto-excités.

Les processus auto-excités sont des processus stochastiques dont l'intensité ne dépend pas uniquement du temps mais également du nombre et de la date des événements passés. Ils sont très fréquemment utilisés pour analyser les données de tremblements de terre. En effet, un tremblement de terre est caractérisé par des secousses majeures (ou principales) de forte intensité suivies de répliques mineures (ou secondaires) d'intensité plus faible et décroissante dans le temps. Il nous a semblé intuitivement que les phénomènes d'orages s'apparentaient aux phénomènes de tremblements de terre. A des impacts de foudre de charges électriques importantes succèdent des épisodes d'impacts de moindres charges ; la charge électrique jouant le rôle de la magnitude de d'une secousse.

Nous avons donc considéré un processus auto-excité dont l'intensité est de la forme :

$$\lambda^*(t) = \mu + \alpha \sum_{t_i<t} \left(\frac{z_i}{z_0}\right)^\eta e^{-\beta(t-t_i)} \tag{1}$$

où z_i est la charge du $i^{\text{ème}}$ impact. En analysant les données disponibles, nous avons constaté que la charge électrique était distribuée suivant une loi log-normale. Nous proposons des procédures d'estimation classiques et bayésiennes pour les paramètres du modèle (1) qui est donc l'intensité de ce que avons appelé un *processus auto-excité à covariables*.

Application : Données d'Orages

Dans le chapitre 6, nous décrivons les données d'orages et proposons une définition (statistique) de l'orage. Nous obtenons des représentations des impacts correspondants. Les trois figures suivantes donnent une représentation spatiale (5) , une représention de l'intensité empirique cumulée (5) et une représentation de la charge des impacts (5) pour un exemple d'orage qui a eu lieu aux alentours de Valence, le 10 avril 2008. Cette orage a duré 37 minutes et 160 impacts de foudre se sont produits.

Nous appliquons alors les méthodes proposées au chapitre précédent pour analyser une sélection d'orages extraite par notre procédure de la base de données.

FIGURE 3 : *Représentation spatiale d'un orage (Valence, 10 avril 2008)*

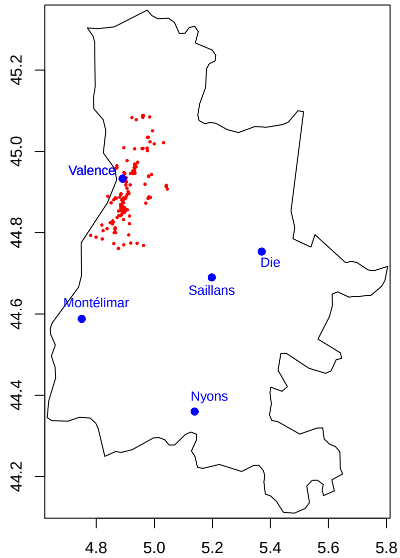


FIGURE 4 : *Représentation de l'intensité empirique cumulée pour un orage (Valence, 10 avril 2008)*

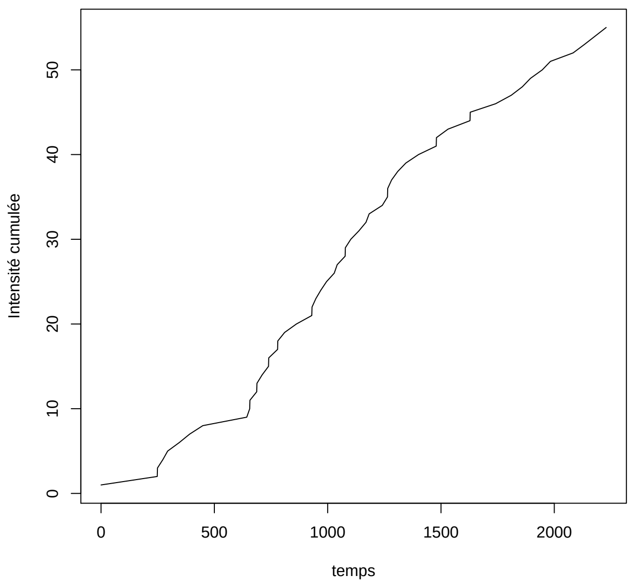
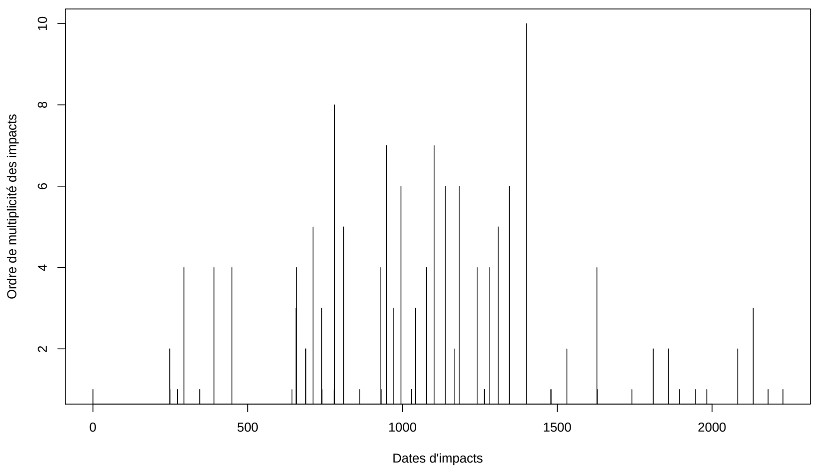


FIGURE 5 : *Représentation de la charge des impacts d'un orage (Valence, 10 avril 2008)*



Conclusion et Perspectives

Dans nos travaux, nous nous sommes essentiellement intéressé à une modélisation temporelle, la prise en compte de l'aspect spatial s'est effectuée de manière empirique et déterministe. Une direction de travail que nous envisageons est la prise en compte de l'aspect spatial. Il s'agira alors d'étudier des processus auto-excités spatio-temporels à covariables.

A notre connaissance, il existe peu de travaux sur les techniques de tests d'hypothèses non paramétrique dans un cadre bayésienne pour les PAE. Cette direction de recherche est à envisager.

Le problème que nous avons abordé dans la thèse est très riche. De nombreux modèles sont envisageables et à imaginer pour les données d'orages. Le développement d'un utilitaire afin d'automatiser l'extraction des séquences d'impacts qui caractérisent les orages est un de nos projets. Il pourrait permettre le traitement spatio-temporelle mais également la classification des orages.

De nombreuses questions restent ouvertes et de nombreux points abordés dans la thèse sont l'objet de projet de publication.

A l'heure où l'évolution de climat est devenue une question importante, notre travail sur la modélisation de l'activité orageuse pourrait contribuer à une meilleure compréhension de ce phénomène.

Chapter 1

General Introduction

Context of the thesis

"Tous les ans, environ 500 000 impacts de foudre sont enregistrés en France."

Thunderstorms are phenomena that happen in the Earth's atmosphere. A thunderstorm creates a series of impacts nearby a place and within a period of time. In each thunderstorm, a lightning strike may trigger the preferable condition of the atmosphere then more strikes come after and nearby. That is to say, impacts of a thunderstorm are grouped in time and clustered in space.

Considering the thunderstorm data, we are looking for statistical models to fit the data. The data comprises of lightning impacts with dates, times, longitudes, latitudes, electrical charges, etc. Self-exciting point processes (SEPP) appear as good candidates since they have been widely used to model clustering point patterns.

While investigating deeply the data, we also find out that, given a fixed amplitude-threshold of electrical charges, thunderstorm impacts may occur more or less frequently over time. This kind of point pattern may follow the power-law process (PLP) or exponential-law process (ELP) that have been used to model reliability of repairable systems.

At first sight the link between those three processes is not obvious. The PLP and the ELP belong to the class of nonhomogeneous Poisson processes while the SEPP belongs to the family of doubly stochastic Poisson processes. The PLP has intensity function that depends on time according a power-law and for the ELP the dependency is the exponential-law. These intensity functions are independent to the previous history of the processes. The SEPP has conditional intensity function depending not only on time but also on the previous history. Therefore, the conditional intensity function is a random quantity itself. This form of conditional intensity function captures the feature "self-excited" of the process.

The PLP has been used intensively in reliability to model repairable systems and also finds its application in many other domains. The convenient form of its intensity function makes it popular in the literature. As mentioned by Rigdon and Basu [51]:

'it is also possible that PLP could also be used to model [...] the occurrences of earthquakes of a given magnitude, or other events that occur at random points in time.'

The ELP can be also used to model repairable systems but with a faster grow rate of reliability than the PLP. This process has form of intensity as simple as of the PLP but the maximum likelihood estimators can not be obtained analytically. By using numerical techniques to maximize the likelihood for the ELP, we get some clues for the same techniques employed to the SEPP.

The SEPP has been used to model many phenomena such as earthquakes (Ogata [43], Vere-Jones [61]), seismic data, social media analytics (Zadeh & Sharda [62]), ultra-high frequency financial data (Chen & Hall [7]), conversation event sequences (Masuda et al. [37]), crime (Molher et al. [40]), neural activities (Kazemipour et al. [29]). However, we have not found any application of self-exciting point process to thunderstorms.

We choose to present the three processes in this thesis. We will give a brief introduction to each of the process in the sequel.

The power-law process

The PLP was first introduced by Duane in 1964 ([17]). In the literature, there exists many terminologies for this process: Duane process, Weillbull process, AMSAA process. In fact, it is a nonhomogeneous Poisson process with power-law intensity function parametrized by two parameters. We use the term *p*ower-law process as suggested by Guida et al. ([6]). Thank to the power-law form of the intensity, the compensator is also in power-law form. That makes possible to conduct a simple graphical test such as Duane plot ([17]) and a goodness-of-fit test (Gaudoin et al. [20]). The likelihood of the

PLP is tractable and gives closed-forms estimators (MLEs). The MLEs can be transformed to pivotal quantities that facilitate the construction of confidence intervals for the parameters. In practice, two schemes of observation can be considered: *event truncation* and *time truncation*. If we set a fixed number of event beforehand, say n, and stop our observation at the date of occurrence of the $n-$th event then the data are *event truncated*. For time truncation, we fix the time C, end of the observation. In the first scheme, the ending time is a random variable while the number of events is fixed. In the second scheme, the ending time is no more random but the number of events is. Therefore inference on maximum likelihood estimators is different but it will not be the case for Bayesian inference. In many papers, authors focus on the event truncation case. We treat in details the time truncation case for maximum likelihood estimators of the PLP.

Bayesian approach for the PLP has been considered by many authors. Guida et al. (1989, [23]), Calibria et al. (1990, [6]), Bar-lev et al. (1992, [3]), Huang and Bier (1998, [27]) and Olivera et al. (2012, [45]) are amongst those who contributed to the Bayesian inference for the PLP. They consider many different parametrizations for the intensity function of the PLP. Guida et al. use the form $\lambda(t) = (\beta/\alpha)(t/\alpha)^{\beta-1}$. Huang and Bier reparametrize $\mu = 1/\alpha^{\beta}$ to get the form $\lambda(t) = \mu\beta t^{\beta-1}$. Olivera et al. consider time truncation scheme of observation in the time window $[0, C]$ and reparametrize $\eta = \Lambda(C) = (C/\alpha)^{\beta}$ so the intensity is $\lambda(t) = \mu\beta t^{\beta-1}$. Different choices of priors are proposed corresponding to the parametrization. The table 1.1 summarizes some of the prior choices for Bayesian analysis of the PLP.

Table 1.1: Prior choices for the Bayesian analysis of the PLP

Reference	Prior
Guida et al. [23] and Calabria et al. [6]	Noninformative joint prior for (α, β) Noninformative for α, uniform for β gamma for α given β, uniform for β
Bar-Lev et al. [3]	general noninformative joint prior for (α, β)
Kuo & Yang [31]	any distribution for α, gamma for β (independent)
Huang & Bier [27]	natural conjugate prior for (μ, β)
Sen [54]	Noninformative joint prior for (α, β) gamma for α given β, uniform for β
Oliveira et al. [45]	independent conjugate prior for (η, β)

Conjugate prior is a very convenient choice of prior for Bayesian analysis since the prior and the posterior belong to the same family distributions. We seek for a conjugate prior for the process since it lessens the computation of posterior. Huang and Bier propose a prior that they call the natural conjugate prior for the PLP. In fact, their prior is not exactly a conjugate for the PLP since the posterior distribution is different from prior distribution and they need numerical method to analyze the posterior. In our work, we start from Jeffreys' rule for noninformative prior then consider a reparametrization that leads to an independent conjugate prior. The posterior distribution obtained from those priors can be decomposed into a marginal gamma distribution and a conditional gamma distribution. That type of combination leads us to consider a joint conjugate prior for the PLP. We obtain a bivariate distribution that is a conjugate prior for the PLP. Huang and Bier ([27]) introduce the same idea but they still need numerical method to approximate the posterior since the posterior does not belong to the same family as the prior. We name that natural conjugate prior the H-B distribution. This bivariate distribution is a product of a marginal gamma distribution and a conditional gamma distribution. It is a uni-modal distribution. The expectations, variances of each components can be obtained explicitly as well as the covariance between the two components. Those convenient closed-forms of moments make it easy to integrate prior information for prior elicitation. We also suggest many scenarios of prior information and the associated prior elicitation strategies for the hyperparameters.

The simulation algorithm for the PLP is available for a comparison between Bayesian estimates and maximum likelihood estimates.

The exponential-law process

As the PLP, the ELP has been introduced in reliability analysis. It is normally referred to as a nonhomogeneous Poisson process with exponential law intensity function (Huang & Bier [28]). We use the term *exponential-law process* mimicking the name for the *power-law process*. The ELP, despite the simple form of intensity function, is less popular than the PLP. The literature for this process is limited, for example (Huang & Bier [28]). It is considered as an alternative model for reliability when the grow rate changes very fast according an exponential-law. The compensator (integrated intensity) is not exponential and it is not possible to obtain a simple graphical test or goodness-of-fit test for the adequacy of the process. The likelihood is also tractable but it requires numerical methods to get maximum likelihood estimates. We propose a procedure to maximize the likelihood of the ELP that is original.

For Bayesian inference on the ELP, we follow the same roadmap as for the PLP. First, we consider the noninformative prior and then the independent conjugate prior by a reparametrization. We obtain a joint conjugate prior for the ELP as a combination of gamma distributions and what we name *Modified-Gumbel* distribution. We name that natural conjugate prior the *Gamma-Modified-Gumbel* distribution. The expectation, variance of each component of this bivaritate distribution can not be obtained in closed-forms but require numerical method. That leads to the fact that any prior elicitation requires a trials and errors procedure. Then we introduce some elicitation strategies and conduct a simulation study to compare Bayesian estimates and maximum likelihood estimates.

To compare the PLP and the ELP, we give an application on real data.

Self-exciting point processes

Self-exciting point processes (SEPP) are useful statistical models for point patterns that have a temporal clustering feature. Since the seminal paper of Hawkes ([24]), SEPP and their extensions have found applications in a wide range of areas, such as earthquake occurrence modeling (Ogata [43]) and prediction (Vere-Jones [61]), neuron firing process modeling (Chornoboy et al. [8]), triggered optical emission modeling (Teich & Saleh [59]), credit rating transition modeling (Koopmana et al. [30]), general ultra-high frequency financial data modeling (Chen & Hall [7]), and social network interaction modeling (Crane & Sornette [12]).

An important problem in applications of point processes is estimating model parameters. Both maximum likelihood inference and Bayesian inference for SEPP have been proposed. Some works on estimation for SEPP and associated asymptotic theory can be found in the literature. Ogata [43] establishes consistency and asymptotic normality of maximum likelihood estimators under stationarity and ergodicity conditions. Chornoboy et al. [8] derives consistency and asymptotic normality of maximum likelihood estimators for the multivariate SEPP under regularity conditions on the excitation components of the intensity processes. They also proposes an expectation-maximization procedure to calculate maximum likelihood estimators, and establishes the convergence of the procedure. Rathbun [49] studies asymptotic properties of maximum likelihood estimators (MLEs) for spatial-temporal SEPPs under stationarity conditions.

We introduce the power-law covariate self-exciting point process (PLC-SEPP) that takes into account covariates associated with jumps of the process (here, electrical charges of thunderstorm impacts). Writing the likelihood for such a process, the MLE can not be obtained in closed-forms so it requires numerical techniques. We present a reduced maximum likelihood procedure for the Hawkes process and the power-law covariate self-exciting point process. Efficiency of the procedures is studied through simulations.

Objectives

The objective of the thesis is to develop some statistical models to fit the thunderstorm data and propose some procedures to make inference on the parameters of the models using maximum likelihood approach and Bayesian approach.

Main results

We introduce and study conjugate priors for Bayesian analysis of the PLP and the ELP. The conjugate prior for the PLP is H-B distribution while the conjugate prior for the ELP is G-M-G distribution. The two distributions are investigated in details with some practical elicitation strategies to obtain values for their parameters from prior information. Relying on the abundant literature on the PLP, we suggest a state-of-art for this process related to many aspects: interevents distribution, maximum likelihood estimators, data truncation, Bayesian prior choices, parametrizations, etc. By contrast, the literature on the ELP is very limited. We develop a maximum likelihood procedure for the process and then propose several Bayesian prior choices for inference. Our approaches have been investigated by simulation studies for the two processes and by application on real data.

We propose a reduced maximum likelihood procedure for the Hawkes process that reduced the objective optimizing function from three dimensions to two dimensions.

We propose the power-law covariate self-exciting point process (PLC-SEPP) that allows us to model thunderstorm with dates of jumps and the associated electrical charges. A maximum likelihood procedure for the PLC-SEPP is given and is tested by a simulation study. We also set the basis for the Bayesian approach for this model.

We introduce a method to define thunderstorms by specifying time-threshold and by localizing impacts on the map of the two regions. Some thunderstorms are selected to fit in the models.

Structure of the thesis

The dissertation is organized as follos.

In chapter 2, we review some basic statistical concepts and methods for stochastic processes. The conditional intensity function is our main tool to make inference.

In chapter 3 and chapter 4, we study the PLP and the ELP, respectively. We summarize the state-of-art in the literature for each process. We consider both maximum likelihood approach and Bayesian approach for the processes. Strategies to obtain the natural conjugate priors for the PLP and the ELP are presented from noninformative prior and independent conjugate prior to conjugate prior. We introduce conjugate priors for the PLP and the ELP and compare Bayesian estimates to maximum likelihood estimates by simulation study.

In chapter 5, we study the self-exciting point process (SEPP). We reconsider some classical models of the process and some well-known types of conditional intensity function. A reduced maximum likelihood procedure for Hawkes process is proposed. Then we propose the model that we name power-law covariate self-exciting point process.

In chapter 6, we analyze the thunderstorm data. We define a thunderstorm in term of a sequence of impacts that occur in a fixed spatial window and temporal window. Some typical thunderstorms are chosen to make inference applying Hawkes process and the power-law covariate self-exciting point process.

In chapter 7, we present the general conclusion of our works. Some perspectives are given for future research.

Some proofs and simulation algorithms can be found in the appendices.

Chapter 2

Preliminary: Temporal Point Processes

In this chapter, we present some related concepts and notions as well as some techniques to deal with stochastic processes. Some important theorems and propositions are introduced and the readers can find the detail proofs in the appendices.

2.1 Introduction

A temporal point pattern is basically a list of jumps of events. Mathematically it is called a *stochastic point process*. Many real phenomena produce data that can be represented as a temporal point pattern, for instance failures of a repairable system, impacts of a thunderstorms to be listed a few. Usually complex mechanisms are behind these seemingly random times, for instance lightening strikes cause new lightening strikes by triggering the atmosphere. We do not know how many events will occur, or at what times they will occur. An essential tool for dealing with these mechanisms, for example in predicting future events, is a stochastic point process modeling the point pattern: a temporal point process. The term *point* is used since we may think of an event as being date and this date can represent as a point on the real line. Hence, the terms *point* and *event* or *jump* will be used interchangeably throughout this text. Often there is more information available associated with an event. This information is known as marks. The marks may be of separate interest or may simply be included to make a more realistic model of the event times. For example, for thunderstorm it is of practical relevance to know the position and the electrical charge associated with a lightening strike, not just its occurrence date. In the same time, we may think that the electrical charge of an impact influences the number and nature of future impacts.

2.2 Temporal Point Processes

There are many ways of treating temporal point processes. In this text, we will explore one approach based on the so-called conditional intensity function. To understand this concept, we first have to understand the concept of evolutionary. Usually we think of time as having an evolutionary character: what happens now may depend on what happened in the past, but not on what is going to happen in the future. This order of time is also a natural starting point for defining practically useful temporal point processes. Roughly speaking, we can define a point process by specifying a stochastic model for the time of the next event given we know all the times of previous events. The term evolutionary point process is used for processes defined in this way. The past in a point process is captured by the concept of the history of the process. If we consider the time t, then the history $\mathscr{H}_t$ is the list of all the dates of events $(t_1, t_2, \ldots, t_n)$ up to but not including time t. Note that we assume that we have a simple point process, i.e. a point process where no points coincide, such that the points can be strictly ordered in time.

2.2.1 Jumps and interevents

To specify a temporal point process we can use many different approaches. In this text, we will consider two of those approaches.

One rely on the distribution of the occurrence times of jumps or the time lengths between subsequent events. The second one rely on the number of events occurring in a given time-interval.

The lengths of the time intervals between subsequent events are called the *interevents*. We can define a temporal point process by specifying the distribution of these interevents.

In the case of i.i.d. exponential interevents with parameter λ, the process is the well-known homogeneous Poisson process (HPP) with intensity λ. The process is also obtained assuming that the number of events in a given interval of length t is a Poisson distribution with parameter λt. Another classical hypothesis for the interevents is a Weibull distribution. In this case, the inference for the process boils down to be the classical inference for i.i.d. Weibul r.v. Remark that in this case, the alternative definition through the numbers of events is not possible.

Now, suppose that given t_i, the date of the i^{th} jump , the time to the next jump is a Weibull distribution with parameters (α, β), the interevents are no longer independent and has distribution:

$$f(t \mid t_i, \ldots, t_1) = (\beta/\alpha)(t/\alpha)^{\beta-1} \exp\left\{-(t/\alpha)^{\beta} + (t_i/\alpha)^{\beta}\right\},$$

which is a left-truncated Weibull distribution with support $[t_i, +\infty[$.

This is an alternative definition of the power-law process (PLP) which is defined as we will see in the next chapter, through the distribution of the number of events in a given interval $[s,t]$ as a Poisson distribution with parameter $(t/\alpha)^\beta - (s/\alpha)^\beta$.

2.2.2 Conditional intensity function

The conditional intensity function is an intuitive and convenient way of specifying how the present depends on the past in an evolutionary point process. The behavior of a simple temporal point process is typically modeled by specifying its conditional intensity $\lambda^*(t)$, which represents the infinitesimal rate at which events are expected to occur around a particular time t, conditionally on the history of the point process prior to time t.

Definition 2.2.1. – *The conditional intensity associated with a temporal point process $N(t)$ may be defined via the limiting conditional probability*

$$\lambda^*(t) = \lim_{\Delta t \to 0} \frac{P[N(t, t+\Delta t) > 0|\mathscr{H}_t]}{\Delta t},$$

(provided this limit exists), where $\mathscr{H}_t$ is the history of the point process $N(t)$ over all times strictly prior to time t.

Some authors instead define $\lambda(t)$ as a limit of the conditional expectation ([54]):

$$\lambda^*(t) = \lim_{\Delta t \to 0} \frac{E[N(t, t+\Delta t] \mid \mathscr{H}_t]}{\Delta t},$$

since the two definitions are equivalent for orderly point processes.

Therefore assuming that there are no coincided events, one can also interpret the conditional intensity function as the mean number of events occurring in an infinitisimal length of time around t:

$$\begin{aligned}
\lambda^*(t)\Delta t &= P[\text{next event in } (t, t+\Delta t) \mid \mathscr{H}_t)] \\
&= E[N(t, t+\Delta t) \mid \mathscr{H}_t].
\end{aligned}$$

As all finite-dimensional distributions of $N(t)$ are uniquely determined by the conditional intensity, in modeling a temporal point process, it suffices to prescribe a model via its conditional intensity. $\lambda(t)$ may be estimated non-parametrically or via a parametric model. When $N(t)$ is a simple stochastic point process, however, $\lambda(t)$ is deterministic, i.e. $\lambda(t)$ depends only on t.

For a temporal point process originating at time 0, one may define the conditional compensator $\Lambda(t)$ as the integral of the conditional intensity from time 0 to time t.

$$\Lambda^*(t) = \int_0^t \lambda^*(u)du.$$

The compensator may equivalently be defined as the unique non-negative non-decreasing predictable process $\Lambda(t)$ such that $N[0,t) - \Lambda(t)$ is a martingale (see [38]).

Daley & Vere-Jones [16] establish that there is a one-to-one correspondence between regular point process on $\mathbb{R}^+$ and the family of conditional probability density functions $p_n(t \mid t_{n-1}, \dots, t_1)$ defined on $[t_{n-1}, +\infty[$, for $0 < t_1 < \dots < t_{n-1} < t$. We can consider the associated survivors functions

$$S_n(t \mid t_{n-1}, \dots, t_1) = 1 - \int_{t_{n-1}}^{t} p_n(u \mid t_{n-1}, \dots, t_1) du, \ (t > t_{n-1}) \tag{2.1}$$

and the hazard function:

$$h_n(t \mid t_{n-1}, \dots, t_1) = \frac{p_n(u \mid t_{n-1}, \dots, t_1)}{S_n(u \mid t_{n-1}, \dots, t_1)}. \tag{2.2}$$

The conditional intensity function is therefore the representative function $\lambda^*(.)$ defined piecewise by [1]:

$$\lambda^*(t) = \begin{cases} h_1(t), & \text{for } 0 < t \leqslant t_1, \\ h_n(t \mid t_{n-1}, \dots, t_1) & \text{for } t_{n-1} < t \leqslant t_n, \ n \geqslant 2. \end{cases}$$

This result is useful to simulate point processes.

2.2.3 Poisson processes

One of the most popular family of stochastic point processes is Poisson process, which is a simple point process $\{N(t), t \geq 0\}$ such that the number of points in any set follows a Poisson distribution and the numbers of points in disjoint sets are independent. That is, $\{N(t), t \geq 0\}$ is a Poisson process if $N(A_1), \dots, N(A_k)$ are independent Poisson random variables, for any disjoint and measurable subsets $A_1, \dots, A_k$ of measure space S.

The Poisson process is characterized by the number of points in disjoint sets being independent. The conditional intensity function inherits this independence. A Poisson process is a process satisfying the following properties:

1. The numbers of changes in non-overlapping intervals are independent for all intervals.

2. The probability of exactly one change in a sufficiently small interval $h = 1/n$ is $P = \mu h = \mu/n$, where μ is the probability of one change and n is the number of trials.

3. The probability of two or more changes in a sufficiently small interval h is essentially 0.

In the limit of the number of trials becoming large, the resulting distribution is called a Poisson distribution. In all settings, the Poisson point process has the property that each point is stochastically independent to all the other

[1]([16], pp.231)

points in the process, which is why it is sometimes called a purely or completely random process. Despite its wide use as a stochastic model of phenomena representable as points, the inherent nature of the process implies that it does not adequately describe phenomena in which there is sufficiently strong interaction between the points. This has sometimes led to the overuse of the point process in mathematical models and has inspired other point processes, some of which are constructed via the Poisson process, that seek to capture this interaction.
The formal definition of a Poisson process is

Definition 2.2.2. *A Poisson process $\{N(t), t \geq 0\}$ is a stochastic point process satisfies the following conditions*

(1) $P[N(0) = 0] = 1$,

(2) $P[N(t, t+dt) = 1] = \lambda(t)dt + o(dt)$,

(3) $P[N(t, t+dt) > 1 = o(dt)$.

From this basic definition many properties can be derived. One of the most important is the following:

Proposition 2.2.1. *– Let the point process $\left\{N(t), t \geq 0\right\}$ be a nonhomogeneous Poisson process with intensity function $\lambda(t)$ then the r.v. $N(t)$, number of jumps between 0 and t, follows a Poisson distribution with parameter $\Lambda(t)$.*

Hence, the probability that there is no point in the interval $[a, b]$ is

$$Pr[N(a,b) = 0] = \exp\{\Lambda(a,b)\}.$$

Let us consider some examples of Poisson processes.

Example 2.2.1. Homogeneous Poisson process (HPP)
A Poisson process is homogeneous (or stationary) if the intensity function is constant, that is $\lambda(t) = \mu$ where $\mu > 0$.

It can be shown that the number of events $N(s,t)$ in a time-interval $[s,t]$ follows the Poisson distribution with parameter μt, and interevents are identically independent distributed as exponential distribution with parameter μ. The HPP has constant intensity which indicates that the rate of an event is the same at all times, regardless of how frequently such events have occurred previously. When the intensity function is not constant the Poisson process is said to be *nonhomogeneous*. We now consider some examples of such processes.

Example 2.2.2. Power-law process (PLP)
The PLP is a nonhomogeneous Poisson process with intensity function $\lambda(t) = \beta t^{\beta-1}/\alpha^{\beta}$ where $\alpha, \beta > 0$. In this case, the number of events $N(s,t)$ follows

the Poisson distribution with parameter $(t/\alpha)^{\beta}$. The first jump T_1 follows the Weibull distribution with parameter (α,β); the conditional distribution of the k^th jump T_k is a left truncated Weibull distribution with parameter (α,β) ans support $[t_{k-1},+\infty[$. Roughly speaking, its intensity indicates that the number of occurring events grows more densely over time when $\beta > 1$ (convex function) while when $\beta < 1$ the growth slows down with time.

Example 2.2.3. Exponential-law process (ELP)
Similarly to the PLP, one can define a exponential-law process that the intensity function depends on time by the exponential law. As the PLP, this process can be also used to model both aging and improving reliability of repairable systems but with faster rate than the PLP.

The ELP is a non-homogeneous Poisson process with intensity function $\lambda(t) = \alpha e^{\beta t}$ where $\alpha > 0, \beta \in \mathbb{R}$. The number of events $N(s,t)$ follows the Poisson distribution with parameter $\alpha\left(e^{\beta t} - e^{\beta s}\right)/\beta$. The first jump T_1 follows the distribution with density $f^*(t_1) = \alpha e^{\beta t_1} \exp\left\{-\alpha(e^{\beta t_1} - 1)/\beta\right\}$. If $\beta < 0$, the rate of an event decreases or increases over time depending on whether $\beta < 1$ or $\beta > 1$.

In the above two examples of Poisson processes (PLP and ELP), the conditional intensity function depends only on time t and but not on the history $\mathscr{H}_t$ that in this case only contains the dates of previous events.

We now turn to the case where $\lambda^*(t)$ depends not only on times t but also on the number of jumps of preceding events. Therefore the numbers of jumps in different intervals are no longer independent and the intensity is itself a stochastic process.

Stochastic processes with such conditional intensity functions were introduced by Cox [10] who named it *doubly stochastic point processes*. They are also referred to as *Cox processes*.

2.2.4 Cox processes

As mentioned before *Cox processes* also called *doubly stochastic Poisson processes*. are Poisson processes where the intensity function is itself a stochastic process.

A Cox process $\{N(t), t \geq 0\}$ is called self-exciting (resp. self-correcting) if $\mathrm{cov}[N(s,t), N(t,u)] > 0$ for $s < t < u$ (resp. $\mathrm{cov}[N(s,t), N(t,u)] < 0$).

Thus the occurrence of points in a self-exciting point process causes other points to be more likely to occur, whereas in a self-correcting point process, the points have an inhibitory effect.

We will use the term **self-exciting point process (SEPP)** for both cases. Those processes are often used to model events that are temporally clustered.

The formal definition for a SEPP is:

Definition 2.2.3. *A self-exciting point process is a Cox process* $\{N(t), t \geq 0\}$ *satisfying the following conditions*

(1) $P[N(0)=0]=1$,

(2) $P[N(t,t+dt)=1 \mid \mathscr{H}_t]=\lambda^*(t)dt+o(dt)$,

(3) $P[N(t,t+dt)>1 \mid \mathscr{H}_t]=o(dt)$,

where the conditional intensity function depends on the history $\mathscr{H}_t$ *by the form*

$$\lambda^*(t)=\mu+\int_0^t g(t-u)dN(u)$$

where g *is the response function.*

Example 2.2.4. Hawkes process (HaP)
A commonly used model of the SEPP is the Hawkes process, where the conditional intensity is given by

$$\lambda^*(t)=\mu+\sum_{t_i<t}\alpha e^{-\beta(t-t_i)},$$

where $\mu,\alpha,\beta>0$. Parameter μ represents the deterministic background rate of a homogeneous Poisson process and the function g governs the clustering density. Note that each time a new event arrives in this process, the conditional intensity function grows by α and then decreases exponentially by rate β towards μ. In other words, an event increases the chance of getting other events immediately after creating clustering point pattern. Thus a SEPP can be interpreted as clustered point process.

Hawkes process is commonly used in seismology, where they are sometimes called epidemic-type aftershock sequence (ETAS) models, encompassing the notion that earthquakes can have aftershocks, and those aftershocks can have aftershocks, etc. A form of the clustering density g that is commonly used in modeling earthquake aftershocks is the Omori function:

$$g(t)=\frac{\kappa}{(t+c)^p},$$

which corresponds to power-law decay in the clustering behavior over time.

Alternative versions of clustered point processes are formed by generating a sequence of parents and then placing clusters of points (offsprings) around each parent. For example, the Neyman-Scott process ([11], section 3.4) suggests that the offspring points are independently and identically distributed around the parents. An other example is the Bartlett-Lewis process ([9]) where the offspring points are each generated via a renewal process originating at the corresponding parent.

Example 2.2.5. Self-correcting point process (SCPP)
Self-correcting point process is a Cox process with intensity

$$\lambda^*(t)=\exp\left\{\mu t-\sum_{t_i<t}\alpha\right\}$$

A self-correcting models are used in ecology, forestry and other fields to model occurrences that are *well-dispersed*. Such models may be useful in describing births of species, for example, or in seismology for modeling earthquake catalogs after aftershocks have been removed.

Note that the models in the above examples are specified simply by choosing a particular form on the conditional intensity and then give an interpretation for it. Some creativity and common sense can lead to many new models using conditional intensity function.

2.2.5 Marked point processes

The conditional intensity function can be generalized to marked point processes. We can specify the distribution of the mark κ associated with jump t by its conditional density function $f^*(\kappa \mid t) = f^*(\kappa \mid t, \mathscr{H}_t)$ where the history $\mathscr{H}_t = ((t_1,\kappa_1),\ldots,(t_n,\kappa_n))$ now includes information of both jump and marks of past events. We define the conditional intensity function of marked point process as ([16], p. 238):

$$\lambda^*(t,\kappa) = \lambda^*(t) f^*(\kappa \mid t)$$

where $\lambda^*(t)$ is called *ground intensity* which now depends on the marks of the past event also. We can rewrite the above expression as

$$\lambda^*(t,\kappa) = \frac{f^*(t,\kappa)}{1-F^*(t,\kappa)} = \frac{f^*(t)}{1-F^*(t,\kappa)} f^*(\kappa \mid t)$$

where $f^*(t,\kappa)$ is the joint density of the jump and mark and $F^*(t,\kappa)$ is the conditional cumulative distribution function of t given the past jumps and marks. Therefore

$$\lambda^*(t,\kappa) = E\left[N(dt,d\kappa) \mid \mathscr{H}_t)\right],$$

that is, the conditional intensity function of marked point process can be interpreted as the mean number of events in a small time interval dt associated with $d\kappa$, a small variation of the mark.

Sometimes we can assume the independence on the marks. A mark is unpredictable if it is independent to the history. An independent mark is an stronger assumption, which means κ_i is independent of everything else except t_i.

Example 2.2.6. Power-law covariate self-exciting point process (PLC-SEPP)

We will consider in the sequel a particular case of marked-process, a process that we have called power law covariate self-exciting point process. It has an intensity of the form:

$$\lambda^*(t) = \mu + \sum_{t_i<t} \left(\frac{z_i}{z_0}\right)^{\eta} \alpha e^{-\beta(t-t_i)}$$

where $\mu, \alpha, \beta, \eta > 0$ and the mark z has log-normal distribution with density

$$f^*(z \mid t) = \frac{1}{z\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}[\log(z) - m]^2\right\}.$$

The PLC-SEPP can be used to model thunderstorm data where the mark z denotes the amplitude of a lightning strike. The idea behind using this model is that lightning strikes of a thunderstorm trigger the atmosphere to conditions for more impacts nearby in time and space. The bigger the amplitude is the greater influence it makes on the atmosphere.

2.3 Inference

There are many possibilities for estimating the parameters of a process specified by its conditional intensity function, among them the maximum likelihood method and Bayesian approaches are the most common choices. Both two methods are based on the expression of the likelihood function.

2.3.1 Likelihood

Assume that we have observed a point pattern $(t_1, \ldots, t_n)$ on $[0, C]$ for some given $C > 0$, and if we are in the marked case, also its accompanying marks $(\kappa_1, \ldots, \kappa_n)$. Denote $\tau = t_n$ for a event truncation scheme and $\tau = C$ for a time truncation scheme. Then the likelihood function is given by the following proposition (see [16]).

Proposition 2.3.1. *Given an unmarked point pattern $(t_1, \ldots, t_n)$ on an observation interval $[0, C]$, the likelihood function for a parameter of interest θ is given by*

$$L(\theta) = \left(\prod_{i=1}^{n} \lambda^*(t_i;\, \theta)\right) \exp\left\{-\Lambda^*(\tau;\, \theta)\right\}.$$

Given a marked point pattern $((t_1, \kappa_1), \ldots, (t_n, \kappa_n))$ on $[0, C] \times \mathbb{M}$, where $\mathbb{M}$ is a set of marks, the likelihood function is given by

$$\begin{aligned} L(\theta) &= \left(\prod_{i=1}^{n} f^*(\kappa_i \mid t_i;\, \theta)\right)\left(\prod_{i=1}^{n} \lambda^*(t_i;\, \theta)\right) \exp\left\{-\Lambda^*(\tau;\, \theta)\right\} \\ &= \left(\prod_{i=1}^{n} \lambda^*(t_i, \kappa_i;\, \theta)\right) \exp\left\{-\Lambda^*(\tau;\, \theta)\right\} \end{aligned}$$

2.3.2 Maximum likelihood estimation

Depending on the complexity of the likelihood function of a Poisson process, it is not always possible to express the maximum likelihood estimate (MLE) in closed-form. Amongst the examples in the previous sections, the HPP and the PLP have simple closed-forms for the MLE whereas the ELP, the HaP and the PLC-SEPP have not. For most cases, it requires numerical methods to obtain estimates, such as Newton-Raphson for the MLE and Markov Chain Monte Carlo for Bayesian estimates.

2.3.3 Bayesian estimation

Let us consider a sample $\underline{X} = (X_1, \ldots, X_n)$ of i.i.d.r.v. with probability density function $f(\underline{x} \mid \theta)$, $\theta \in \Theta$. Let $\pi(\theta)$ be the prior distribution of the parameter θ and $\pi(\theta \mid ux)$ the posterior distribution. Applying the Bayes' rule, we have:

$$\pi(\theta \mid \underline{x}) = \frac{f(\underline{x} \mid \theta)\pi(\theta)}{\int_{\Theta} f(\underline{x} \mid \theta)\pi(\theta)d\theta}.$$

In other words, the posterior distribution is proportional to the product of the likelihood and the prior distribution.

$$\pi(\theta \mid \underline{x}) \propto f(\underline{x} \mid \theta)\pi(\theta).$$

2.4 Simulation

Simulation of a point process can be performed based on its conditional intensity via two approaches, the *inverse method* and *Ogata's modified thinning procedure*. Two schemes of observation are considered. The *event-truncation* is the case where we stop observing the process as soon as a fixed given number of events – say n – has occurred whereas the *time-truncation* scheme is the case where we stop observing the process at a given known time C. Let us first of all consider the classical inverse method.

2.4.1 Inverse Method

The inverse method is a well-known procedure to generate realizations of a r.v. with a cumulative distribution function F provided that F^{-1} is explicitly known. It relies on the fact that if U is a uniform distribution on $[0,1]$, then $F^{-1}(U)$ has distribution function F. Also, if X has distribution function F, then $F(X)$ is uniformly distributed on $[0,1]$.

The basic algorithm is:

1. Generate $u \sim \mathscr{U}[0,1]$,

2. Compute $F^{-1}(u)$.

Let us consider a stochastic point process with intensity $\lambda^*(t)$. We can express the survivor function for any T_k using (2.1).
Since $P(T_k > t \mid \mathscr{H}_t) = P(N(t_{k-1}, t) = 0)$, we have: $1 - F^*(t) = \exp\left\{-\left[\Lambda^*(t) - \Lambda^*(t_{k-1})\right]\right\}$.

From this relation, provided that Λ^{*-1} exists, we obtain the expression of the inverse c.d.f. given t_{k-1} as :

$$F^{*-1}(t) = \Lambda^{*-1}\Big(\Lambda^*(t_{k-1}) - \log(1-t)\Big).$$

The algorithm is therefore:

1. `repeat`
   (a) `Generate` $u \sim \mathscr{U}[0,1]$,
   (b) `Compute` $t_k = \Lambda^{*-1}\Big(-\log(u)\Big)$
2. `Until` $k = n$ `for` n`-event truncation`
   `or` $\sum t_k > C$ `for time truncation.`

An alternative method relies on the following proposition.

Proposition 2.4.1. – *If* $\{s_i\}_{i\in\mathbb{N}}$ *is a unit rate Poisson process on* $\mathbb{R}$.
Then $\{t_i\}_{i\in\mathbb{N}}$ *where* $t_i = \Lambda^{*-1}(s_i)$, *is a point process with intensity* $\lambda^*(t_i)$.

We will discuss this proposition in the section on graphical test.

Therefore, to simulate a point process with intensity $\lambda^*(t)$ we can simulate a unit rate Poisson process and transform the dates with the function Λ^{*-1} provided this latter exists. Since for a unit rate Poisson process, the interevents are i.i.d. exponential, the algorithm can be described as follows:

1. `repeat`
   (a) `Generate` $w_k \sim \mathscr{E}(1)$,
   (b) `Compute` $t_k = \Lambda^{*-1}\big(\sum_{j=1}^{k} w_j\big))$
2. `Until` $k = n$ `for` n`-event truncation`
   `or` $\sum t_k > C$ `for time truncation.`

We give the details simulation algorithm for some stochastic processes in the appendices (see Appendix I) using the inverse method. In the following section, we introduce the thinning method for simulation.

2.4.2 Thinning method

Simulation by thinning was first introduced in 1979 by Lewis and Shedler ([34]) that is now called the *Lewis' thinning algorithm*. The Lewis' thinning algorithm is based on the following theorem.

Theorem 2.4.1. – *Let* $\lambda(t)$ *be an intensity function associated to a nonhomogeneous Poisson process. Consider a homogeneous Poisson process* $\{\bar{N}(t), t \geq 0\}$ *with intensity function* $\bar{\lambda}$. *Let* $\bar{t}_1, \ldots, \bar{t}_{\bar{N}(C)}$ *be the jumps of the process in the interval* $[0, C]$. *Suppose that for* $0 \leqslant t \leqslant C$, $0 \leqslant \lambda(t) \leqslant \bar{\lambda}$. *For* $k = 1, \ldots, \bar{N}(C)$, *delete the point* $\bar{t}_k$ *with probability* $1 - \lambda(\bar{t}_k)/\bar{\lambda}$; *then the remaining points form a nonhomogeneous Poisson process* $\{N(t), t \geq 0\}$ *with intensity* $\lambda(t)$.

The algorithm can be described as follows:

Input: $\lambda(t), C$

Initialize $n = m = 0, t_0 = s_0 = 0, \bar{\lambda} = sup_{0 \leqslant t \leqslant C} \lambda(t)$;

while $s_m < C$ do

Generate $u \sim$ uniform(0,1);

Let $w = -\log(u)/\bar{\lambda}$;

Set $s_{m+1} = s_m + w$;

Generate $D \sim$ uniform(0,1);

If $D \leqslant \lambda(s_{m+1})/\bar{\lambda}$ then

$t_{n+1} = s_{m+1}$;

$n = n + 1$;

end

$m = m + 1$;

end

if $t_n \leqslant C$ then

return $\{t_k\}_{k=1,\ldots,n}$

else

return $\{t_k\}_{k=1,\ldots,n-1}$

end

In 1981, Ogata ([42]) suggested to modify the Lewis' algorithm in the following way in order to deal more efficiently with "complicated" intensity. He proposed that instead of simulating a homogeneous Poisson process on a time window $[0, C]$, simulate a homogeneous Poisson process on some interval $[t, t + h(t)]$ for some chosen function $h(t)$ (this is the maximum distance we may go forward in time from t and it may be infinite). This HPP has a chosen constant intensity $m(t)$ on $[t, t + h(t)]$, which fulfill the condition:

$$m(t) \geqslant \sup_{u \in [t, t+h(t)]} \lambda^*(u)$$

Actually we only need to simulate the first point t_k of this HPP. There are now two possibilities: If $t_k > h(t)$, then there is no point in $[t, t + h(t)]$, so restart at any point in $[t, t + h(t)]$, but if $t_k \leqslant h(t)$, there might be a point at t_k in

$[t, t+h(t)]$. In the latter case we need to figure out whether to keep this point or not. By independent thinning, we keep it with probability $\lambda^*(t_k)/m(t)$. Whether or not we keep it, we start all over at t_k.

This algorithm called *the Ogata's modified thinning algorithm* can be described as follow:

1. Set $t = 0$ and $n = 0$.
2. Repeat until $t > C$:
   (a) Compute $m(t)$ and $h(t)$.
   (b) Generate independent random variables $s \sim \text{Exp}(m(t))$ and $U \sim \text{Unif}(0,1)$.
   (c) If $s > h(t)$, set $t = t + h(t)$.
   (d) Else if $t + s > T$ or $U > (t+s)/m(t)$, set $t = t + s$.
   (e) Otherwise, set $n = n + 1$, $t_n = t + s$, $t = t + s$.
3. Output is $(t_1, \ldots, t_n)$.

2.5 Model Checking

In practice, after fitting a model to a data set we normally would like to assess whether the model provides an adequate fit. Checking the adequacy of a model is the next step of data analysis. We can only use the data in the first half of the observation interval to fit a model, and then simulate predictions of the second half to see if this corresponds to the second half of the observed data. Or we can use all of the data, and compare with simulations of the whole dataset. In addition to the model checking approaches based on simulation, there are some particular kinds of model checking associated with the conditional intensity function. For some models we can build a graphical test, as for power-law process. If a graphical test is not available, residual test is a more general tool to assess the model.

2.5.1 Graphical test

Graphical test is a visual tool for model checking. It gives us an immediate view whether the model fit the data or not. For example, plotting time (t_i, i) on the plane can give us the first idea of the process. If all the point are nearby a straight line we can guess that the point pattern follows the homogeneous Poisson process. Duane plot is a graphical test for the power-law process. It is said that if the PLP fits the point pattern $(t_1, \ldots, t_n)$ then plotting $(\log(i), \log(t_i)$ on the plane gives an image of a straight line. Such a simple graphical test is not available for the exponential-law process and the self-exciting point process. Therefore we need residual test for those models.

2.5.2 Residual test

A useful technique for evaluating point process models relies on proposition 2.4.1 which can considered as a re-scaling method. The method essentially involves re-scaling the time axis of the observed point process N at time t by a factor of the intensity $\lambda(t)$. More specifically, if the point pattern $(t_1,\ldots,t_n)$ are observed from time 0 to time C, then each point t_i is moved to the new time $\Lambda(t_i)$, where Λ is the compensator. The resulting process M is a stationary Poisson process with unit rate, provided the original point process is simple. Similarly, one may inspect residuals obtained by randomly thinning the process: that is, keeping each point t_i independently with probability inversely proportional to $\lambda(t_i)$. As with rescaled residuals, the resulting thinned residuals will be distributed according to a stationary Poisson process ([43]). In practice, one may use the estimated intensity or compensator in place of the true intensity or compensator and inspect the rescaled or thinned process for uniformity. Several tests exist for this purpose, with different uses depending on the alternative hypotheses. Some of the most useful are tests based on second and higher order properties. There are also more general second-order tests for point processes that do not rely on a stationary Poisson null hypothesis.

A very powerful general result due to Papangelou [47] can be applied to test the form of the intensity process. This result already mentioned with proposition 2.4.1, can be stated in the following term [16], p. 23:

> *'Any point process satisfying a simple continuity condition can be transformed into a Poisson process if we allow a random time change in which $\Lambda(t)$ depends on the past of the process up to time t.'*

From this remark, we can state that if $(t_1,\ldots,t_n)$ is a realization of a point process with the compensator $\Lambda(t) = (t/\alpha)^\beta$ then $(\Lambda(t_1),\ldots,\Lambda(t_n))$ is a realization of the homogeneous Poisson Process with rate 1. Thus, the points $(\Lambda(t_i), i)$ should stand on a straight line.

This statement relies on the well-known result that if the r.v. X has a continuous distribution $F(x)$, the $U = F(X)$ has a uniform distribution on $[0,1]$.

Formally, we have the following theorem ([16], pp. 258):

Theorem 2.5.1. *Let N be a simple point process adapted to a history $\mathscr{F}$ with bounded strictly positive conditional $\mathscr{F}$-intensity $\lambda(t)$ and $\mathscr{F}$-compensator $\Lambda(t) = \int_0^t \lambda(u)\,du$ that is not a.s.-bounded. Under the random time change $t \to \Lambda(t)$, the transform process $\tilde{N}(t) = N\big(\Lambda^{-1}(t)\big)$ is a Poisson process with unit rate.*

Conversely, suppose there is a given history $\mathscr{G}$, a $\mathscr{G}$-adapted cumulative process $M(t)$ with a.s. finite, monotonically increasing and continuous trajectories, and a $\mathscr{G}$-adapted simple Poisson process $N_0(t)$. Let $\mathscr{F}$ denote the history *of σ-algebras $\mathscr{F}_t = \mathscr{G}_{M(t)}$. Then $N(t) = N_0\big(M(t)\big)$ is a simple point process the is $\mathscr{F}$-adapted and has $\mathscr{F}$-compensator $M(t)$.*

Hence if a point pattern is a realization of a point process with conditional intensity function $\lambda^*(t)$, then the compensator will transform the pattern into a realization of homogeneous Poisson process with a unit rate. In practice this means that if we model an observed point pattern with a point process, and the model is well selected, then the transformed pattern should closely resemble a unit-rate HPP. In other words, the model checking boils down to checking whether the inter-event times are independent exponential variables with mean one.

On the other hand, if the model does not fit the point pattern, residual test may provide important information on how it does not fit. For example, if the data contains an unrealistically large gap for the model between t_i and t_{i+1}, then the transformed data will contain a large gap between s_i and s_{i+1}, i.e. $s_{i+1} - s_i$ will be to large to realistically come from a unit rate exponential distribution. A bit of creativity in analyzing the residuals can give us all kinds of information about the original point pattern. We can also assess the adequacy of a fitted model by checking the uniformly of the transformed event times $\hat{\Lambda}(t_i)$ on the interval $[0, \hat{\Lambda}(1)]$, where $\hat{\Lambda}$ is obtained by substituting in λ the unknown parameters by their estimates. The uniformity can be visually checked through a QQ-plot or through formal tests, such as the χ^2 test or the Kolmogorov-Smirnov (KS) test. Large $p-$values of the test indicate acceptable model fits. Of course, in interpreting the magnitude of $p-$values we should bear in mind that the transformation function $\hat{\Lambda}$ carries the randomness of the data, and the distribution of the test statistic would be more spread-out than that calculated from a pre-specified Λ, and therefore we should be more tolerant of smaller $p-$values. In any case, the $p-$values should be assessed together with other diagnostic checks, such as the QQ-plot.

2.5.3 Model selection: AIC and BIC

Given a sequence of jumps $(t_1, \ldots, t_n)$, you might have several candidates for a model. The fit of a parametric model can be assessed using a likelihood score such as *Akaike information criterion* (AIC) [1] which is:

$$\mathrm{AIC} = 2p - 2\log L(\theta)$$

where p is the number of parameters in the model. The AIC rewards a model for higher likelihood and penalizes a model for overfitting. The model with the lowest AIC should be selected.

The Bayesian information criterion (BIC) [53] is another possibility. The definition of this criterion is

$$\mathrm{BIC} = p\log(n) - 2\log L(\hat{\theta})$$

where p is still the number of parameters in the model and $\log(\hat{\theta})$ is the log-likelihood at the value $\hat{\theta}$ that maximizes the likelihood.

Chapter 3

Power-Law Process

This chapter is devoted to the study of the power-law process. Due to its convenient intensity function, it is possible to develop a graphical test for the model and to obtain explicit expressions for maximum likelihood estimators of its parameters. Some properties of the model are given. For a Bayesian approach, we review some common used priors then we introcuce a conjugate prior for the model. The distribution is investigated in details and some elicitation strategies are suggested to integrate prior information from experts.

3.1 Introduction

Let us consider a counting process $\{N(t), t \geqslant 0\}$ such that:

$$\Pr\big(N(t+h) - N(t) = 1 \mid N(t) = n\big) = \lambda(t)\, h + o(h), \tag{3.1}$$

$$\Pr\big(N(t+h) - N(t) > 1 \mid N(t) = n\big) = o(h). \tag{3.2}$$

The function $\lambda(t)$ is called the intensity function of the process $\{N(t), t \geqslant 0\}$.

$N(t+h) - N(t)$ can be denoted $N[t, t+h)$. Its represents the number of events in a small interval of length h and the intensity can be defined as:

$$\lambda(t) = \lim_{h \to 0} \frac{Pr\big(N[t, t+h) > 0\big)}{h}.$$

From the assumptions (3.1) and (3.2), it can be shown that:

$$\lambda(t) = \lim_{h \to 0} \frac{E\big(N[t, t+h)\big)}{h}$$

and

$$\Pr(N(t) = k) = \frac{\Lambda(t)^k}{k!} \exp\big\{-\Lambda(t)\big\},\ k \in \mathbb{N}.$$

where $\Lambda(t) = \int_0^t \lambda(s)\,ds$ is the cumulative intensity also called the *compensator*.

Thus the r.v. $N(t)$, number of events between 0 and t, has a Poisson distribution with parameter $\Lambda(t)$ and the expected numbers of events in the interval $[0,t)$, is $E[N(t)] = \Lambda(t)$.

This can be generalized to any interval $[a,b] \subset \mathbb{R}^+$, the number of events in $[a,b]$ – that is to say $N[a,b] = N(b) - N(a)$ – has a Poisson distribution with parameter $\Lambda(b) - \Lambda(a)$.

A classical model for $\lambda(t)$ is the power-law model which is of the form:

$$\lambda(t) = \mu\ \beta\ t^{\beta-1}, \mu, \beta > 0.$$

A stochastic process with such intensity function is called a **power-law process**. Others denominations can be find in the literature (Duane process, Weibull process, AMSAA[1] process, etc.). The denomination Weibull process is misleading and Soland ([57],[58]) calls Weibull process the renewal process with Weibull independent interarrivals.

When $\beta = 1$, $\lambda(t)$ is constant, equal to μ. We have the well-known homogeneous Poisson process (HPP). In this case, the times between consecutive events are independent r.v. following an exponential distribution with the parameter μ. Therefore the HPP with intensity μ is equivalent to a renewal process characterized by exponential distributed interarrivals. The expected number of events in a given interval is the product of μ with the length of the interval. Let T_k be the date of the k^{th} events, $k = 1,2,\ldots$. Relying on the relationship $\{T_k < t\} = \{N(t) > k\}$, it can be proved that the distribution of S_k is an Erlang distribution with parameter (k,μ).

When $\beta > 1$, the intensity depends on time and the power-law process is a nonhomogeneous Poisson process (NHPP).

The time before the first event in this case, has a Weibull distribution but the interarrivals are not independent. The distribution of the time t_i to the i^{th} event given the time t_{i-1} of the $(i-1)^{th}$ event is a left truncated Weibull distribution with support $[t_{i-1}, +\infty[$. The distribution of T_i is a generalized gamma distribution with pdf:

$$\frac{1}{\Gamma(i)} \mu\beta t^{i\beta-1} \exp\left\{-\mu t^{\beta}\right\},\ t > 0.$$

The compensator is $\lambda(t) = \mu\ t^{\beta}$. It is linear on a log-log scale that is $\log\lambda(t) = \log\mu + \beta\log t$ and this observation due to Duane [17] suggests a procedure to derive a graphical test for PLP that we will describe later.

The PLP is very popular because of its mathematical tractability and well-documented inference procedures. It covers many situation in reliability (growth, decay) where the intensity is called the rate of occurrence of

[1] Army Material System Analysis Activity

failure (ROCOF). The literature on PLP is abundant. We are going to give an overview in the sequel.

Event truncation and time truncation

There are two schemes of collecting data: one can stop an observation until a fixed number of event or one can observe the process in a fixed time window. In the first scheme, number of event n is fixed but the last jump t_n is random and we call this *event truncation*. In the second scheme, the ending time C is fixed but number of event n is random and we call this *time truncation*.

Classical inference

In 1964 Duane [17] published a report concerning failure data occurring on different systems during their development programs. He observed that for these systems, the observed cumulative failure rate versus cumulative operating hours is close to a straight line when plotted on log-log paper.

In 1974, Crow [13] interpreted the Duane observation in term of nonhomogeneous Poisson process:

> *'If the cumulative failure rate (expected number of failures at time t divided by t) versus test time is linear on log-log scale, then the system failure times follow a non homogeneous Poisson process with Weibull intensity function $u(t) = \lambda\beta t^{\beta-1}$. If the system reliability is improving, then $u(t)$ is decreasing; i.e., $0 < \beta < 1$.'*

Considering that several systems are observed in a time window, he obtains the MLE of the parameters μ and β, builds confidence intervals and suggests a goodness-of-fit test relying on Cramér-Von Mises statistic.

Using Monte Carlo methods, Crow [14] studies the distribution of $\lambda(t_k)/\hat{\lambda}(t_k)$ where t_k is the date of the k^{th} failure and $\hat{\lambda}(t_k) = \hat{\mu}\hat{\beta}t^{\hat{\beta}-1}$.

In 1976, Finkelstein [19] considers the alternative parametrization of the ROCOF:

$$\lambda(t) = \frac{\beta}{\alpha^{\beta}}\, t^{\beta-1}, \alpha > 0, \beta > 0.$$

He shows that $(\hat{\alpha}/\alpha)^{\hat{\beta}}$ is a pivotal function and has the same distribution as $\hat{\alpha}_{11}^{\hat{\beta}_{11}}$ where $\hat{\alpha}_{11}$ and $\hat{\beta}_{11}$ are the MLE of α and β when the sample is from a Weibull distribution with $\alpha = 1$ and $\beta = 1$. This result relies on a change scale [47]. The distribution of $\hat{\alpha}_{11}^{\hat{\beta}_{11}}$ can be tabulated via Monte Carlo methods and be used to construct confidence intervals for α.

In 1978, Lee and Lee [33] obtains the exact distribution of $(\hat{\alpha}/\alpha)^{\hat{\beta}}$.

In the same year, Engelhart and Bain [18] derives exact prediction intervals based upon maximum likelihood estimation. They propose the exact solution and a simpler approximate prediction limits.

In 1980, they develop conditional inference procedures for the shape parameter β and approximate confidence limits for the scale parameter α [2].

Calabria and al. (1988) [5] consider modified maximum likelihood estimators of the expected number of failures in a given interval and of the failure intensity. They compare with the maximum likelihood computing mean squarred errors.

Sen and Khattree (1998) [55] consider general decision theory, loss function.

From frequentist perspective, one should note that event truncation data and time truncation data are treated differently and results should also be inferenced in different ways.

Bayesian approach

The advantages of the Bayesian approach are well-known. It allows the practitioner to introduce in the inferential procedure prior information. Thus even if the quality of the observation is poor, inference remains possible.

Higgins and Tsokos (1981) [26] suggest a quasi-Bayes strategy to estimate the value of the intensity at the n-th failure date that is $\nu_n = \mu\beta t_n^\beta$. They use a pseudo-likelihood and a gamma prior distribution for ν_n.

Guida et al. (1989) [23] propose Bayes estimators considering different type of priors. They investigate a joint non-informative priors for the parameters of the form $\pi(\alpha,\beta) \propto (\alpha\beta)^{-1}$. This approach leads to non closed form estimators. Then they consider a uniform prior for β and $\pi(\alpha) \propto 1/\alpha$ for α, a non-informative prior. For the informative case they choose a gamma distribution as prior distribution on the mean of failures in a given interval $[0,T]$. Prior knowledge on this quantity is then transformed into a conditional distribution on α. Again the Bayes estimator obtained are not in a closed form.

Bar-Lev et al. (1992) [3] consider a non-informative prior of the form $1/(\alpha\beta^\nu)^{-1}$ dealing with independence between the parameters. We will consider the use of Jeffreys' rule to propose non-informative priors.

From a Bayesian perspective, Guida et al. [23] discuss point and interval estimation for α and β assuming event truncation data and using several different choices of prior, both informative and non-informative. Kyparisis & Singpurwalla [32] analyze both interval and event truncation data by employing informative priors on α and β, and derive prediction distributions of future jumps and the number of jumps in some future time interval. Their predictive and posterior distributions generally require complicated numerical computations. Calabria et al. [6] also derive predictive distributions for future failure times using both informative and non-informative priors and note the numerical equivalence with classical methods when non-informative priors are used. The above three references usually assume that the prior distributions on α and β are independent. Alternatively Calabria et al. [6] consider independent priors on α and the mean value function $m(t)$, with t fixed.

From the Bayesian approach both scenarios can be handle in the same manner and result in the same type of posterior inference on α and β in contrast to the frequency approach in which each case must be treated separately and different types of results are obtained.

3.2 Graphical Test

In this section we propose an overview of different graphical techniques for testing the adequacy to the PLP model on the basis of a n-sample of events occurrences $(t_1,\dots,t_n)$.

Duane's Plot

As mentionned in the introduction, Duane [17] observes that

> *'when the cumulative failure rate (defined as total malfunctions since program start, divided by total operating hours since start) is plotted on log-log-paper as a function of cumulative operating hours, the points tented to line up on a straight line.'*

This was true for many different sets of reliability data and many other engineers had seen the same results.

This heuristic is consolidated using a NHPP approach. If $\{N(t), t \geqslant 0\}$ is a PLP(α,β) then $E[N(t)] = (t/\alpha)^{\beta}$, $t \geqslant 0$ and $E[N(t)/t] = t^{\beta-1}/\alpha^{\beta}$, $t \geqslant 0$. Taking the logarithm on both sides, we obtain:

$$\log E[N(t)/t] = (\beta - 1)\log t - \beta \log \alpha. \tag{3.3}$$

Suppose that we observe $\underline{t} = (t_1,\dots,t_n)$, n dates of jumps of a process.

$N(t_i) = i$ and can be consider as an estimation of the expected numbers of jump at time t_i. Therefore $E[N(t_i)/t_i] \approx i/t_i$.

$t_i = \sum_{j=1}^{i}(t_j - t_{j-1})$ and t_i/i can be interpreted as the mean time between failure at time t_i denoted $MTBF_i$.

Thus the equation (3.3) becomes:

$$\log MTBF_i = (1-\beta)\log t_i + \beta \log \alpha,\ i = 1,\dots,n.$$

The usual reasonning is that if the points $(\log(t_i), \log(MTBF_i))$, $i = 1,\dots,n$ stand roughly on a straight line, it would not be absurd to consider the process to be a PLP. The graph obtained plotting $(\log(t_i), \log(t_i/i))$, $i = 1,\dots,n$ is called a **Duane's plot**.

It is often suggested that graphical estimates of the parameters can be obtained from the graph.

However, in 2002, Rigdon [50] points out that 1. the power-law process does not imply a linear Duane plot and that 2. the linearity of a Duane plot does not imply the power-law process. His argument for 1. is that t_i the date of the $i-th$ jump is a realization of a random variable T_i. Then, the approximation of $E[N(t_i)/t_i]$ by i/t_i is not correct and the non-linearity in plotting $\left(\log t_i, log(t_i/i)\right)$ does not really indicate a departure from a power-law process.

For 2., Rigdon considers a process where jumps occur at every 4 hours. Then the dates of jump are $t_i = 4i$ and the Duane's plot is:

$$\left(log 4i, log(i/4i)\right) = \left(log 4i, log(1/4)\right).$$

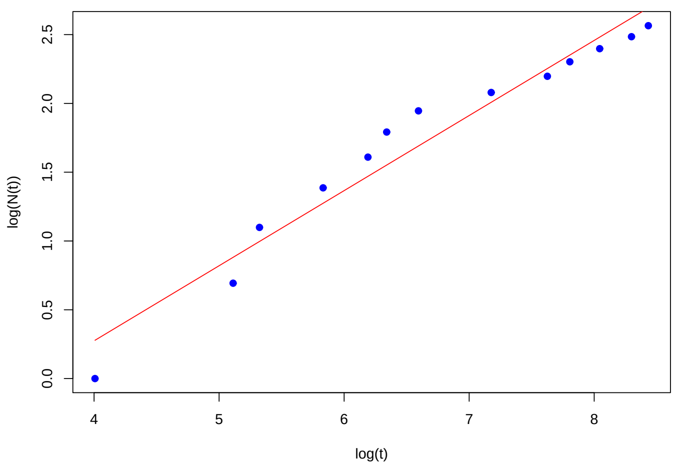
Ridgon says 'These points are linearly related (the line is perfectly flat) but the process is not a power-law process'. If jumps occur every 4 hours, the process is a homogeneous Poisson process (that is a power-law process with $\alpha = 1/4$ and $\beta = 1$).

Some authors consider directly the expression $E[N(t)] = (t/\alpha)^\beta$ and plot $(\log t_i, \log i)$.

In 2003, Gaudoin et al. [20] suggest a goodness-of-fit test based on R^2 showing that this statistic is a pivotal function (see Appendix A).

Example The figure 3.1 displays the Duane's plot for the data of example 3.1. From the graph, we can deduce the following estimations for the param-

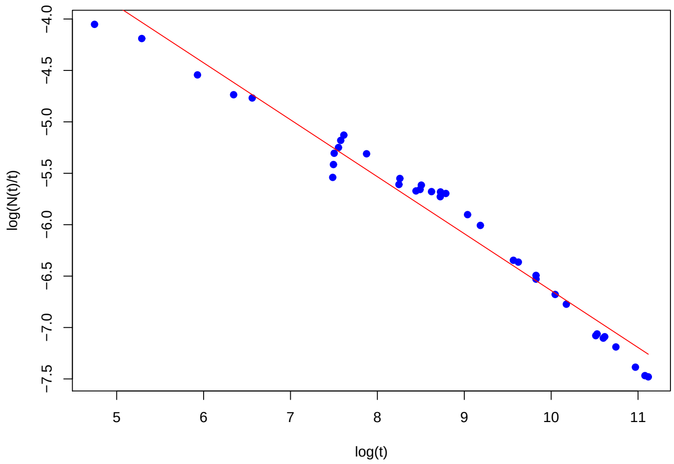
Figure 3.1: *Duane's plot for aircraft generator failure times*



eters : $\alpha_g = 0.546$ and $\beta_g = 33.037$.

The figure 3.2 displays the Duane's plot for the data of example 3.2. From the graph, we can deduce the following estimations for the parameters : $\alpha_g = 0.446$ and $\beta_g = 11.850$.

Figure 3.2: *Duane's plot for aircraft generator failure times*



R-square Test

This example concerns a complex type of aircraft generator. They were deduced from Figure 2 in Duane [17] by Ridgon and Basu [51].

Table 3.1: *Failure times in hours for aircraft generator*

i	t_i	i	t_i
1	55	8	1308
2	166	9	2050
3	205	10	2453
4	341	11	3115
5	488	12	4017
6	567	13	4596
7	731		

The data of example 2 was presented in Musa [41] and Kyparisis and Singpurwalla [32]. It involves software failures times.

The following sections are devoted to inference for PLP. First of all, we consider classical frequentist method before considering the Bayesian approach.

Table 3.2: *Software failure times in seconds*

i	t_i	i	t_i	i	t_i
1	115	14	3821	27	18494
2	115	15	3861	28	18500
3	198	16	4649	29	23061
4	376	17	4871	30	26229
5	570	18	4943	31	36800
6	706	19	5558	32	37363
7	1783	20	6147	33	40133
8	1798	21	6162	34	40785
9	1813	22	6552	35	46378
10	1905	23	8415	36	58074
11	1955	24	9752	37	64798
12	2026	25	14260	38	67344
13	2632	26	15094		

3.3 Maximum Likelihood Method

3.3.1 Likelihood

Assume that we have observed a point process $\{N(t), t \geqslant 0\}$ in a time window $[0, C]$ for some given $C > 0$. The number of $\underline{t}_n$ is a realization of the n-dimensional r.v. $(t_1, \dots, t_n)$ where n is a realization of the r.v. $N(C)$. This situation corresponds to a situation of *time truncation*. When the observation stops when the n^{th} event occurs , we say to have a *n-event truncation*. $(C \equiv t_n)$[2]. We are going to investigate both situations in the sequel.

Let us consider the distribution of the first event T_1. The survival function of T_1 is

$$Pr(T_1 > t) = Pr(N(0,t) = 0) = \exp\{-\Lambda(t)\} = \exp\left\{-(t/\alpha)^\beta\right\}$$

and its probability density function is then

$$f_{T_1}(t) = \frac{\beta}{\alpha^\beta} t^{\beta-1} \exp\left\{-(t/\alpha)^\beta\right\}.$$

Hence the first event T_1 has a Weibull distribution with scale parameter α and shape parameter β.

Let us denote $\underline{t}_i = (t_1, \dots, t_i)$ the history of the PLP until the i^{th} event. For any $t \geqslant t_i$, the conditional survival function of T_i given $\underline{t}_i$ is

$$\begin{aligned} Pr(T_{i+1} > t \mid \underline{t}_i) &= Pr(N(t_i, t) = 0) = \exp\{-\Lambda(t_i, t)\} \\ &= \exp\left\{-(t/\alpha)^\beta + (t_i/\alpha)^\beta\right\}. \end{aligned}$$

[2] n.b. C is fixed while T_n is a r.v.

Denote $f^*_{T_i}(t) = f_{T_i|\underline{t}_i}(t)$ the conditional probability density function of T_i given $\underline{t}_i$. We have

$$f^*_{T_i}(t) = \frac{\beta}{\alpha^\beta} t^{\beta-1} \exp\left\{-(t/\alpha)^\beta + (t_i/\alpha)^\beta\right\} \tag{3.4}$$

which is a left truncated Weibull distribution with support $[t_i, +\infty[$.

Event truncation

For event truncation, the likelihood is the product of the conditional distributions. We obtain:

$$L(\alpha,\beta) = \frac{\beta^n}{\alpha^{n\beta}} \prod_{i=1}^n t_i^{\beta-1} \exp\left\{-(t_n/\alpha)^\beta\right\}.$$

Time truncation

In case of time truncation, since the observation end up at C, we know that if $N(C) = n$, T_{n+1} given $\underline{t}_n$ is greater than C and the contribution to the likelihood is:

$$Pr(T_{n+1} > C \mid \underline{t}_n) = \exp\left\{-(C/\alpha)^\beta + (t_n/\alpha)^\beta\right\} \tag{3.5}$$

The likelihood is then the product of the conditional distributions (3.4) for $i = 1, \dots, n$ times the term (3.5). We obtain:

$$L(\alpha,\beta) = \frac{\beta^n}{\alpha^{n\beta}} \prod_{i=1}^n t_i^{\beta-1} \exp\left\{-(C/\alpha)^\beta\right\}.$$

Let us resume both situations in the following expression for the likelihood:

$$L(\alpha,\beta) = \frac{\beta^n}{\alpha^{n\beta}} \prod_{i=1}^n t_i^{\beta-1} \exp\left\{-(\tau/\alpha)^\beta\right\}$$

with $\tau = t_n$ for event truncation and $\tau = C$ with $N(C) = n$ for time truncation. The log-likelihood is then:

$$\log L(\alpha,\beta) = n\ln\beta - n\beta\ln\alpha + (\beta-1)\sum_{i=1}^n \ln t_i - (\tau/\alpha)^\beta.$$

3.3.2 Maximum likelihood estimation

To find the maximum likelihood estimators of α and β, we solve the likelihood equations:

$$\begin{cases} \dfrac{\partial}{\partial\alpha}\log L(\alpha,\beta) = 0 \\[2ex] \dfrac{\partial}{\partial\beta}\log L(\alpha,\beta) = 0 \end{cases} \iff \begin{cases} -\dfrac{n\beta}{\alpha} + \beta\dfrac{\tau^\beta}{\alpha^{\beta+1}} = 0 \\[2ex] \dfrac{n}{\beta} - n\log\alpha + \displaystyle\sum_{i=1}^n \log t_i - \log\left(\frac{\tau}{\alpha}\right)\left(\frac{\tau}{\alpha}\right)^\beta = 0 \end{cases}$$

The solution $\hat{\alpha}, \hat{\beta}$ to these equations are obtained in closed form:

$$\hat{\beta} = \frac{n}{\sum_{i=1}^{n} \log(\tau/t_i)} \text{ and } \hat{\alpha} = \frac{\tau}{n^{1/\hat{\beta}}}.$$

Remark that the terms of the Hessian matrix are:

$$\begin{aligned}
H_{11}(\alpha,\beta) &= \frac{\partial^2}{\partial\alpha^2} \log L(\alpha,\beta) = \frac{n\beta}{\alpha^2} - \frac{\beta(\beta+1)}{\alpha^2}\left(\frac{\tau}{\alpha}\right)^{\beta}, \\
H_{12}(\alpha,\beta) &= \frac{\partial^2}{\partial\alpha\partial\beta} \log L(\alpha,\beta) \\
&= -\frac{n}{\alpha} + \frac{1}{\alpha}\left(\frac{\tau}{\alpha}\right)^{\beta} + \frac{1}{\alpha}\left(\frac{\tau}{\alpha}\right)^{\beta} \log\left[\left(\frac{\tau}{\alpha}\right)^{\beta}\right], \\
H_{22}(\alpha,\beta) &= \frac{\partial^2}{\partial\beta^2} \log L(\alpha,\beta) \\
&= -\frac{n}{\beta^2} + \frac{1}{\beta^2}\left(\frac{\tau}{\alpha}\right)^{\beta} \log^2\left[\left(\frac{\tau}{\alpha}\right)^{\beta}\right].
\end{aligned}$$

At the point $(\hat{\alpha}, \hat{\beta})$, the values of the Hessian matrix terms are:

$$\begin{aligned}
H_{11}(\hat{\alpha},\hat{\beta}) &= -n\left(\frac{\hat{\beta}}{\hat{\alpha}}\right)^2, \\
H_{12}(\hat{\alpha},\hat{\beta}) &= \frac{n \log n}{\hat{\alpha}}, \\
H_{22}(\hat{\alpha},\hat{\beta}) &= -\frac{n(1+\log^2 n)}{\hat{\beta}^2},
\end{aligned}$$

and the determinant is

$$\det H(\hat{\alpha},\hat{\beta}) = \frac{n^2}{\hat{\alpha}^2}.$$

Since $H_{11}(\hat{\alpha},\hat{\beta}) < 0$ and $\det H(\hat{\alpha},\hat{\beta}) = \frac{n^2}{\hat{\alpha}^2} > 0$, $H(\hat{\alpha},\hat{\beta})$ is negative definite and the solution $(\hat{\alpha},\hat{\beta})$ is unique [36].

3.3.3 Properties of maximum likelihood estimators

Since event truncation and time truncation give different inferences for maximum likelihood estimators, we use the notations $\hat{\alpha}(T_n), \hat{\beta}(T_n)$ and $\hat{\alpha}(C), \hat{\beta}(C)$ to indicate the maximum likelihood estimators of α, β for event truncation and time truncation, respectively.

The following theorem can be used to derive inference on α and β.

Theorem 3.3.1. *–Denote*

$$
\begin{aligned}
Z(T_n) &= \frac{2n\beta}{\hat{\beta}(T_n)} \\
Z(C) &= \frac{2n\beta}{\hat{\beta}(C)} \\
U_n &= \left(\frac{T_n}{\alpha}\right)^{\beta} \\
W &= (\hat{\alpha}(T_n)/\alpha)^{\hat{\beta}(T_n)}.
\end{aligned}
$$

Then the following statements hold:

(i) $Z(T_n)$ *has a chi-square distribution with* $2(n-1)$ *degrees of freedom.*

(ii) $Z(C)$ *has a chi-square distribution with* $2n$ *degrees of freedom.*

(iii) U_n *has a gamma distribution with parameter* $(n,1)$*, hence* $2U_n$ *has chi-square distribution with* $2n$ *degree of freedom.*

(vi) The cdf of W is

$$
Pr(W \leqslant w) = \int_0^{\infty} G[(nw)^{z/2n}]\, g(z) dz,
$$

where $g(z)$ *is the chi-square density with (n-1) degrees of freedom:*

$$
g(z) = \frac{1}{2^{n-1}(n-2)!} z^{n-2} \exp\left\{-\frac{z}{2}\right\}
$$

and $G(z)$ *the incomplete gamma distribution with parameters* $(n,1)$*:*

$$
G(x) = \frac{1}{(n-1)!} \int_0^x y^{n-1} e^{-y} dy.
$$

Proof 1. The proof of (i) is given in Appendix C. It relies on the distribution of $\underline{U}_n = (U_1, \ldots, U_{n-1}, U_n)$ where $U_i = \log(T_n/T_i), i = 1, \ldots n-1$ and $U_n = (T_n/\alpha)^{\beta}$. It is shown that the U_i, $i = 1, \ldots, n-1$ are i.i.d random variables of exponential distribution with parameter β. Therefore $\sum_{i=1}^{n-1} U_i$ has a gamma distribution with parameters $(n-1, \beta)$ and $2n\beta/\hat{\beta} = 2\beta \sum_{i=1}^{n-1} U_i$ has a gamma distribution with parameters $(n-1, 1/2)$ that is a chi-square distribution with $2(n-1)$ degrees of freedom.

To prove (ii) we need to find the conditional distribution of $(T_1, \ldots, T_n)$ given $N(C) = n$. Denote $U_i = C/T_i, i = 1, \ldots, n$, it can be shown (see Appendix C) that $U_1, \ldots, U_n$ are independent r.v. and have an exponential distributions with parameter β which leads to the chi-square distribution with n degrees of freedom for Z in the case of time truncation.

(iii) follows from the fact that $U_n = (T_n/\alpha)^\beta$ in the case of event truncation has a gamma distribution with parameters $(n,1)$ which means that $S = 2U_n$ has a gamma distribution parameters $(2n/2, 1/2)$ that is a Chi-square distribution with $2n$ degrees of freedom.

To prove (iv), we express W as a function of S and Z. $\hat{\alpha} = T_n/n^{1/\hat{\beta}}$ thus

$$W = \left(\frac{T_n/n^{1/\hat{\beta}}}{\alpha}\right)^{\hat{\beta}} = \frac{1}{n}\left(\frac{T_n}{\alpha}\right)^{\hat{\beta}} = \frac{1}{n}\left[\left(\frac{T_n}{\alpha}\right)^{\beta}\right]^{\hat{\beta}/\beta} = \frac{1}{n}\left(\frac{S}{2}\right)^{\hat{\beta}/\beta}.$$

But $\hat{\beta} = 2n\beta/Z$, then $W = (1/n)(S/2)^{2n/Z}$.
The expression of $Pr(W \leqslant w) = Pr\big((1/n)(S/2)^{2n/Z} \leqslant w\big)$ is then obtained by conditioning on Z.

$\square$

Unbiased estimator

We need to find the expectation of $\hat{\beta}$ to check if it is the biased estimator of the parameter β. For a event truncated realization, $\hat{\beta} = n/U$ where $U = \sum_{i=1}^{n-1} U_i$ has gamma distribution of parameters $(n-1,\beta)$. Since $1/U$ has inverse gamma distribution of parameters $(n-1,\beta)$, we have

$$E(\hat{\beta}) = nE(1/U) = n\frac{\beta}{n-2} = \frac{n}{n-2}\beta.$$

Hence $\hat{\beta}$ is a biased estimator of β. The unbiased estimator is

$$\hat{\beta}^* = \frac{n-2}{n}\hat{\beta} = \frac{n-2}{\sum_{i=1}^{n-1}\log(T_n/T_i)}.$$

For a time truncated realization, follow the same reasoning, the unbiased estimator is

$$\hat{\beta}^* = \frac{n-1}{n}\hat{\beta} = \frac{n-1}{\sum_{i=1}^{n-1}\log(C/T_i)}.$$

Consistency

Applying the law of large number for a series of i.i.d random variables for a event truncated realization we get

$$\frac{1}{\hat{\beta}} = \frac{n}{n-1}\frac{\sum_{i=1}^{n-1}U_i}{n-1} \xrightarrow{a.s} \frac{1}{\beta}.$$

And for a time truncated observation, we have

$$\frac{1}{\hat{\beta}} = \frac{\sum_{i=1}^{n}U_i}{n} \xrightarrow{a.s} \frac{1}{\beta}$$

so $\hat{\beta} \xrightarrow{a.s} \beta$. Thus, $\hat{\beta}$ is a consistent estimator for β.

3.4 Bayesian Approach

For Bayesian analysis, one can consider different choices of priors such as non-informative prior, conjugate prior. Choosing the prior distribution is an important matter. Guida et al. [23] propose different choice : a joint non informative prior of the form $(\alpha\beta)^{-1}$, a uniform distribution for β and $1/\alpha$ for α. Then considering a gamma prior distribution on $m(t)$, the number of expected failures, they express a distribution for α given β. Bar-Lev et al. [3] consider a joint prior for $(\alpha\beta)$ of the form $(\alpha\beta^{\nu})^{-1}$. They obtain a chi-square distribution for β posterior distribution but a cumbersome expression for α posterior distribution. Sen & Khattree [55] study specifically the Bayesian estimator of $m(t)$ considering different lost functions. Let us consider non-informative prior by applying Jeffreys' rule.

Jeffreys' rule

We consider in this section the construction of noninformative prior using the Jeffreys' rule. The method is to choose $\pi(\alpha,\beta) \propto [\det I(\alpha,\beta)]^{-1}$ where $I(\alpha,\beta)$ is the Fischer information matrix that is the (2×2) matrix with element:

$$
\begin{aligned}
I_{1,1}(\alpha,\beta) &= E\left[-\frac{\partial^2}{\partial\alpha^2}\log L(\alpha,\beta)\right],\\
I_{1,2}(\alpha,\beta) &= I_{2,1}(\alpha,\beta) = E\left[-\frac{\partial^2}{\partial\alpha\partial\beta}\log L(\alpha,\beta)\right],\\
I_{2,2}(\alpha,\beta) &= E\left[-\frac{\partial^2}{\partial\beta^2}\log L(\alpha,\beta)\right].
\end{aligned}
$$

We have previously compute the second derivatives of the log-likelihood ot obtain:

$$
\begin{aligned}
\frac{\partial^2}{\partial\alpha^2}\log L(\alpha,\beta) &= \frac{n\beta}{\alpha^2} - \beta(\beta+1)\frac{\tau^\beta}{\alpha^{\beta+2}},\\
\frac{\partial^2}{\partial\alpha\partial\beta}\log L(\alpha,\beta) &= -\frac{n}{\alpha} + \frac{1}{\alpha}\left(\frac{\tau}{\alpha}\right)^\beta + \frac{\beta}{\alpha}\log\left(\frac{\tau}{\alpha}\right)\left(\frac{\tau}{\alpha}\right)^\beta,\\
\frac{\partial^2}{\partial\beta^2}\log L(\alpha,\beta) &= -\frac{n}{\beta^2} - \left[\log\left(\frac{\tau}{\alpha}\right)\right]^2\left(\frac{\tau}{\alpha}\right)^\beta.
\end{aligned}
$$

Event truncation

Recall that, for failure truncation, $U_n = (T_n/\alpha)^\beta$ follows the gamma distribution with parameter $(n,1)$. Then $E(U_n) = n$.

$$\begin{aligned}
I_{11}(\alpha,\beta) &= -\frac{n\beta}{\alpha^2} + \frac{\beta(\beta+1)}{\alpha^2}E(U_n) = \frac{n\beta^2}{\alpha^2},\\
I_{12}(\alpha,\beta) &= I_{21}(\alpha,\beta) = \frac{n}{\alpha} - \frac{1}{\alpha}E(U_n) - \frac{1}{\alpha}E[U_n\log(U_n)] = \frac{m_1}{\alpha},\\
I_{22}(\alpha,\beta) &= \frac{n}{\beta^2} - \frac{1}{\beta}E[U_n\log^2(U_n)] = \frac{n+m_2}{\beta^2},
\end{aligned}$$

where $m_1 = E[U_n\log(U_n)]$ and $m_2 = E[U_n\log^2(U_n)]$ are free from α,β and we know that $1/e \le m_1 \le 2n^2, 0 \le m_2 \le 2n^2$.

The determinant of the Fisher information matrix is therefore

$$\det[I(\alpha,\beta)] = \frac{n^2 + m_2 n - m_1^2}{\alpha^2}.$$

The Jeffreys' prior for vector-parameter θ is proportional to the square-root of the determinant of the Fisher information matrix

$$\pi(\theta) \propto \sqrt{\det[I(\theta)]}.$$

Therefore Jeffreys' non-informative prior for (α,β) is

$$\pi(\alpha,\beta) \propto \frac{1}{\alpha}.$$

Time truncation

For time truncation, $n = N(C)$ is a random variable that follows a Poisson distribution with parameter $(C/\alpha)^\beta$ and $E[N(C)] = (C/\alpha)^\beta$. Then

$$\begin{aligned}
I_{11}(\alpha,\beta) &= -\frac{\beta}{\alpha^2}E[N(C)] + \frac{\beta(\beta+1)}{\alpha^2}\left(\frac{C}{\alpha}\right)^\beta\\
&= \frac{\beta^2}{\alpha^2}\left(\frac{C}{\alpha}\right)^\beta,\\
I_{12}(\alpha,\beta) &= I_{21}(\alpha,\beta)\\
&= \frac{1}{\alpha}E[N(C)] - \frac{1}{\alpha}\left(\frac{C}{\alpha}\right)^\beta - \frac{1}{\alpha}\left(\frac{C}{\alpha}\right)^\beta \log\left(\frac{C}{\alpha}\right)^\beta\\
&= -\frac{1}{\alpha}\left(\frac{C}{\alpha}\right)^\beta \log\left(\frac{C}{\alpha}\right)^\beta,\\
I_{22}(\alpha,\beta) &= \frac{1}{\beta^2}E[N(C)] + \frac{1}{\beta^2}\left(\frac{C}{\alpha}\right)^\beta \log^2\left[\left(\frac{C}{\alpha}\right)^\beta\right]\\
&= \frac{1}{\beta^2}\left(\frac{C}{\alpha}\right)^\beta\left[1 + \log^2\left(\frac{C}{\alpha}\right)^\beta\right].
\end{aligned}$$

The determinant of the Fisher information matrix is therefore

$$\det[I(\alpha,\beta)] = \left[\frac{1}{\alpha}\left(\frac{C}{\alpha}\right)^{\beta}\right]^2. \tag{3.6}$$

Remark that at time C, n events occurred. Therefore n can be considered as an approximation of the expected number of jumps in $[0,C]$ that is

$$E[N(C)] = \Lambda(C) = \left(\frac{C}{\alpha}\right)^{\beta} \approx n.$$

If we replace $(C/\alpha)^{\beta}$ by n into (3.6) we get $\det[I(\alpha,\beta)] \approx n^2/\alpha^2$. Thus we can suggest $\pi(\alpha,\beta) \propto 1/\alpha$ as a noninformative prior for (α,β).

Reparametrization

This reparametrization will be used in the next section for Bayesian analysis of the PLP with the natural conjugate prior. Let $\mu = \alpha^{-\beta}$ so with new parametrization of $\theta = (\mu,\beta)$ the PLP has intensity function

$$\lambda(t) = \mu\beta t^{\beta-1},$$

and the compensator

$$\Lambda(t) = \mu t^{\beta}.$$

The likelihood becomes

$$L(\mu,\beta) = (\mu\beta)^n \left(\prod_{i=1}^{n} t_i\right)^{\beta-1} \exp\left\{-\mu\tau^{\beta}\right\},$$

and the log-likelihood is

$$\log L(\mu,\beta) = -\mu\tau^{\beta} + n\log(\mu) + n\log(\beta) + (\beta-1)\sum_{i=1}^{n}\log(t_i).$$

The gradients are

$$\begin{aligned}
J_1 &= \frac{\partial}{\partial\mu}\log L(\mu,\beta) = -\tau^{\beta} + \frac{n}{\mu},\\
J_2 &= \frac{\partial}{\partial\beta}\log L(\mu,\beta) = -\mu\tau^{\beta}\log(\tau) + \frac{n}{\beta} + \sum_{i=1}^{n}\log(t_i).
\end{aligned}$$

The maximum likelihood estimates for μ,β are

$$\hat{\beta} = \frac{n}{\sum_{i=1}^{n}\log(\tau/t_i)}, \tag{3.7}$$

$$\hat{\mu} = \frac{n}{n^{\hat{\beta}}}. \tag{3.8}$$

The Hessian matrix is

$$\begin{aligned}
H_{11}(\mu,\beta) &= \frac{\partial^2}{\partial\mu^2}\log L(\mu,\beta) = -\frac{n}{\mu^2},\\
H_{12}(\mu,\beta) &= \frac{\partial^2}{\partial\mu\partial\beta}\log L(\mu,\beta) = -\tau^{\beta}\log(\tau),\\
H_{22}(\mu,\beta) &= \frac{\partial^2}{\partial\beta^2}\log L(\mu,\beta) = -\frac{n}{\beta^2} - \mu\tau^{\beta}\log^2(\tau).
\end{aligned}$$

At the point $(\hat{\mu},\hat{\beta})$, the values of the Hessian matrix are

$$\begin{aligned}
H_{11}(\hat{\mu},\hat{\beta}) &= -\frac{n}{\hat{\mu}^2} < 0,\\
H_{12}(\hat{\mu},\hat{\beta}) &= -\frac{n\log(\tau)}{\hat{\mu}},\\
H_{22}(\hat{\mu},\hat{\beta}) &= -\frac{n}{\hat{\beta}^2} - n\log(\tau),
\end{aligned}$$

and determinant of the Hessian matrix at that point is

$$\det[H(\hat{\mu},\hat{\beta}] = \left(\frac{n}{\hat{\mu}\hat{\beta}}\right)^2 > 0.$$

That means $H(\hat{\mu},\hat{\beta})$ is negative defined and $(\hat{\mu},\hat{\beta})$ is the unique maximal of the log-likelihood. The Fisher information matrix

$$\begin{aligned}
I_{11}(\mu,\beta) &= \frac{n}{\mu^2},\\
I_{12}(\mu,\beta) &= \frac{m_1 - n\log(\mu)}{\mu\beta},\\
I_{22}(\mu,\beta) &= \frac{m_2 - 2m_1\log(\mu) + n(1+\log^2(\mu))}{\beta^2},
\end{aligned}$$

where $m_1 = E[U_n\log(U_n)]$ and $m_2 = E[U_n\log^2(U_n)]$ are free from μ,β and we know that $1/e \le m_1 \le 2n^2, 0 \le m_2 \le 2n^2$. The determinant of the Fisher information matrix is therefore

$$\det[I(\alpha,\beta)] = \frac{n^2 + m_2 n - m_1^2}{\mu^2\beta^2}.$$

Thus Jeffreys non-informative prior for (μ,β) is

$$\pi(\mu,\beta) \propto \frac{1}{\mu\beta}.$$

3.4.1 Noninformative prior

We consider Bayesian inference for the PLP with noninformative prior. While the maximum likelihood estimates are interpreted differently according to event truncation or time truncation scheme, Bayesian estimates give the same interpretation for both schemes.

Theorem 3.4.1. *– Let $\underline{t} = (t_1, \dots, t_n)$ be an observation of a PLP observed in a time window $[0, C]$. Denote $\tilde{\mu}$and$\tilde{\beta}$ the Bayesian estimates of μ and β. Assuming a quadratic loss and a Jeffreys' noninformative prior $\pi(\mu, \beta) \propto (\mu\beta)^{-1}$, we obtain the following results:*

(i) the posterior density is

$$\pi(\mu, \beta \mid \underline{t}) = \frac{s_n^n}{\Gamma(n)^2} \mu^{n-1} \beta^{n-1} p_n^{\beta} \exp\left\{-\mu\tau^{\beta}\right\},$$

where $p_n = \prod_{i=1}^n t_i$ and $s_n = \sum_{i=1}^n \log(\tau/t_i)$,

(ii) the posterior marginal distribution of β is a gamma distribution with parameter (n, s_n).

(iii) the posterior conditional distribution of μ given β is a gamma distribution with parameter (n, τ^{β}),

(iv) the Bayesian estimators are:

$$\tilde{\beta} = \frac{n}{s_n}, \tag{3.9}$$

$$\tilde{\mu} = n\left(\frac{s_n}{s_n + \log(\tau)}\right)^n. \tag{3.10}$$

Proof 2. *Proof of (i):* Recall that the probability density of an observation of the PLP is

$$\begin{aligned} f(\underline{t} \mid \mu, \beta) &= \mu^n \beta^n p_n^{\beta-1} \exp\{-\mu\tau^{\beta}\} \\ &\propto \mu^n \beta^n p_n^{\beta} \exp\{-\mu\tau^{\beta}\}. \end{aligned}$$

With prior density $\pi(\mu, \beta) \propto (\mu\beta)^{-1}$, and applying the Bayes'theorem, the posterior density is

$$\pi(\mu, \beta \mid \underline{t}) = K(\underline{t}) \mu^{n-1} \beta^{n-1} p_n^{\beta} \exp\{-\mu\tau^{\beta}\}.$$

The normalizing constant $K(\underline{t})$ can be computed by taking double-integral and applying Fubini's theorem

$$
\begin{aligned}
K(\underline{t})^{-1} &= \int_0^\infty \int_0^\infty \mu^{n-1}\beta^{n-1}p_n^\beta \exp\{-\mu\tau^\beta\}d\mu d\beta \\
&= \int_0^\infty \beta^{n-1}p_n^\beta \left(\int_0^\infty \mu^{n-1}\exp\{-\mu\tau^\beta\}d\mu\right)d\beta \\
&= \int_0^\infty \beta^{n-1}p_n^\beta \frac{\Gamma(n)}{(\tau^\beta)^n}d\beta \\
&= \Gamma(n)\int_0^\infty \beta^{n-1}e^{-\beta s_n}d\beta \\
&= \frac{\Gamma(n)^2}{s_n^n}.
\end{aligned}
$$

Proof of (ii): The posterior marginal distribution of β is obtained by integrating out μ from the posterior joint distribution of (μ,β)

$$
\begin{aligned}
\pi(\beta \mid \underline{t}) &= \int_0^\infty \frac{s_n^n}{\Gamma(n)^2}\mu^{n-1}\beta^{n-1}p_n^\beta \exp\{-\mu\tau^\beta\}d\mu \\
&= \frac{s_n^n}{\Gamma(n)}\beta^{n-1}e^{-\beta s_n}.
\end{aligned}
$$

Thus $\beta \mid \underline{t}$ has gamma distribution with parameter (n, s_n).
Proof of (iii): The posterior conditional marginal distribution of μ given β is:

$$
\pi(\mu \mid \beta, \underline{t}) = \frac{\pi(\mu,\beta \mid \underline{t})}{\pi(\beta \mid \underline{t})} = \frac{(\tau^\beta)^n}{\Gamma(n)}\mu^{n-1}\exp\{-\mu\tau^\beta\}.
$$

Hence $\mu \mid \beta, \underline{t}$ has a gamma distribution with parameter (n, τ^β).
Proof of (iv): With quadratic loss, the Bayesian estimates are the expectations of the posterior marginal distributions. The posterior expectation of β is obvious since its distribution is gamma:

$$
E(\beta \mid \underline{t}) = \frac{n}{s_n}.
$$

Thus we get Bayesian estimate for β as in (iv). The posterior conditional expectation of μ given β is

$$
E(\mu \mid \beta, \underline{t}) = \frac{n}{\tau^\beta}.
$$

The posterior expectation of μ is obtained by taking the expectation of the posterior conditional expectation of μ given β.

$$
\begin{aligned}
&= E(\mu \mid \underline{t}) = E(E(\mu \mid \beta, \underline{t})) = \int_0^\infty E(\mu \mid \beta, \underline{t})p(\beta \mid \underline{t})d\beta \\
&= \int_0^\infty \frac{s_n^n}{\Gamma(n)}\frac{n}{\tau^\beta}\beta^{n-1}e^{-\beta s_n}d\beta = \frac{n s_n^n}{\Gamma(n)}\int_0^\infty \beta^{n-1}\exp\left\{-\beta(s_n+\log(\tau)\right\}d\beta \\
&= n\left(\frac{s_n}{s_n+\log(\tau)}\right)^n.
\end{aligned}
$$

Hence we get a Bayesian estimate for μ as in (iv). Recall that the maximum likelihood estimates of β and μ are:

$$\hat{\beta} = \frac{n}{s_n}, \quad \hat{\mu} = \frac{n}{\tau^{\hat{\beta}}}.$$

Therefore the Bayesian estimate of β is equal to its maximum likelihood estimate which is the classical result. Now we consider the relationship between Bayesian estimate of μ and its maximum likelihood estimate. Substitute $s_n = n/\tilde{\beta}$ from the equation (3.9) to the equation (3.10) we obtain

$$\tilde{\mu} = n\left(\frac{s_n}{s_n + \log(\tau)}\right)^n = n\left(1 + \frac{\log(\tau)}{s_n}\right)^{-n} = n\left(1 + \frac{\log(\tau)^{\tilde{\beta}}}{n}\right)^{-n}.$$

When n is large enough, the last part of the above equation can be approximated by

$$n\left(1 + \frac{\log(\tau)^{\tilde{\beta}}}{n}\right)^{-n} \approx \frac{n}{\tau^{\tilde{\beta}}}.$$

Thus we have

$$\tilde{\mu} \approx \frac{n}{\tau^{\tilde{\beta}}}.$$

Therefore, when n is very big, the Bayesian estimate of μ can be derived from the Bayesian estimate of β by the same function $\mu = n/\tau^{\beta}$ as for maximum likelihood estimates. The consequence is that, when we have no prior information about the parameter, their Bayesian estimates and their maximum likelihood estimates are getting closed to each other when the sample size is getting larger.

□

3.4.2 Independent conjugate priors (Oliveira et al. [45])

When we have an observation of the PLP observed from time truncation scheme in the fixed time window $[0, C]$, value of the compensator at the ending time can be considered as a parameter of the process with some prior information, that is $\eta = \Lambda(C) = \mu C^{\beta}$. The new parameter η can be interpreted as the expectation of $N(C)$. If we have some prior information about η then we can conduct Bayesian analysis on the PLP. This procedure is given in the following theorem.

Theorem 3.4.2. *– Let $\underline{t} = (t_1, \ldots, t_n)$ a realization of the PLP with parameter (η, β) that we observe in the time window $[0, C]$. Denote $\tilde{\eta}, \tilde{\beta}$ the Bayesian estimates of η, β respectively. A natural joint conjugate prior for (η, β) is a product of two independent prior for each parameter where:*

(i) The natural conjugate for η is gamma distribution with parameter (a,b) and the natural conjugate for β is gamma distribution with parameter (c,d).

(ii) $\eta \mid \underline{t}$ and $\beta \mid \underline{t}$ are independent; $\eta \mid \underline{t}$ has gamma distribution with parameter $(a+n, b+1)$ and $\beta \mid \underline{t}$ has gamma distribution with parameter $(c+n, d+s_n)$.

(iii) The Bayesian estimates for η and β are:

$$\tilde{\eta} = \frac{a+n}{b+1},$$
$$\tilde{\beta} = \frac{c+n}{d+s_n}.$$

Proof 3. *Proof of (i):* With the new parametrization of (η, β) the probability density of an observation becomes

$$\begin{aligned} f(\underline{t} \mid \eta, \beta) &= \eta^n C^{-n\beta} \beta^n p_n^{\beta-1} e^{-\eta} \\ &\propto \eta^n e^{-\eta} \times \beta^n e^{-\beta s_n}. \end{aligned}$$

It follows that η and β are orthogonal and the natural joint conjugate prior is simply a product of two independent gamma distributions $\pi(\eta, \beta) = \pi(\eta) \times \pi(\beta)$ where

$$\pi(\eta) = \frac{1}{\Gamma(a}\eta^{a-1}e^{-b\eta},$$
$$\pi(\beta) = \frac{1}{\Gamma(c}\beta^{c-1}e^{-d\beta}.$$

Applying the Bayes' theorem, the posterior distribution is

$$\pi(\eta, \beta \mid \underline{t}) \propto \eta^{a+n-1} e^{-(b+1)\eta} \times \beta^{a+n} e^{-(b+s_n)\beta}.$$

Proof of (ii): Thus $\eta \mid \underline{t}$ has a gamma distribution with parameters $(a+n, b+1)$ and $\beta \mid \underline{t}$ has a gamma distribution with parameters $(c+n, d+s_n)$. Hence we obtain the result in *(ii)*.

Proof of (iii): Assuming quadratic loss, the Bayesian estimate is the posterior expectation of the parameter. The posterior expectation for η, β are easily obtained since the posterior distributions are gamma. The Bayesian estimates are then:

$$\tilde{\eta} = \frac{a+n}{b+1},$$
$$\tilde{\beta} = \frac{c+n}{d+s_n}.$$

□

Prior elicitation

Prior elicitation is easily deduced from prior information of each parameter. Let $g_{\eta,1}$ is a guess for the value of η and $g_{\eta,2}$ is a guess for standard deviation associated with $g_{\eta,1}$; let $g_{\beta,1}$ is a guess for the value of β and $g_{\beta,2}$ is a guess for standard deviation associated with $g_{\beta,1}$. Since the expectation and the variance of a gamma distribution are available in closed-form expressions, one can easily obtained values for a,b,c,d as following:

$$\frac{a}{b} = g_{\eta,1}, \quad \frac{a}{\sqrt{b}} = g_{\eta,2},$$
$$\frac{c}{d} = g_{\beta,1}, \quad \frac{c}{\sqrt{d}} = g_{\beta,2}.$$

Therefore we get

$$a = \frac{g_{\eta,2}^2}{g_{\eta,1}}, \quad b = \left(\frac{g_{\eta,2}}{g_{\eta,1}}\right)^2,$$
$$c = \frac{g_{\beta,2}^2}{g_{\beta,1}}, \quad d = \left(\frac{g_{\beta,2}}{g_{\beta,1}}\right)^2.$$

3.5 Conjugate Prior: the H-B Distribution

Our purpose is to consider a conjugate prior for the Bayesian analysis of the PLP. This problem has already been addressed in paper Oliveira et al. [45] and in paper Huang & Bier [27]. While Olivera et al. propose a reparametrization with independent conjugate priors for each parameter, Huang and Bier propose a joint conjugate prior that allows independency between two parameters of the model. However, their choice of prior is not really a conjugate prior for the PLP since the posterior distributon and the prior distribution are not in the same family. In adition, they do not give any elicitation strategy that is practical.

We define a 5-parameter bivariate distribution that we name H-B. Properties of this distribution are given and it is shown that this distribution is a natural conjugate prior for the PLP. The Bayes estimates are then obtained and we suggest a technique to elicit the hyperparameters of the prior distribution. This technique is very attractive and simple since the practitioner has only to give a prior guess on β and a standard deviation associated with his guess. To end with and before concluding, we apply the method on simulated data and on data from aircraft generator.

3.5.1 Prior information and conjugate priors

Let's consider Bayesian inference for the PLP with the parametrization (μ,β) and the intensity $\lambda(t) = \mu\beta t^{\beta-1}$. The probability density of an observation

$\underline{t} = (t_1, \dots, t_n)$ in the time window $[0, C]$ of the PLP with parameter (μ, β) is

$$f(\underline{t} \mid \mu, \beta) \propto (\mu\beta)^n p_n^\beta \exp\left\{-\mu\tau^\beta\right\}$$

where $\tau = t_n$ for event truncated data and $\tau = C$ for time truncated data.

Mimicking the above functional form, the prior density should be in the form

$$\pi(\mu, \beta) \propto (\mu\beta)^{a-1} b^\beta \exp\left\{-\mu\tau^\beta\right\}.$$

This prior density belong to a new bivariate distribution with three parameters (a, b, τ). Since the last parameter is fixed to be τ it remains two parameter a, b to be elicitated. It requires, for instance, a guess on the value of β and a guess on standard deviation associated to the first guess.

If we have one more prior guess on the value of μ, we need a conjugate prior with four parameters (a, b, τ, d) that allow three free parameters to be elicitated. We then consider a prior distribution with the density of the form

$$\pi(\mu, \beta) \propto (\mu\beta)^{a-1} b^\beta \exp\left\{-d\mu\tau^\beta\right\}.$$

At the last step, if we have four prior guesses on expectations and standard deviations of both β and μ then we consider a prior distribution with four parameters (a, b, τ, d, m) with the density of the form

$$\pi(\mu, \beta) \propto \mu^{m-1} \beta^{a-1} b^\beta \exp\left\{-d\mu\tau^\beta\right\}.$$

All of the three forms suggest us a family of bivariate distribution that we will propose in the next section.

3.5.2 H-B distribution

We now introduce a new family of bi-variate distributions we name H-B distribution referring to the work of Huang and Bier. The definition of a H-B distribution is given below.

Definition 3.5.1. *–A bivariate r.v. $(X, Y) \in \mathbb{R}^+ \times \mathbb{R}^+$ has a H-B distribution with parameters (a, b, c, d, m) where $a, b, c, d, m > 0$ and such that $b < c^m$, if it has a p.d.f. of the form:*

$$f_{X,Y}(x, y) = K\, x^{m-1} y^{a-1} b^y \exp\left\{-dxc^y\right\}$$

where

$$K = \frac{d^m \left[\log(c^m/b)\right]^a}{\Gamma(m)\Gamma(a)}. \tag{3.11}$$

We denote: $(X, Y) \sim \text{H-B}(a, b, c, d, m)$.

K is obtained by computing:

$$\int_0^\infty \int_0^\infty x^{m-1} y^{a-1} b^y \exp\left\{-dxc^y\right\} dx\, dy$$

Applying the Fubini's theorem we have

$$\begin{aligned}
&\int_0^\infty \int_0^\infty x^{m-1} y^{a-1} b^y \exp\left\{-dxc^y\right\} dxdy \\
&= \int_0^\infty y^{a-1} b^y \left(\int_0^\infty x^{m-1} \exp\left\{-dxc^y\right\} dx \right) dy \\
&= \int_0^\infty y^{a-1} b^y \frac{\Gamma(m)}{(dc^y)^m} dy \\
&= \frac{\Gamma(m)}{d^m} \int_0^\infty y^{a-1} \exp\left\{-y \log\left(c^m/b\right)\right\} dy \\
&= \frac{\Gamma(m)\Gamma(a)}{d^m \log^a\left(c^m/b\right)}.
\end{aligned}$$

Marginal distributions and conditional distributions

One of the two components and the conditional distribution of the other one of a H-B distribution have gamma distributions. The following theorem provides a conditional decomposition of a H-B distribution.

Theorem 3.5.1. – *Let* $(X,Y) \sim H\text{-}B(a,b,c,d,m)$ *then:*

(i) the marginal distribution of Y *is a gamma distribution with parameters* $\left(a, log(c^m/b)\right)$*,*

(ii) the conditional distribution of X *given* $Y = y$ *is a gamma distribution with parameters* (m, dc^y)*.*

Proof 4. –

Proof of (i): The marginal density function of Y is obtained by integrating out X from the joint density of (X,Y)

$$\begin{aligned}
f_Y(y) &= \int_0^\infty K x^{m-1} y^{a-1} b^y \exp\left\{-dxc^y\right\} dx \\
&= K y^{a-1} b^y \int_0^\infty x^{m-1} \exp\left\{-dxc^y\right\} dx \\
&= K y^{a-1} b^y \frac{\Gamma(m)}{(dc^y)^m} \\
&= \frac{\log^a\left(c^m/b\right)}{\Gamma(a)} y^{a-1} \exp\left\{-y \log\left(c^m/b\right)\right\}.
\end{aligned}$$

Therefore Y has a gamma distribution with parameters $\left(a, \log(c^m/b)\right)$.

Proof of (ii): The conditional density function of X given $Y = y$ is

$$\begin{aligned} f_{X|Y=y}(x) &= f_{X,Y}(x,y)/f_Y(y) \\ &= \frac{K x^{m-1} y^{a-1} b^y \exp\{-dxc^y\}}{\left[\log^a(c^m/b)/\Gamma(a)\right] y^{a-1} \exp\{-y\log(c^m/b)\}} \\ &= \frac{(dc^y)^m}{\Gamma(m)} x^{a-1} \exp\{-dc^y x\}. \end{aligned}$$

Thus $X \mid Y = y$ has a gamma distribution with parameters (m, dc^y).

□

Conditional expectation and conditional variance

From the above theorem, we have

$$E(X \mid Y = y) = \frac{m}{dc^y},$$
$$Var(X \mid Y = y) = \frac{m}{d^2 c^{2y}}.$$

Expectation, variance and covariance

The previous theorem allows us to compute the expectation and the variance of X and Y. Let $k = \log(c^m/b)$ then we have

Theorem 3.5.2. – *Let $(X,Y) \sim H\text{-}B(a,b,c,d,m)$ and denote $k = \log(c^m/b)$ then we have the following results.*

(i) *The expectation of Y is*
$$E(Y) = a/k.$$

(ii) *The variance of Y is*
$$Var(Y) = a/k^2.$$

(iii) *If $b < min(c^m, c^{m+1})$, then the expectation of X is*
$$E(X) = \frac{m}{d}\left[\frac{k}{k+\log(c)}\right]^a.$$

(iv) *If $b < min(c^m, c^{m+2})$, the variance of X is*
$$Var(X) = \frac{m(m+1)}{d^2}\left[\frac{k}{k+2\log(c)}\right]^a - \frac{m^2}{d^2}\left[\frac{k}{k+\log(c)}\right]^{2a}.$$

(v) *If $b < min(c^m, c^{m+1})$, the covariance between X and Y is*
$$Cov(X,Y) = -\frac{am}{dk}\frac{\log(c)k^a}{[k+\log(c)]^{a+1}}$$

Proof 5. –

Proof of (i) and (ii) : The expectation and the variance of Y is easily obtained since it has gamma distribution.

Proof of (iii) : To compute $E(X)$ we consider the conditional expectation and compute $E[E(X \mid Y)]$ to obtain:

$$\begin{aligned}
E(X) &= \int_0^\infty \frac{m}{dc^y} f_Y(y) dy \\
&= \int_0^\infty \frac{m}{dc^y} \frac{k^a}{\Gamma(a)} y^{a-1} e^{-ky} dy \\
&= \frac{m}{d} \frac{k^a}{\Gamma(a)} \int_0^\infty y^{a-1} \exp\{-y(k+\log(c))x\} dy \\
&= \frac{m}{d} \left[\frac{k}{k+\log(c)} \right]^a .
\end{aligned}$$

Proof of (iv) : A similar reasoning provides:

$$\begin{aligned}
E(X^2) &= \int_0^\infty \frac{m(m+1)}{d^2 c^{2y}} f_Y(y) dy \\
&= \int_0^\infty \frac{m}{d^2 c^{2y}} \frac{k^a}{\Gamma(a)} y^{a-1} e^{-ky} dy \\
&= \frac{m(m+1)}{d^2} \left[\frac{k}{k+2log(c)} \right]^a .
\end{aligned}$$

Proof of (v) : We have

$$\begin{aligned}
E(XY) &= \int_0^\infty \int_0^\infty xy f_{X,Y}(x,y) dx dy \\
&= \frac{am}{kd} \left[\frac{k}{k+\log(c)} \right]^{a+1} .
\end{aligned}$$

Hence the covariance is

$$Cov(X,Y) = E(XY) - E(X)E(Y) = -\frac{am}{dk} \frac{\log(c) k^a}{[k+\log(c)]^{a+1}}$$

□

Thus, when $c > 1$ the two components are negative correlated; when $0 < c < 1$ the two components are positive correlated and when when $c = 1$, it is interesting to remark that the two components are independent as in the following theorem.

Theorem 3.5.3. – *Let* $(X,Y) \sim H\text{-}B(a,b,c,1,m)$*. The two components X and Y are independent; Y has a gamma distribution with parameters* $(a, \log(1/b))$ *and X has a gamma distribution with parameters* (m,d)*.*

Proof 6. When $c = 1$, clearly

$$\begin{aligned} f_{X,Y}(x,y) &= \frac{\log^a(1/b)d^m}{\Gamma(a)\Gamma(m)} x^{m-1} y^{a-1} b^y \exp\{-dx\} \\ &= \frac{[\log(1/b)]^a}{\Gamma(a)} y^{a-1} \exp\{-\log(1/b)y\} \times \frac{d^m}{\Gamma(m)} x^{m-1} \exp\{-dx\} \\ &= f_Y(y) \times f_X(x). \end{aligned}$$

Thus we obtain the results. Note that in this case, the expectations and the variances are easily obtained. We have:

$$\begin{aligned} E(X) &= m/d, \quad Var(X) = m/d^{2\text{'}} \\ E(Y) &= a/\log(1/b), \quad Var(Y) = a/[\log(1/b)]^2. \end{aligned}$$

□

Mode

If $a > 1, m > 1$ and $b < min\left(c^m, c^{m-1}\right)$ then the H-B distribution with parameter (a,b,c,d,m) is uni-modal with the mode at (x_{mod}, y_{mod}) where

$$\begin{aligned} y_{mod} &= \frac{a-1}{k - \log(c)}, \\ x_{mod} &= \frac{m-1}{dc^{y_m}}. \end{aligned}$$

Graphical illustration

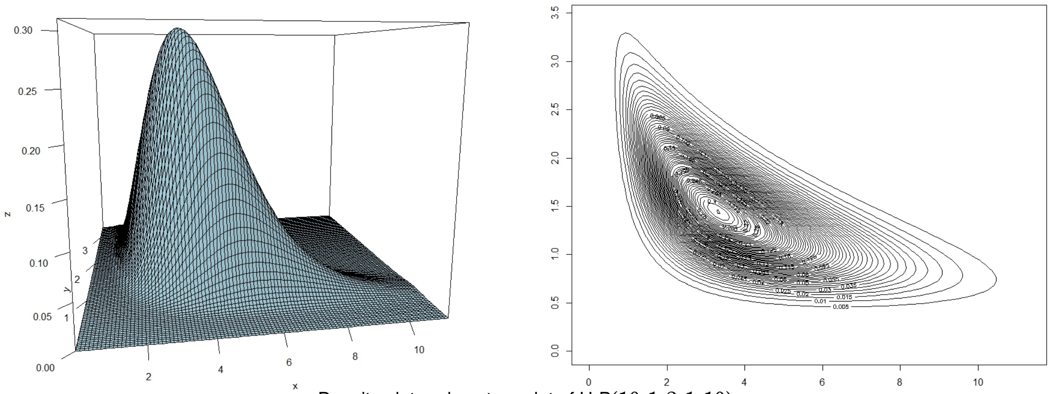
The figure 3.3 displays 3-D plots of density functions of H-B distribution with different parameters. Wet fix $m = a = 10, c = 2, d = 1$ and choose different b's values such that $b < min(c^m, c^{m-1})$ so the densities have unique mode.

3.5.3 The H-B distribution as a conjugate prior

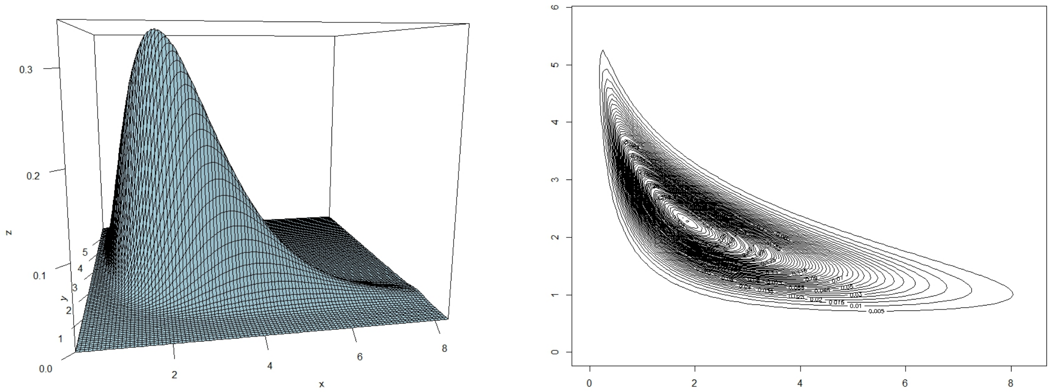
We now consider the Bayesian inference for the PLP with the parametrization (μ,β) and the intensity $\lambda(t) = \mu\beta t^{\beta-1}$. A prior distribution for (μ,β) is needed. We consider three variants of the H-B distribution introduced previously as a natural conjugate priors for the PLP. These variants are such that some of their parameters are depending on each other reducing the number of parameters characterizing the H-B distribution. The elicitation of the hyperparameters will be depending on the available information that we characterized by the *number of guesses*.

Theorem 3.5.4. *– Let $\underline{t} = (t_1, \ldots, t_n)$ be the jump dates of a PLP with intensity $\mu\beta t^{\beta-1}$ observed in a time window $[0,\tau]$. The H-B distribution with parameters (a,b,τ,d,m) is a natural conjugate prior for the PLP. The posterior is a*

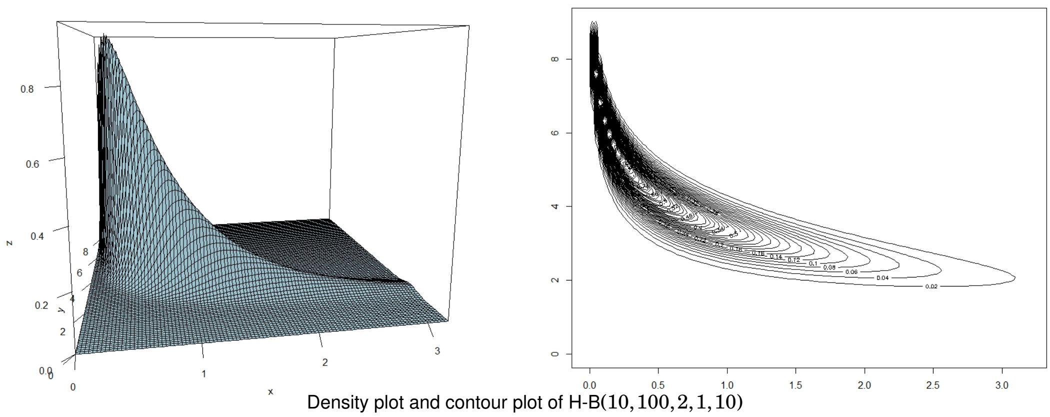
Figure 3.3: *Density functions of H-B distribution*



Density plot and contour plot of H-B$(10, 1, 2, 1, 10)$



Density plot and contour plot of H-B$(10, 10, 2, 1, 10)$



Density plot and contour plot of H-B$(10, 100, 2, 1, 10)$

H-B distribution with parameters $(a+n, bp_n, \tau, d+1, m+n)$. Denote $\tilde{\mu}, \tilde{\beta}$ the Bayesian estimates of μ, β respectively, the Bayesian estimates are:

$$\tilde{\beta} = \frac{a+n}{k+s_n},$$
$$\tilde{\mu} = \left(\frac{m+n}{d+1}\right)\left(\frac{k+s_n}{k+s_n+\log(\tau)}\right)^{a+n}.$$

Proof 7. Recall that the probability density of an observation of the PLP is

$$f(\underline{t} \mid \mu, \beta) \propto (\mu\beta)^n p_n^{\beta} \exp\{-\mu\tau^{\beta}\}$$

where $p_n = \prod_{i=1}^n t_i$.

Proof of (i): Let us consider H-B$(a, b, \tau, 1, a)$ as the prior distribution with density

$$\pi(\mu, \beta) \propto (\mu\beta)^{a-1} b^{\beta} \exp\{-\mu\tau^{\beta}\}.$$

Applying the Bayes' theorem, the posterior distribution is:

$$\begin{aligned} \pi(\mu, \beta \mid \underline{t}) &\propto f(\underline{t} \mid \mu, \beta)\pi(\mu, \beta) \\ &\propto (\mu\beta)^{a+n-1}(bp_n)^{\beta} \exp\{-2\mu\tau^{\beta}\} \end{aligned}$$

That is to say a H-B distribution with parameters $(a+n, bp_n, \tau, 2, a+n)$.

Assuming a quadratic loss, the Bayes estimators are the expectation of the posterior distributions. Since the expectations of a H-B distribution has given in the theorem 3.5.2, we can easily obtain the Bayesian estimates of μ, β:

$$\begin{aligned} \tilde{\beta} = E(\beta \mid \underline{t}) &= \frac{a+n}{\log(\tau^{a+n}/(bp_n))} = \frac{a+n}{\log(\tau^a/b) + \log(\tau^n/p_n)} = \frac{a+n}{k+s_n}, \\ \tilde{\mu} = E(\mu \mid \underline{t}) &= (a+n)\left(\frac{\log(\tau^{a+n}/(bp_n))}{\log(\tau^{a+n}/(bp_n)) + \log(\tau)}\right)^{a+n} \\ &= (a+n)\left(\frac{k+s_n}{k+s_n+\log(\tau)}\right)^{a+n}. \end{aligned}$$

Proof of (ii): Now, we take H-B(a, b, τ, d, a) as the prior distribution

$$\pi(\alpha, \beta) \propto (\mu\beta)^{a-1} b^{\beta} \exp\{-d\mu\tau^{\beta}\}.$$

The posterior distribution is then

$$\begin{aligned} \pi(\mu, \beta \mid \underline{t}) &\propto f(\underline{t} \mid \mu, \beta)\pi(\mu, \beta) \\ &\propto (\mu\beta)^{a+n-1}(bp_n)^{\beta} \exp\{-(d+1)\mu\tau^{\beta}\}. \end{aligned}$$

That is to say a H-B distribution with parameters $(a+n, bp_n, \tau, d+1, a+n)$. Similar reasoning gives us the Bayesian estimates for μ, β as in (ii).

Proof of (iii): Finally, we take H-B(a,b,c,τ,m) as the prior distribution

$$\pi(\alpha,\beta) \propto \lambda^{m-1}\beta^{a-1}b^{\beta}\exp\{-d\mu\tau^{\beta}\}.$$

Hence, the posterior distribution is

$$\begin{aligned}\pi(\mu,\beta \mid \underline{t}) &\propto f(\underline{t} \mid \mu,\beta)\pi(\mu,\beta)\\ &\propto \mu^{m+a-1}\beta^{a+n-1}(bp_n)^{\beta}\exp\{-(d+1)\mu\tau^{\beta}\}\end{aligned}$$

That is to say a H-B distribution with parameters $(a+n, bp_n, \tau, d+1, m+n)$. Similar reasoning gives us the Bayesian estimates for μ,β as in (iii).

□

Relation between Bayesian estimates and maximum likelihood estimates:

With the prior H-B(a,b,τ,d,a), the Bayesian estimates μ,β are:

$$\tilde{\beta} = \frac{a+n}{k+s_n}, \tag{3.12a}$$

$$\tilde{\mu} = \left(\frac{a+n}{d+1}\right)\left(\frac{k+s_n}{k+s_n+\log(\tau)}\right)^{a+n}. \tag{3.12b}$$

whereas maximum likelihood estimates for μ,β are

$$\begin{aligned}\hat{\beta} &= \frac{n}{s_n},\\ \hat{\mu} &= \frac{n}{\tau^{\hat{\beta}}}.\end{aligned}$$

One can see that $\tilde{\beta}$ can be expressed as a convex combination of the MLE and the expectation of the prior distribution:

$$\tilde{\beta} = q_n\hat{\beta} + [1-q_n)E(\beta),$$

where

$$q_n = \frac{s_n}{k+s_n}.$$

This remark will be useful to choose the hyperparameters (a,b,d) in the sequel.

A relationship between $\tilde{\mu}$ and $\hat{\mu}$ can be proposed. From (3.12a) we get $k+s_n = (a+n)/\tilde{\beta}$. Substituting in (3.12b) we get

$$\tilde{\mu} = \frac{a+n}{d+1}\left[1+\frac{\log\left(\tau^{\tilde{\beta}}\right)}{a+n}\right]^{-(a+n)}$$

which can be approximated by:

$$\tilde{\mu} \approx \frac{a+n}{d+1}\frac{1}{\tau^{\tilde{\beta}}}.$$

Therefore $\tilde{\mu}$ can be expressed as a convex combination of the MLE and the prior expectation of μ given $\beta = \tilde{\beta}$:

$$\tilde{\mu} = (1-\xi)\hat{\mu} + \xi \frac{a}{d\tau^{\tilde{\beta}}},$$

where $\xi = \dfrac{d}{d+1}$. This approximation will be used in the next section to elicit prior parameters.

3.5.4 Prior elicitation

We suggest some strategies to elicitate the prior parameters according to provided prior guesses. In each scenario, different prior guesses lead to different prior elicitation strategies. We also consider some strategies relying on the relation between maximum likelihood estimates and maximum likelihood estimates. Some strategies require trials and errors procedure to obtain values for prior parameters..

Scenario 1

In this scenario, we are at disposal of two prior guesses. They might be a guess on the value of β and a guess on the confidence of the first guess (strong confidence, moderate confidence or weak confidence). They might be also a guess on the value of β and a guess on the value of μ. We now employ H-B$(a, b, \tau, 1, a)$ as conjugate prior for the PLP. The values for a, b need to be provided.

Elicitation strategy 1:

Suppose that the practitioner has a guess $g_{\beta,1}$ at the value of β and a guess $g_{\beta,2}$ at the standard deviation associated with $g_{\beta,1}$. The value for a and $k = \log(\tau^a/b)$ can be obtained by solving the system of two equations:

$$\begin{cases} a/k = g_{\beta,1}, \\ \sqrt{a}/k = g_{\beta,2}. \end{cases}$$

We have: $a = g_{\beta,1}^2/g_{\beta,2}$ and $k = a/g_{\beta,1}$. Therefore $b = \tau^a e^{-k}$.

Elicitation strategy 2:

Suppose that the practitioner has a guess g_β at the value of β and a guess g_μ at the value of μ. The value for a, k can be obtained by solving the system of two equations:

$$\begin{cases} a/k = g_\beta, \\ a\left(\dfrac{k}{k+\log(\tau)}\right)^a = g_\mu. \end{cases}$$

From the first equation we get $k = a/g_\beta$ then replace it in the second equation we obtain

$$a\left(\frac{a}{a+g_\beta \log(\tau)}\right)^a = g_\mu.$$

This equation can not be solved explicitly but need trials and errors procedure. One can start at $a_0 = n$. A value for k is then deduced from the value for a. Hence a value for b can be obtained as $b = y^a e^{-k}$.

Scenario 2

In this scenario, we are at disposal of three prior guesses. Besides the prior information in the scenario 1, strategy 1, we are provided one more prior guess g_μ at the value of μ. We now employ H-B(a, b, τ, d, a) as conjugate prior for the PLP. The values for a, b, d need to be provided. The value for a and b can be obtained as in scenario 1, strategy 1. Since the prior expectation of μ is

$$E^\pi(\mu) = \frac{a}{d}\left(\frac{k}{k+\log(\tau)}\right)^a$$

the value for d is obtained by

$$d = \frac{a}{g_{\mu,1}}\left(\frac{k}{k+\log(\tau)}\right)^a.$$

Scenario 3

In this scenario, we are at disposal of four prior guesses. Suppose that the practitioner has a guess $g_{\beta,1}, g_{\mu,1}$ at the value of β, μ and a guess $g_{\beta,2}, g_{\mu,2}$ at the standard deviation associated with $g_{\beta,1}, g_{\mu,1}$ respectively. We now employ H-B(a, b, τ, d, m) as conjugate prior for the PLP. The values for a, b, d, m need to be provided. The value for a and b can be obtained as in scenario 1 and scenario 2. The values for m and d can be found by solving the equations:

$$\frac{m}{d}\left(\frac{k}{k+\log(\tau)}\right)^a = g_{\mu,1},$$
$$\frac{m}{d^2}\left(\frac{k}{k+\log(\tau)}\right)^a = g_{\mu,1}^2 + g_{\mu,2}^2.$$

Then we get

$$m = \frac{g_{\mu,1}^2}{g_{\mu,1}^2 + g_{\mu,2}^2} \frac{[k+\log(\tau)]^{2a}}{k^a\,[k+2\log(\tau)]^a},$$
$$d = \frac{g_{\mu,1}}{g_{\mu,1}^2 + g_{\mu,2}^2} \left[\frac{k+\log(\tau)}{k+2\log(\tau)}\right]^a.$$

Scenario 4

In this situation, we want to apply H-B(a,b,τ,d,a) as conjugate prior for the PLP providing two prior guesses.

Elicitation strategy 1:

We suggest a first strategy to choose the values for three prior parameters a,b,d. Suppose that the practitioner has a guess $g_{\beta,1}$ at the value of β and a guess $g_{\beta,2}$ at the standard deviation associated with $g_{\beta,1}$. Then a value for a can be obtained by solving the system:

$$\begin{cases} a/k = g_{\beta,1}, \\ \sqrt{a}/k = g_{\beta,2}. \end{cases}$$

We have:
$a = \left[g_{\beta,1}/g_{\beta,2}\right]^2$ and $k = a/g_{\beta,1}$. Then (3.12a) can be computed.

According to (3.5.3), $\dfrac{a+n}{n(d+1)}$ can be interpreted as a confidence or corrective factor α associated with the MLE. A value for d can be obtained solving the equation:

$$\frac{n+a}{n(d+1)} = \alpha \text{ to obtain } d = \frac{a+n}{n\alpha} - 1,$$

with $\alpha = q_n$ for example.

Elicitation strategy 2:

A second strategy consists in considering a guess at ξ and a guess at β, g_β. From the guess at ξ, a value for b can be deduced. Setting $n = a/d$, a value for a is obtained. The guess at β provides a value for k since $k = a/g_\beta$. The results using this strategy are displayed in table 2.

3.5.5 Application

Simulation study

In order to investigate the behavior of the H-B natural conjugate prior, we make a comparison between Bayesian estimation and maximum likelihood estimation relying on simulated data from PLP. The table 3.3 describe the results of estimation based on the data generated by the PLP with true parameters $\beta = 2.0, \mu = 0.001$. The sample sizes vary from small size $n = 13$ to medium size $n = 100$ and then to large size $n = 1000$. We repeat the simulation 1000 times to get the mean values of estimation and the mean square of errors (in bracket).

Scenario 4, elicitation strategy 1 is used for choosing the values of the prior hyperparameters. Three different values of prior mean for β are considered: case [1] prior mean underestimates the input value, case [2] prior mean overestimates the input value, and case [3] prior mean is relatively close to the input value. For underestimated prior guess, accurate prior guess and overestimated prior guess, we choose respectively $g_{\beta,1} = 0.9, g_{\beta,1} = 1.9, g_{\beta,1} = 2.9$. For each prior guess $g_{\beta,1}$, computations are carried out using three incertitude values of variability $g_{\beta,2}$ according to the scheme: $g_{\beta,2} = \rho g_{\beta,1}$, where $\rho = 0.3, 0.6, 0.9$ are the coefficient of variation.

With large sample size, it is not surprising that Bayesian estimates are relatively close to the maximum likelihood estimates (MLEs) and the two approach give very good estimates close to the input values of the two parameters whatever the prior guess for β.

With medium sample size, in most of the scenarios of prior guess for β, the Bayesian estimates of both β, μ are more accurate than the MLEs. For example, provided a underestimated guess associated with moderate confidence or weak confidence on the value of β, the Bayesian estimates of both β, μ are more accurate than the MLEs. Provided a overestimated guess on the value of β, the Bayesian estimates of both β, μ are more accurate than the MLEs whatever the associated confidence. Maximum likelihood approach only give better estimate for β than Bayesian approach in case of underestimated guess associated with strong confidence ($g_{\beta,1} = 0.9, g_{\beta,2} = 0.3 g_{\beta,1} = 0.27$).

With small sample size, one can see that the Bayesian approach seems to outperform the maximum likelihood approach in most of the case. Providing a precise-guess for β, the Bayesian estimates of both β, μ are always more accurate than the MLEs. This also happen with a over-guess for β. Only in the case of under-guess associated with strong confidence for β, the MLE is more accurate than the one of Bayesian estimate of β. The small size case is in favor of showing the advantage of Bayesian approach.

The table 3.4 illustrates simulation results for scenario 4, elicitation strategy 2. This time, the PLP is generated with $\beta = 1.38, \mu = 0.0008$. We choose different values for ξ depending on the confidence we might have in the data. We set $\xi = 0.3, 0.6, 0.8, 0.95$.

With small and medium sample size, it turns out that if we are provided a precise guess for β then the Bayesian estimates are better than the MLEs of μ.

With large sample size, the Bayesian estimates of β, μ are good only when provided a precise guess for β. We get bad estimates for β when the true value of β is overestimated or underestimated. We remark that in general, the results obtained by the elicitation strategy 2 are worse than that of elicitation strategy 1. However, for some schemes of prior, Bayesian estimates of β with elicitation strategy 2 are closer to the input value than that with elicitation strategy 1. With strategy 2, we observe more dispersion on Bayesian estimates.

Real data

The table 3.5 gives data that has been discussed many times in the literature [3]. Those are failure times in hours for a complex type of aircraft generator.

The MLE for β and μ are easily obtained: $\tilde{\beta} = 0.5690$ and $\tilde{\mu} = 0.10756$. We compare the MLE with the Bayesian estimates in the table 3.6 for strategy 1 and in table 3.7 for strategy 2.

With elicitation strategy 1, the Bayesian estimate is close to the MLE when the guess on β is 0.5 associated with a small standard deviation. Bayesian estimates with elicitation strategy 2 is unstable whatever is the guess on β provided. Again the only case where the Bayesian estimate of β close to the MLE is when $g_\beta = 0.5$.

3.6 Concluding Remarks

Through out this chapter, we study the PLP and make inference on the process. The adequacy of the PLP model can be verified graphically by plotting the MTBF versus time on log-log scale or by R^2 indicator for a simple regression line.

The power-law form of its intensity form makes the PLP a tractable likelihood and closed-forms of maximum likelihood estimators then some probabilistic properties of the maximum likelihood estimators are deduced. This classical point of view leads to different inference for event truncation and time truncation schemes.

For Bayesian approach, we summary several choices of prior including Jeffrey's rule for non-informative prior and independent conjugate priors. A joint conjugate prior for two parameters of the PLP would allow the dependency between the two parameters and lessen the cumbersome of calculation of posterior distribution. Although the form of the likelihood does not belong to the exponential family, we search a possibility of such a natural conjugate prior by mimicking the form of its likelihood. A new bi-variate distribution naming PLP distribution is introduced based on the seminal paper of Huang and Bier in 1998. This distribution has some good properties facilitating both prior elicitation and posterior calculation. A simulation study is conducted in order to compare the Bayesian estimates with conjugate prior and other prior choices as well as the maximum likelihood estimates. We suggest two strategies that are easy to implement, relying on expert guessing. The results show that the choice of the elicitation strategy is very sensitive. We introduce in

Table 3.3: *Mean of the Bayes estimates with elicitation strategy 1 for simulated data from the PLP with input parameter values* $\beta = 2.0$ *and* $\mu = 0.001$.

Sample-size	Prior guess		Bayes estimates	
n	$g_{\beta,1}$	$g_{\beta,2}$	$\tilde{\beta}$	$\tilde{\mu}$
13	0.9	0.27	1.3498 (0.6607)	0.0521 (0.0572)
		0.54	1.7538 (0.4395)	0.0281 (0.0377)
		0.81	2.0748 (0.4928)	0.0183 (0.0329)
	1.9	0.57	2.1211 (0.3250)	0.0048 (0.0066)
		1.14	2.1894 (0.5642)	0.0122 (0.0216)
		1.71	2.2508 (0.6546)	0.0129 (0.0201)
	2.9	0.87	2.5743 (0.7771)	0.0017 (0.0025)
		1.74	2.4533 (0.7722)	0.0072 (0.0123)
		2.61	2.3774 (0.7327)	0.0102 (0.0177)
	MLE		2.4667 (0.8927)	0.0024 (0.0070)
100	0.9	0.27	1.7933 (0.2582)	0.0071 (0.0088)
		0.54	1.9757 (0.1962)	0.0034 (0.0042)
		0.81	2.0054 (0.1926)	0.0028 (0.0032)
	1.9	0.57	2.0043 (0.1833)	0.0028 (0.0039)
		1.14	2.0524 (0.1941)	0.0021 (0.0021)
		1.71	2.0190 (0.1684)	0.0024 (0.0031)
	2.9	0.87	2.0697 (0.1892)	0.0019 (0.0021)
		1.74	2.0638 (0.2176)	0.0021 (0.0023)
		2.61	2.0853 (0.2684)	0.0025 (0.0037)
	MLE		2.0783 (0.2677)	0.0015 (0.0025)
1000	0.9	0.27	1.9798 (0.0625)	0.0014 (0.0007)
		0.54	1.9962 (0.0637)	0.0012 (0.0006)
		0.81	1.9974 (0.0606)	0.0012 (0.0005)
	1.9	0.57	2.0052 (0.0607)	0.0012 (0.0005)
		1.14	2.0111 (0.0631)	0.0011 (0.0005)
		1.71	2.0044 (0.0663)	0.0012 (0.0006)
	2.9	0.87	2.0096 (0.0655)	0.0011 (0.0005)
		1.74	2.0040 (0.0618)	0.0012 (0.0005)
		2.61	2.0029 (0.0635)	0.0012 (0.0005)
	MLE		2.021 (0.0635)	0.0011 (0.0005)

Table 3.4: *Mean of the Bayesian estimates with elicitation strategy 2 for simulated data from the PLP with input parameter values* $\beta = 1.38$ *and* $\mu = 0.0008$

Sample-size	Prior guess		Bayesian estimates	
n	g_β	ξ	$\tilde{\beta}$	$\tilde{\mu}$
10	0.90	0.30	1.3084	0.0172384
		0.60	1.0856	0.0187112
		0.80	0.9821	0.0205621
		0.95	0.9189	0.0224425
	1.40	0.30	1.5894	0.0080820
		0.60	1.4833	0.0033231
		0.80	1.4350	0.0016563
		0.95	1.4076	0.0009654
	2.10	0.30	1.7735	0.0057058
		0.60	1.8691	0.0008951
		0.80	1.9654	0.0001251
		0.95	2.0617	0.0000164
	MLE		1.4343	0.001604
150	0.90	0.30	1.1988	0.0060447
		0.60	1.0488	0.0177631
		0.80	0.9686	0.0329844
		0.95	0.9162	0.0500529
	1.40	0.30	1.4018	0.0012170
		0.60	1.4000	0.0009408
		0.80	1.3996	0.0008128
		0.95	1.3998	0.0007406
	2.10	0.30	1.5625	0.0003637
		0.60	1.7534	0.0000530
		0.80	1.9103	0.0000106
		0.95	2.0489	0.0000026
	MLE		1.3995	0.001082
2000	0.90	0.30	1.1944	0.0064560
		0.60	1.0475	0.0298960
		0.80	0.9681	0.0687878
		0.95	0.9161	0.1189907
	1.40	0.30	1.3912	0.0008157
		0.60	1.3949	0.0007593
		0.80	1.3974	0.0007268
		0.95	1.3993	0.0007047
	2.10	0.30	1.5447	0.0001643
		0.60	1.7420	0.0000196
		0.80	1.9042	0.0000034
		0.95	2.0474	0.0000007
	MLE		1.3803	0.000834

Table 3.5: *Failure times in hours for aircraft generator*

Failure	Time	Failure	Time
1	55	8	1308
2	166	9	2050
3	205	10	2453
4	341	11	3115
5	488	12	4017
6	567	13	4596
7	731		

Table 3.6: *Bayesian estimates with elicitation strategy 1 for aircraft generator data*

Prior guess		Bayesian estimates	
$g_{\beta,1}$	$g_{\beta,2}$	$\tilde{\beta}$	$\tilde{\mu}$
	0.075	0.3583	0.2561
0.25	0.15	0.4646	0.2642
	0.225	0.5123	0.2457
	0.15	0.5350	0.1054
0.5	0.30	0.5555	0.1730
	0.45	0.5623	0.1959
	0.225	0.6402	0.0604
0.75	0.45	0.5943	0.1441
	0.675	0.5812	0.1797
	MLE	0.5690	0.1072

Table 3.7: *Bayesian estimates with elicitation strategy 2 for aircraft generator data*

Prior guess		Bayesian estimates	
g_β	ξ	$\tilde{\beta}$	$\tilde{\mu}$
0.25	0.30	0.4115	0.5399
	0.60	0.3223	0.9559
	0.80	0.2816	1.2621
	0.95	0.2572	1.4992
0.5	0.30	0.5464	0.2120
	0.60	0.5255	0.2041
	0.80	0.5124	0.1981
	0.95	0.5031	0.1934
0.75	0.30	0.6134	0.1355
	0.60	0.6653	0.0735
	0.80	0.7051	0.0439
	0.95	0.7383	0.0277
	MLE	0.5690	0.1072

this work a 4-parameter H-B distribution. More investigations concerning the properties of this distribution need to be carried out. In particular a better understanding of the properties will be helpful to elicit prior parameters. One can develop a 5-parameter H-B distribution that allows more prior information to be integrated.

More need to be done in order to improve the accuracy of the estimates. Other strategies should be investigated. We are working in this direction in the present time.

Chapter 4

Exponential-Law Process

In this chapter, we study the exponential-law process which is a non-homogeneous Poisson process with intensity function depending on time by an exponential-law. Despite its simple form of intensity and its tractable likelihood, the maximum likelihood estimation can not be obtained by explicit expressions and requires numerical approximation. For a Bayesian approach of the ELP inference, we consider different type of priors: non-informative, independent conjugate priors. In particular, we are interested in the natural conjugate prior of the ELP by mimicking its functional form of likelihood function.

4.1 Introduction

Huang & Bier [28] consider a Poisson process with a exponential-law intensity. We call this process the *exponential-law process* (ELP). The intensity function is expressed as $\lambda(t) = \alpha e^{\beta t}$, $\alpha > 0$, $\beta \in \mathbb{R}$. It can be used to model both improving and aging repairable systems. This model is close to the Goel-Okumoto model [21]. We are going to investigate this process in details following the same roadmap as for the PLP, allowing us to make comparison.

In section 1, we give a short introduction of the exponential-law process. Section 2 is devoted to maximum likelihood procedure of two parameters of the ELP. In section 3, we define a bivariate distribution that we name the *ELP distribution*. Properties of this distribution are given and it is shown that this distribution is a natural conjugate prior for Bayesian analysis of the ELP. The Bayesian estimates are then obtained and we suggest some strategies to elicit the parameters of the prior distribution. Practitioners need to provide only some guesses on values of the two parameters.

Simulated data sets are used to examine the Bayesian estimates using ELP distribution and the results are then compared with other priors. Our method is applied on simulated data sets to examine the behavior of Bayesian estimates in different scenarios of prior guess and prior confidence. The results show some advantages of our strategies comparing to the MLE in case of small sample size. We then conduct our method on some real data sets that have been addressed in the literature for comparison.

4.2 Maximum Likelihood Method

4.2.1 Likelihood

Let $\{t_1,\ldots,t_n\}$ be a sequence of jumps in an interval $[0,\tau]$ of the ELP.
Denote $s_n = \sum_{i=1}^n t_i$, the likelihood is:

$$L(\alpha,\beta) = \alpha^n \exp\left\{\beta s_n - \frac{\alpha}{\beta}\left(e^{\beta\tau}-1\right)\right\}$$

where $\tau = t_n$ for event truncation and $\tau = C$ for time truncation. The log-likelihood is

$$\log L(\alpha,\beta) = n\log\alpha + \beta s_n - \frac{\alpha}{\beta}\left(e^{\beta\tau}-1\right).$$

4.2.2 Maximum likelihood estimation

The likelihood equations are:

$$\begin{cases} \dfrac{n}{\alpha} - \dfrac{e^{\beta\tau}-1}{\beta} = 0 & (4.1)\\ s_n - \dfrac{\alpha}{\beta^2}\left[(\beta\tau-1)e^{\beta\tau}+1\right] = 0 & (4.2) \end{cases}$$

From (4.1), we have:

$$\alpha = \frac{n\beta}{e^{\beta\tau}-1}, \tag{4.3}$$

and injecting (4.3) in (4.2), we obtain an equation that only depends on β:

$$\beta - \frac{e^{\beta\tau}-1}{(\tau-\bar{t})e^{\beta\tau}+\bar{t}} = 0. \tag{4.4}$$

where $\bar{t} = s_n/n$.
There is no explicit solution to equation (4.4). We use a numerical method to solve it. Let us denote φ the function defined by:

$$\varphi(\beta) = \beta - \frac{e^{\beta\tau}-1}{(\tau-\bar{t})e^{\beta\tau}+\bar{t}},\ \beta\in\mathbb{R}.$$

The derivative of this function is

$$\varphi'(\beta) = 1 - \frac{\tau^2 e^{\beta\tau}}{\left[(\tau-\bar{t})e^{\beta\tau}+\bar{t}\right]^2}.$$

It vanishes for $\beta_1 = 0$ and for $\beta_2 = \dfrac{2}{\tau}\log\left(\dfrac{\bar{t}}{\tau-\bar{t}}\right)$. Remark that $\bar{t} < \tau$.

If $\bar{t} < \tau/2$, then $\beta_2 > 0$ and $\varphi(\beta_2) < 0$.
We have $\lim_{\beta\to-\infty}\varphi(\beta) = -\infty$, $\lim_{\beta\to+\infty}\varphi(\beta) = +\infty$ and $\varphi(0) = 0$. φ is decreasing for all β in $[0,\beta_2]$ and increasing for β in $[\beta_2,+\infty[$.

Therefore there exists a unique solution to the equation $\varphi(\beta) = 0$ which can be obtained applying a Newton-Raphson algorithm with starting point β_2. If $\bar{t} > \tau/2$, using a similar reasoning we end up with the same conclusion.

4.3 Bayesian Approach

For Bayesian analysis, one can consider different choices of priors such as non-informative prior, conjugate prior. Let us consider non-informative prior by applying the Jeffreys' rule.

4.3.1 Fisher information matrix

We consider in this section the construction of non-informative prior for Bayesian analysis of the exponential-law process using the Jeffreys' rule.

Event truncation

Denote $m_0 = E\left(e^{\beta T_n}\right)$, $m_1 = E\left(T_n e^{\beta T_n}\right)$ and $m_2 = E\left(T_n^2 e^{\beta T_n}\right)$ then the Fisher matrix information is

$$\begin{aligned}
I_{11}(\alpha,\beta) &= -\frac{n\beta}{\alpha^2},\\
I_{22}(\alpha,\beta) &= -\frac{2\alpha}{\beta^3}E\left(e^{\beta T_n}-1\right)+\frac{2\alpha}{\beta^2}E\left(T_n e^{\beta T_n}\right)-\frac{\alpha}{\beta}E\left(T_n^2 e^{\beta\tau}\right),\\
I_{12}(\alpha,\beta) = I_{21}(\alpha,\beta) &= \frac{1}{\beta^2}E\left(e^{\beta T_n}-1\right)-\frac{1}{\beta}E\left(T_n e^{\beta\tau}\right).
\end{aligned}$$

Now we need to calculate m_0, m_1, m_2. Firstly, we find the distribution of last jump T_n. The probability density function of the last jump is

$$\begin{aligned}
f_{T_n}(t) &= \frac{\lambda(t)[\Lambda(t)]^{n-1}}{\Gamma(n)}\exp\{-\Lambda(t)\}\\
&= \frac{\alpha^n}{\beta^{n-1}\Gamma(n)}e^{\beta t}\left(e^{\beta t}-1\right)^{n-1}\exp\left\{-\alpha\frac{e^{\beta t}-1}{\beta}\right\}.
\end{aligned}$$

Denote $U_n = \alpha\left(e^{\beta T_n}\right)/\beta$. The pdf of U_n is then

$$f_{U_n}(u) = \frac{1}{\Gamma(n)}u^{n-1}e^{-u}.$$

Hence U_n has gamma distribution with parameter $(n,1)$. Since $E(U_n) = n$ and $E(U_n^2) = n(n+1)$ we have

$$m_0 = E\left(e^{\beta T_n}\right) = \frac{\beta}{\alpha}E(U_n) + 1 = \frac{n\beta}{\alpha} + 1.$$

Note that $\beta T_n \le e^{\beta T_n} - 1$ for $\beta \in \mathbb{R}$ so $T_n \ge U_n/\beta$ and $T_n e^{\beta T_n} \ge (U_n + U_n^2)/\beta$. Transfer the inequality into expectation we obtain

$$m_0 = E\left(T_n e^{\beta T_n}\right) \le \frac{E(U_n^2 + U_n)}{\beta} = \frac{n(n+1)\beta}{\alpha^2} + \frac{n}{\alpha}.$$

Hence

$$0 \le m_0 \le \frac{n(n+1)\beta}{\alpha^2} + \frac{n}{\alpha}.$$

Similarly

$$0 \le m_1 \le m_0 \le \frac{n(n+1)\beta}{\alpha^2} + \frac{n}{\alpha}.$$

The Fisher matrix information is then

$$\begin{aligned}
I_{11}(\alpha,\beta) &= -\frac{n\beta}{\alpha^2},\\
I_{22}(\alpha,\beta) &= \frac{2n - 2\alpha m_1 + \alpha\beta m_2}{\beta^2},\\
I_{12}(\alpha,\beta) = I_{21}(\alpha,\beta) &= \frac{n - \alpha m_1}{\alpha\beta}.
\end{aligned}$$

The determinant of the Fisher information matrix is

$$\det[I(\alpha,\beta)] = \frac{n^2 + n\alpha\beta m_2 - \alpha^2 m_1^2}{\alpha^2\beta^2}.$$

Thus the Jeffreys non-informative prior for (α,β) is

$$\pi(\alpha,\beta) \propto (\alpha\beta)^{-1}.$$

Time truncation

In this case, $N(C)$ is a random variable having Poisson distribution with parameter $\Lambda(C) = \alpha\left(e^{\beta C} - 1\right)/\beta$ thus

$$\begin{aligned}
I_{11}(\alpha,\beta) &= \frac{1}{\alpha^2}E[N(C)] = \frac{e^{\beta C} - 1}{\alpha\beta},\\
I_{22}(\alpha,\beta) &= \frac{2\alpha\left(e^{\beta C} - 1\right) - 2\alpha\beta C e^{\beta C} + \alpha\beta^2 C^2 e^{\beta C}}{\beta^3},\\
I_{12}(\alpha,\beta) &= \frac{\beta C e^{\beta C} - e^{\beta C} + 1}{\beta^2}.
\end{aligned}$$

The determinant of the Fisher information matrix is therefore

$$\det[I(\alpha,\beta)] = \frac{e^{2\beta C} + (\beta^2C^2+2)e^{\beta C}+1}{\beta^4} = \frac{1}{\beta^4}\left[\left(e^{\beta C}-1\right)^2 - \beta^2C^2e^{\beta C}\right].$$

At time C we have $\Lambda(C) = \alpha\left(e^{\beta C}-1\right)/\beta \approx n$. Replace $\alpha\left(e^{\beta C}-1\right)/\beta$ by n we have

$$\det[I(\alpha,\beta)] \approx \frac{n^2 - n\alpha\beta - \alpha^2C^2}{\alpha^2\beta^2}.$$

Hence Jeffreys non-informative prior for (α,β) is

$$\pi(\alpha,\beta) \propto (\alpha\beta)^{-1}.$$

4.3.2 Non-informative prior

We consider Bayesian inference for the ELP with Jeffreys non-informative prior. While the maximum likelihood estimates are interpreted differently according to event truncation or time truncation scheme, Bayesian estimates give the same interpretation for both schemes.

Theorem 4.3.1. – *Let $\underline{t} = (t_1,\ldots,t_n)$ a realization of the ELP that we observe in the time window $[0,C]$. Denote $\tilde{\alpha},\tilde{\beta}$ the Bayesian estimates of α,β. With Jeffreys non-informative prior $\pi(\alpha,\beta) \propto (\alpha\beta)^{-1}$ and assuming quadratic loss, we obtain the following results:*

(i) The posterior density is

$$\pi(\alpha,\beta \mid \underline{t}) = \frac{\tau^{n+1}}{\Gamma(n)I(s_n/\tau,n)}\alpha^{n-1}\beta^{-1}\exp\left\{\beta s_n - \alpha\left(e^{\beta\tau}-1\right)/\beta\right\}$$

where

$$I(\alpha,\beta) = \int_0^\infty u^{\alpha-1}\left[\frac{\log(u)}{u-1}\right]^\beta, \quad 1 \le \alpha \le \beta - 1.$$

(ii) The posterior marginal distribution of β, say $\beta \mid \underline{t}$ has probability density function

$$\pi(\beta \mid \underline{t}) = \frac{\tau^{n+1}}{I(s_n/\tau,n)}\left(\frac{\beta}{e^{\beta\tau}-1}\right)^n e^{\beta s_n}.$$

(iii) The posterior conditional density of α given β, say $\alpha \mid \beta,\underline{t}$, belongs to gamma family with parameter $(n,\left(e^{\beta\tau}-1\right)/\beta)$.

(iv) The Bayesian estimates are:

$$\tilde{\beta} = \frac{1}{\tau}\frac{J_1(s_n/\tau,n)}{J_0(s_n/\tau,n)},$$
$$\tilde{\alpha} = \frac{n}{\tau}\frac{I(s_n/\tau,n+1)}{I(s_n/\tau,n)},$$

where

$$J_m(\alpha,\beta) = \int_0^\infty u^{\alpha-1}[\log(u)]^m\left[\frac{\log(u)}{u-1}\right]^\beta, \quad 1 \le \alpha \le \beta - 1.$$

Proof of (i): Recall that the probability density of a observation of the ELP is

$$f(\underline{t} \mid \alpha,\beta) = \alpha^n \exp\left\{\beta s_n - \frac{\alpha}{\beta}\left(e^{\beta\tau}-1\right)\right\}.$$

With prior density $\pi(\alpha,\beta) \propto (\alpha\beta)^{-1}$, and applying the Bayes'theorem, we obtain the posterior density as

$$\pi(\alpha,\beta \mid \underline{t}) = K(\underline{t})\alpha^{n-1} \exp\left\{\beta s_n - \frac{\alpha}{\beta}\left(e^{\beta\tau}-1\right)\right\}.$$

The normalizing constant $K(\underline{t})$ can be computed by taking double-integral and applying Fubini's theorem

$$\begin{aligned}
K(\underline{t})^{-1} =& \int_0^{\infty}\int_{-\infty}^{+\infty} \alpha^{n-1} \exp\left\{\beta s_n - \frac{\alpha}{\beta}\left(e^{\beta\tau}-1\right)\right\} d\alpha d\beta \\
=& \int_{-\infty}^{+\infty} e^{\beta s_n} \left(\int_0^{\infty} \alpha^{n-1} \exp\left\{-\frac{\alpha}{\beta}\left(e^{\beta\tau}-1\right)\right\} d\alpha\right) d\beta \\
=& \int_{-\infty}^{+\infty} e^{\beta s_n} \frac{\Gamma(n)}{\left[\left(e^{\beta\tau}-1\right)/\beta\right]^n} d\beta \\
=& \Gamma(n)\int_{-\infty}^{+\infty} e^{\beta s_n} \left(\frac{\beta}{e^{\beta\tau}-1}\right)^n d\beta \\
=& \frac{\Gamma(n)}{\tau^{n+1}}\int_0^{\infty} u^{s_n/\tau-1}\left[\frac{\log(u)}{u-1}\right]^n du \\
=& \frac{\Gamma(n)I(s_n/\tau,n)}{\tau^{n+1}}.
\end{aligned}$$

Hence the normalizing constant $K(\underline{t})$ is given by

$$K(\underline{t}) = \frac{\tau^{n+1}}{\Gamma(n)I(s_n/\tau,n)}.$$

Proof of (ii): By integrating out α from the posterior joint density, we obtain the posterior marginal density of β as

$$\begin{aligned}
\pi(\beta \mid \underline{t}) =& \int_0^{\infty} \frac{\tau^{n+1}}{\Gamma(n)J_0(s_n/\tau,n)} \alpha^{n-1} \exp\left\{\beta s_n - \frac{\alpha}{\beta}\left(e^{\beta\tau}-1\right)\right\} d\alpha \\
=& \frac{\tau^{n+1}}{I(s_n/\tau,n)}\left(\frac{\beta}{e^{\beta\tau}-1}\right)^n e^{\beta s_n}.
\end{aligned}$$

This density belongs to a family of distributions that we will study in detail in the next section. We name it modified-Gumbel distribution (M-G distribution in short).

Proof of (iii): The posterior conditional density of α given β is

$$\pi(\alpha \mid \beta,\underline{t}) = \frac{\pi(\alpha,\beta \mid \underline{t})}{\pi(\beta \mid \underline{t})} = \left(\frac{e^{\beta\tau}-1}{\beta}\right)^n \frac{1}{\Gamma(n)}\alpha^{n-1}\exp\left\{-\alpha\left(\frac{e^{\beta\tau}-1}{\beta}\right)\right\}.$$

That is, α given $\beta,\underline{t}$ has a gamma distribution with parameter $(n,\left(e^{\beta\tau}-1\right)/\beta)$.

Proof of (iv): Assuming a quadratic loss, the Bayesian estimators are the expectations of the posterior marginal distributions. Since $\beta \mid \underline{t}$ has M-G distribution with parameter (n, s_n, τ), one can compute its expectation and obtain

$$\tilde{\beta} = E(\beta \mid \underline{t}) = \frac{\alpha}{\tau} \frac{J_1(s_n/\tau, n)}{J_0(s_n/\tau, n)}.$$

Thus we get Bayesian estimate for β as in (iv).
The posterior conditional expectation of α given β is

$$E(\alpha \mid \beta, \underline{t}) = \frac{n\beta}{\tau e^{\beta\tau} - 1}.$$

The posterior expectation of μ is obtained by taking expectation of posterior conditional expectation of α given β.

$$\begin{aligned}
\tilde{\alpha} = E(\alpha \mid \underline{t}) = E(E(\alpha \mid \beta, \underline{t})) &= \int_{-\infty}^{+\infty} E(\alpha \mid \beta, \underline{t}) \pi(\beta \mid \underline{t}) d\beta \\
&= \int_{-\infty}^{+\infty} \frac{n\beta}{\tau e^{\beta\tau} - 1} \frac{\tau^{n+1}}{I(s_n/\tau, n)} \left(\frac{\beta}{e^{\beta\tau} - 1} \right)^n e^{\beta s_n} d\beta \\
&= \frac{n\tau^{n+1}}{I(s_n/\tau, n)} \int_{-\infty}^{+\infty} \left(\frac{\beta}{e^{\beta\tau} - 1} \right)^{n+1} e^{\beta s_n} d\beta \\
&= \frac{n\tau^{n+1}}{I(s_n/\tau, n)} \frac{1}{\tau^{n+2}} \int_0^{+\infty} u^{s_n/\tau - 1} \left[\frac{\log(u)}{u-1} \right]^{n+1} du \\
&= \frac{n}{\tau} \frac{I(s_n/\tau, n+1)}{I(s_n/\tau, n)}.
\end{aligned}$$

Hence we get Bayesian estimate for α as in (iv).
Recall that the maximum likelihood estimate $\hat{\alpha}$ is the solution of the equation

$$\beta - \frac{e^{\beta\tau} - 1}{(\tau - \bar{t})e^{\beta\tau} + \bar{t}} = 0$$

and $\hat{\alpha}$ is derived from $\hat{\alpha}$ as

$$\hat{\alpha} = \frac{n\hat{\beta}}{e^{\hat{\beta}\tau} - 1}.$$

The classical result is that the Bayesian estimate with non-informative prior is equal or closed to the maximum likelihood estimate. Therefore $(\tilde{\alpha}, \tilde{\beta})$ can give us a good initial value to obtain the numerical approximation value for $(\hat{\alpha}, \hat{\beta})$ when applying an approximation method such as Newton-Raphson or other package in R software.

It can be shown that $J_m(\alpha, \beta)$ converges when $(1 \leq \alpha \leq \beta - 1)$ (see Appendix) and note that $I(\alpha, \beta) = J_0(\alpha, \beta)$. Thus the above definitions are valid.

□

4.3.3 Modified-Gumbel distribution

We introduce a new family of uni-variate distributions that we name modified-Gumbel distribution (M-G distribution in short) for Bayesian analysis of the ELP. A M-G distribution have four parameters. It requires numerical integrals to compute the expectations and its variances. Firstly, we give definition to this distribution.

M-G distribution

Definition 4.3.1. – *A random variable $X \in \mathbb{R}$ has a M-G distribution with parameters (a,b,c,m) where $a,b,c > 0, m \in \mathbb{Z}$ and such that $c \le b \le (a-1)c$, if it has a probability density function (p.d.f) of the form:*

$$f_X(x) = K\, x^m \left(\frac{x}{e^{cx}-1}\right)^a e^{bx}.$$

The normalizing constant K can be computed as

$$K = \frac{c^{a+m+1}}{J_m(b/c, a)}$$

where

$$J_m(\alpha,\beta) = \int_0^\infty u^{\alpha-1}[\log(u)]^m \left[\frac{\log(u)}{u-1}\right]^\beta, \quad 1 \le \alpha \le \beta - 1.$$

We denote: $X \sim M\text{-}G(a,b,c,m)$.

Proof of validity of M-G distribution:
The normalizing constant K can be computed as

$$K^{-1} = \int_{-\infty}^{+\infty} x^m \left(\frac{x}{e^{cx}-1}\right)^a e^{bx} dx.$$

By changing variable as $u = e^{cx}$ we get

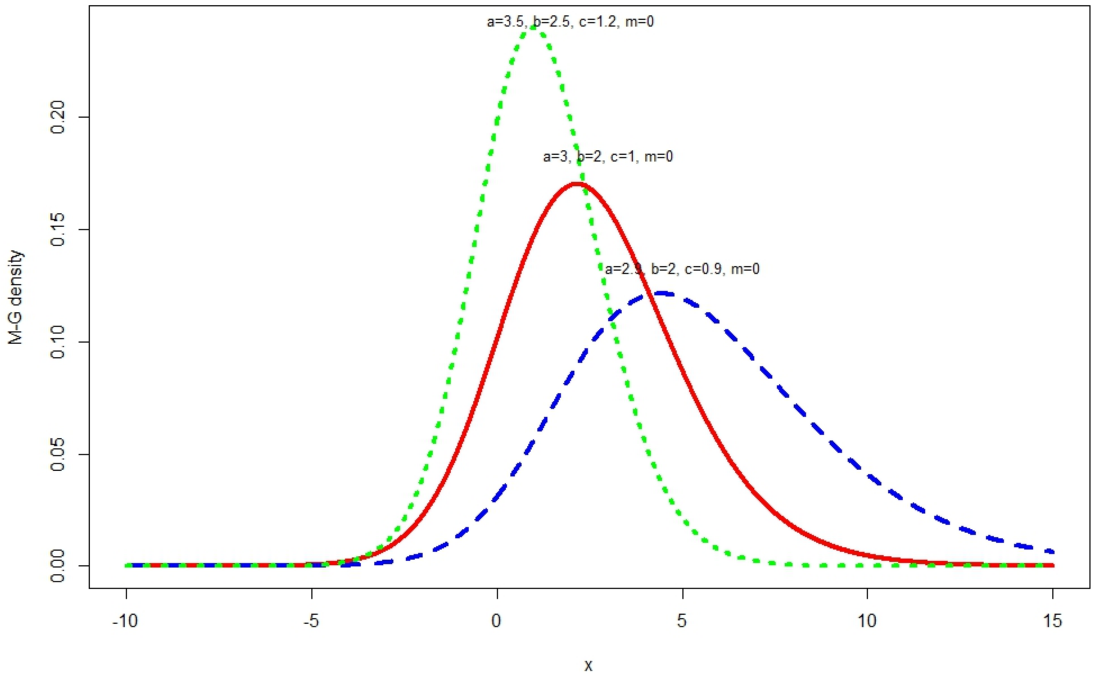
$$K^{-1} = \frac{1}{c^{a+m+1}} \int_0^\infty u^{b/c-1}[\log(u)]^m \left[\frac{\log(u)}{u-1}\right]^a dx = \frac{J_m(b/c,a)}{c^{a+m+1}}$$

where

$$\Gamma(a) = \int_0^\infty x^{a-1}e^{-x}dx,$$
$$J_m(\alpha,\beta) = \int_0^\infty x^{\alpha-1}\log^m(x)\left(\frac{\log(x)}{x-1}\right)^\beta dx.$$

The figure 4.1 displays the p.d.f. of M-G distributions with different parameters.

Figure 4.1: *P.d.f. of M-G distributions with different parameters values*



Expectation and variance

The expectation and variance of M-G distributions can be obtained but it requires numerical integrals.

Theorem 4.3.2. *Let $X \sim M\text{-}G(a,b,c,m)$. The expectation and variance of X always exist and are given as following:*

$$E(X) = \frac{1}{c}\frac{J_{m+1}(b/c,a)}{J_m(b/c,a)},$$
$$Var(X) = \frac{1}{c^2}\frac{J_m(b/c,a)J_{m+2}(b/c,a) - J^2_{m+1}(b/c,a)}{J_m(b/c,a)}.$$

Proof 8. Using the same technique for computing the constant K we can compute the expectation of X as

$$\begin{aligned}
E(X) &= \int_{-\infty}^{+\infty} x \frac{c^{a+1}}{J_m(b/c,a)} x^m \left(\frac{x}{e^{cx}-1}\right)^a e^{bx} \\
&= \frac{c^{a+1}}{J_m(b/c,a)} \frac{1}{c^{a+2}} \int_0^{\infty} u^{b/c-1} [\log(u)]^{m+1} \left[\frac{\log(u)}{u-1}\right]^a du \\
&= \frac{1}{c}\frac{J_{m+1}(b/c,a)}{J_m(b/c,a)}
\end{aligned}$$

and the second moment of X as

$$
\begin{aligned}
E(X^2) &= \int_{-\infty}^{+\infty} x^2 \frac{c^{a+1}}{J_m(b/c,a)} x^m \left(\frac{x}{e^{cx}-1}\right)^a e^{bx} \\
&= \frac{c^{a+1}}{J_m(b/c,a)} \frac{1}{c^{a+3}} \int_0^{\infty} u^{b/c-1} [\log(u)]^{m+2} \left[\frac{\log(u)}{u-1}\right]^a du \\
&= \frac{1}{c^2} \frac{J_{m+2}(b/c,a)}{J_m(b/c,a)}.
\end{aligned}
$$

Hence the variance and the standard deviation of X are:

$$
\begin{aligned}
Var(X) &= \frac{1}{c^2} \frac{J_m(b/c,a) J_{m+2}(b/c,a) - J^2_{m+1}(b/c,a)}{J_m(b/c,a)}, \\
Sd(X) &= \frac{1}{c} \sqrt{\frac{J_m(b/c,a) J_{m+2}(b/c,a) - J^2_{m+1}(b/c,a)}{J_m(b/c,a)}}.
\end{aligned}
$$

$\square$

Mode

We show that the M-G distribution with parameter $(a,b,c,0)$ is an uni-modal distribution. It requires numerical approximation to calculate its mode as in the following theorem.

Theorem 4.3.3. – *Let $X \sim M\text{-}G(a,b,c,0)$. The probability density function has the unique mode x_{mod} that is the solution of the equation*

$$x = \frac{a(e^{cx}-1)}{(ac-b)e^{cx}+b}.$$

When $c > 0$, x_{mod} is bounded as

$$\frac{2}{c}\log\left(\frac{b}{ac-b}\right) < x_{mod} < \frac{a}{ac-b}$$

whereas $c < 0$, x_{mod} is bounded as

$$\frac{-a}{b} < x_{mod} < \frac{2}{c}\log\left(\frac{b}{ac-b}\right).$$

Proof 9. Taking logarithm of the p.d.f. of M-G distribution we have

$$\log(f_X(x)) = \log(K) + a\log(x) - a\log\left(e^{cx}-1\right) + bx$$

Taking derivative of that function gives us

$$\frac{\partial}{\partial x}\log(f_X(x)) = \frac{a}{x} - \frac{ace^{cx}}{e^{cx}-1} + b.$$

The critical value of $\log(f_X(x))$ satisfies the the equation

$$x - \frac{a(e^{cx}-1)}{(ac-b)e^{cx}+b} = 0. \tag{4.5}$$

Denote

$$g(x) = x - \frac{a(e^{cx}-1)}{(ac-b)e^{cx}+b}$$

then

$$g'(x) = 1 - \left[\frac{ace^{cx/2}}{(ac-b)e^{cx}+b}\right]^2.$$

$g(x)$ has two critical values $x_1 = 0$ and

$$x_2 = \frac{2}{c}\log\left(\frac{b}{ac-b}\right).$$

Moreover, the equivalent equation of equation (4.5) is

$$e^{cx} = \frac{a+bx}{a-(ac-b)x}.$$

That gives the lower bound and upper bound for x_{mod} as in the theorem.

□

4.3.4 Independent conjugate priors

Following the same method presented in the previous section for the PLP, we study a possibility of an independent conjugate priors for the ELP. Consider a time truncation data $\underline{t} = (t_1,\ldots,t_n)$ observed in the fixed time window $[0,C]$ of a ELP with compensator $\Lambda(t) = \alpha\left(e^{\beta t}-1\right)/\beta$. Let $\eta = \Lambda(C) = \alpha\left(e^{\beta C}-1\right)/\beta$. The following theorem gives us Bayesian estimates for the ELP with two natural independent conjugate priors for each parameter.

Theorem 4.3.4. *– Let $\underline{t} = (t_1,\ldots,t_n)$ a realization of the ELP with parameter (η,β) that we observe in the time window $[0,C]$. Denote $\tilde{\eta},\tilde{\beta}$ the Bayesian estimates of η,β respectively. A natural joint conjugate prior for (η,β) is a product of two independent prior for each parameter where:*

(i) A natural conjugate for η is a gamma distribution and a natural conjugate for β is a M-G distribution with respective p.d.f.:

$$\pi(\beta) = \frac{C^{a+1}}{I(b/C,a)}\left(\frac{\beta}{e^{\beta C}-1}\right)^a e^{\beta b},$$
$$\pi(\eta) = \frac{1}{\Gamma(a}\eta^{k-1}e^{-\ell\eta}.$$

(ii) *$\beta \mid \underline{t}$ and $\eta \mid \underline{t}$ are independent; $\beta \mid \underline{t}$ has a M-G distribution with parameter $(a+n, b+s_n, C, 0)$ and $\eta \mid \underline{t}$ has a gamma distribution with parameter $(k+n, \ell+1)$.*

(iii) *The Bayesian estimates for β and η are:*

$$\tilde{\beta} = \frac{1}{C}\frac{J_1((b+s_n)/C, a+n)}{J_0((b+s_n)/C, a+n)},$$
$$\tilde{\eta} = \frac{k+n}{\ell+1}.$$

Proof of (i): The probability density of an observation of the ELP with parameters (η, β) becomes

$$f(\underline{t} \mid \eta, \beta) = \eta^n e^{-\eta} \times \left(\frac{\beta}{e^{\beta C}-1}\right)^n e^{\beta s_n}.$$

Therefore η and β are orthogonal and the natural joint conjugate prior is a product of two independent conjugate priors for each parameter

$$\pi(\eta, \beta) \propto \eta^{k-1} e^{-\ell\eta} \times \left(\frac{\beta}{e^{\beta C}-1}\right)^a e^{\beta b}.$$

Proof of (ii): Applying the Bayes' theorem, the posterior distribution is

$$\pi(\eta, \beta \mid \underline{t}) \propto \eta^{k+n-1} e^{-(\ell+1)\eta} \times \left(\frac{\beta}{e^{\beta C}-1}\right)^{a+n} e^{\beta(b+s_n)}.$$

Hence we obtain the result in (ii).

Proof of (iii): Assuming a quadratic loss, the Bayesian estimate is the posterior expectation of the parameter. The posterior expectation for η is easily obtained since the posterior distribution is a gamma distribution. The posterior expectation for β is also available as set up in the previous section for M-G distribution.

□

Prior elicitation

Let $g_{\beta,1}$ is a guess for the value of β and $g_{\beta,2}$ is a guess for standard deviation associated with $g_{\beta,1}$; Let $g_{\eta,1}$ is a guess for the value of η and $g_{\eta,2}$ is a guess for standard deviation associated with $g_{\eta,1}$.

Prior elicitation for prior parameter of η is simple since the expectation and the variance of a gamma distribution are available in closed-form expressions. One can easily obtained values for k, ℓ as following:

$$\frac{k}{\ell} = g_{\eta,1}, \quad \frac{k}{\sqrt{\ell}} = g_{\eta,2}.$$

Therefore we get

$$k = \frac{g_{\eta,2}^2}{g_{\eta,1}}, \quad \ell = \left(\frac{g_{\eta,2}}{g_{\eta,1}}\right)^2.$$

Prior elicitation for prior parameter of β needs trials and errors. We have

$$\begin{aligned} \frac{1}{C}\frac{J_1(b/C,a)}{J_0(b/C,a)} &= g_{\beta,1}, \\ \frac{1}{C}\sqrt{\frac{J_0(b/C,a)J_2(b/C,a) - J_1^2(b/C,a)}{J_0(b/C,a)}} &= g_{\eta,2}. \end{aligned}$$

One can start trying at $a = n$ then vary b such that

$$\begin{aligned} \frac{J_1(b/C,n)}{J_0(b/C,n)} &= C g_{\beta,1}, \\ \frac{J_0(b/C,n)J_2(b/C,n) - J_1^2(b/C,n)}{J_0(b/C,n)} &= C^2 g_{\eta,2}^2. \end{aligned}$$

4.4 Conjugate Prior: the G-M-G Distribution

For Bayesian approach, we seek a possibility of a natural conjugate prior for the ELP. Huang and Bier in the paper [28] propose a family distribution that they consider as the natural conjugate prior for the ELP. In fact, the posterior distribution is different from the prior distribution so it requires numerical method to obtain the Bayesian estimates. In addition, the prior elicitation is not practical for application. We introduce a new family of bi-variate distributions that we name Gamma-modified-Gumbel (G-M-G distribution). The G-M-G prior makes it possible to facilitate the dependence between the two parameters of the ELP by Bayesian point of view. This family of distribution is a natural conjugate prior for the ELP since the posterior has also G-M-G distribution. We propose some elicitation strategies for parameters of the conjugate prior. Simulation study is conducted for comparing the maximum likelihood estimation and Bayesian estimation using this conjugate prior.

4.4.1 Prior information and conjugate priors

Following the same approach as in the previous chapter on the PLP, we now consider Bayesian inference for the ELP with the parametrization (α, β) and the intensity $\lambda(t) = \mu\beta t^{\beta-1}$. The probability density of an observation $\underline{t} = (t_1, \ldots, t_n)$ in the time window $[0, C]$ of the ELP with parameter (α, β) is

$$f(\underline{t} \mid \alpha, \beta) \propto \alpha^n \exp\left\{\beta s_n - \alpha\left(e^{\beta\tau} - 1\right)/\beta\right\}$$

where $y = t_n$ for event truncated data and $y = C$ for time truncated data.

Mimicking the above functional form, the prior density should be in the form

$$\pi(\alpha,\beta) \propto \alpha^{a-1}\exp\left\{\beta b - \alpha\left(e^{\beta\tau}-1\right)/\beta\right\}.$$

This prior density belong to a new bivariate distribution with three parameters (a,b,τ). Since the last parameter is fixed to be τ it remains two parameter a,b to be elicitated. It requires, for instance, a guess on the value of β and a guess on the value of α.

When having three prior guesses including a guess on the value of β and a guess on the confidence of the first guess, an a guess on the value of α, it requires a conjugate prior with four parameters (a,b,τ,d) that allows three free parameters to be elicitated. That prior could be of the form

$$\pi(\mu,\beta) \propto \alpha^{a-1}\exp\left\{\beta b - d\alpha\left(e^{\beta\tau}-1\right)/\beta\right\}.$$

Finally, in case we have four prior guesses on expectations and standard deviations of both β and α then a candidate of the natural conjugate prior should have the density of the form

$$\pi(\mu,\beta) \propto \alpha^{a-1}\beta^{m}\exp\left\{\beta b - d\alpha\left(e^{\beta\tau}-1\right)/\beta\right\}.$$

It is a bi-variate distribution with five parameters (a,b,y,τ,m) that allows four free parameters to integrate the four prior guesses.

All of the three forms potential conjugate priors will be investigate in the following section.

4.4.2 G-M-G distribution

We now introduce and study G-M-G distribution with five parameters. Firstly, we give definition this bivariate distribution.

G-M-G distribution

Definition 4.4.1. *– A bivariate r.v.* $(X,Y)\in\mathbb{R}^+\times\mathbb{R}$ *is said to be distributed as a G-M-G distribution with five parameter* (a,b,c,d,m)*, where* $a\geq 2; b,c,d>0, m\in\mathbb{N}$ *such that* $c\leq b\leq c(a-1)$*, if it has a p.d.f. of the form:*

$$f_{X,Y}(x,y) = K\, x^{a-1}y^{m}\exp\left\{by - dx\left(e^{cy}-1\right)/y\right\}.$$

The normalizing factor is given by

$$K = \frac{d^{a}c^{a+m+1}}{\Gamma(a)J_m(b/c,a)}$$

where

$$\Gamma(a) = \int_0^\infty x^{a-1}e^{-x}dx,$$

$$J_m(\alpha,\beta) = \int_0^\infty x^{\alpha-1}\log^m(x)\left(\frac{\log(x)}{x-1}\right)^{\beta}dx.$$

We denote: $(X,Y) \sim G\text{-}M\text{-}G(a,b,c,d,m)$.

The normalizing constant K is obtain by computing

$$\begin{aligned}
&= \int_0^{+\infty}\int_{-\infty}^{+\infty} x^{a-1}y^m \exp\left\{by - dx\left(e^{cy}-1\right)/y\right\} dxdy.\\
&= \int_{-\infty}^{+\infty} y^m e^{by}\left(\int_0^{+\infty} x^{a-1}\exp\left\{-dx\left(e^{cy}-1\right)/y\right\}dx\right)dy.\\
&= \int_{-\infty}^{+\infty} y^m e^{by}\frac{\Gamma(a)}{[d\,(e^{cy}-1)/y]^a}dy.\\
&= \frac{\Gamma(a)}{d^a}\int_{-\infty}^{+\infty} y^m\left(\frac{e^{cy}-1}{y}\right)^a e^{by}dy.
\end{aligned}$$

By changing variable as $u = e^{cx}$ we get

$$K^{-1} = \frac{\Gamma(a)}{d^a c^{a+m+1}}\int_0^{\infty} u^{b/c-1}[\log(u)]^m\left[\frac{\log(u)}{u-1}\right]^a dx = \frac{\Gamma(a)J_m(b/c,a)}{d^a c^{a+m+1}}.$$

It can be shown that $J_m(\alpha,\beta)$ converges when $(1 \le \alpha \le \beta - 1)$ (see Appendix D) and note that $I(\alpha,\beta) = J_0(\alpha,\beta)$.

Marginal distributions and conditional distributions

The following theorem provides the marginal distribution of Y and the conditional marginal distribution of X.

Theorem 4.4.1. *Let $(X,Y) \sim G\text{-}M\text{-}G(a,b,c,d,m)$ with $a \ge 2; b,c,d > 0, m \in \mathbb{N}$ and such that $c \le b \le c(a-1)$. Then*

(i) The marginal distribution of Y is a m-G distribution with parameters (a,b,c,m)

$$f_Y(y) = \frac{c^{a+m+1}}{J_m(b/c,a)}y^m\left(\frac{y}{e^{cy}-1}\right)^a e^{by}, \quad y \in \mathbb{R}.$$

(ii) The conditional marginal distribution of X given $Y = y$ is a gamma distribution with parameters $(a, d\,(e^{cy}-1)/y)$.

Proof 10. –
Proof of (i): The marginal density of Y is obtained by integrating out x from the joint density of (X,Y)

$$\begin{aligned}
f_Y(y) &= \int_0^{+\infty} K\, x^{a-1}y^m\exp\left\{by - dx\left(e^{cy}-1\right)/y\right\}dx\\
&= K\, y^m e^{by}\int_0^{+\infty} x^{a-1}\exp\left\{-xd\left(e^{cy}-1\right)/y\right\}dx\\
&= K\, y^m e^{by}\frac{\Gamma(a)}{[d\,(cy-1)/y]^a}\\
&= \frac{c^{a+m+1}}{J_m(b/c,a)}y^m\left(\frac{y}{e^{cy}-1}\right)^a e^{by}.
\end{aligned}$$

Proof of (ii): The conditional density of X given Y is

$$\begin{aligned} f_{X|Y=y}(x) &= \frac{f_{X,Y}(x,y)}{f_Y(y)} \\ &= \frac{[d\,(e^{cy}-1)/y]^a}{\Gamma(a)} x^{a-1} \exp\left\{-xd\left(e^{cy}-1\right)/y\right\}. \end{aligned}$$

Thus $X \mid Y = y$ has a gamma distribution with parameters $(a, d\,(e^{cy}-1)/y))$.

□

Conditional expectation and conditional variance

The conditional expectation and conditional variance of X given $Y = y$ are easily obtained since that distribution belongs to gamma family of distributions.

$$\begin{aligned} E(X \mid Y = y) &= \frac{a}{d} \frac{y}{e^{cy}-1}, \\ Var(X \mid Y = y) &= \frac{a}{d^2} \left(\frac{y}{e^{cy}-1}\right)^2. \end{aligned}$$

Expectation, variance and covariance

The previous theorem allows us to compute the expectation and the variance of X and Y. We have

Theorem 4.4.2. *– Let $(X,Y) \sim$ G-M-G(a,b,c,d,m) with $a \geq 2; b,c,d > 0, m \in \mathbb{N}$ and $c \leq b \leq c(a-1)$. Then*

(i) The expectation of Y is

$$E(Y) = \frac{1}{c} \frac{J_{m+1}(b/c,a)}{J_m(b/c,a)}.$$

(ii) The standard deviation of Y is

$$SD(Y) = \frac{1}{c} \sqrt{\frac{J_m(b/c,a) J_{m+2}(b/c,a) - J_{m+1}^2(b/c,a)}{J_m^2(b/c,a)}}.$$

(iii) The expectation of X is

$$E(X) = \frac{a}{cd} \frac{J_m(b/c,a+1)}{J_m(b/c,a)}.$$

(iv) The standard deviation of X is

$$SD(X) = \frac{a}{cd} \sqrt{\frac{J_m(b/c,a) J_m(b/c,a+2) - J_m^2(b/c,a+1)}{J_m^2(b/c,a)}}.$$

(v) The covariance between X and Y is

$$Cov(X,Y) = \frac{a}{c^2 d}\frac{J_m(b/c,a)J_{m+1}(b/c,a+1) - J_m(b/c,a+1)J_{m+1}(b/c,a)}{J_m^2(b/c,a)}.$$

Proof 11. –
Proof of (i) and (ii): Since $Y \sim$ M-G(a,b,c,m) and the expectation and variances of M-G distribution are developed in the previous section, we have

$$E(Y) = \frac{1}{c}\frac{J_{m+1}(b/c,a)}{J_m(b/c,a)},$$
$$SD(Y) = \frac{1}{c}\sqrt{\frac{J_m(b/c,a)J_{m+2}(b/c,a)}{J_m^2(b/c,a)}}.$$

Proof of (iii) and (iv): The expectation of X can be computed as following

$$\begin{aligned}
E(X) &= \int_0^{+\infty}\int_{-\infty}^{+\infty} x f_{X,Y}(x,y)dxdy \\
&= \int_0^{+\infty}\int_{-\infty}^{+\infty} \frac{d^a c^{a+m+1}}{\Gamma(a)J_m(b/c,a)} x^a y^m \exp\left\{by - dx\left(e^{cy}-1\right)/y\right\}dxdy \\
&= \frac{d^a c^{a+m+1}}{\Gamma(a)J_m(b/c,a)}\int_{-\infty}^{+\infty} y^m e^{by}\left(\int_0^{+\infty} x^a \exp\left\{-dx\left(e^{cy}-1\right)/y\right\}dx\right)dy \\
&= \frac{d^a c^{a+m+1}}{\Gamma(a)J_m(b/c,a)}\int_{-\infty}^{+\infty} y^m e^{by}\frac{\Gamma(a+1)}{[d\,(e^{cy}-1)/y]^{a+1}}dy \\
&= \frac{a}{d}\frac{c^{a+m+1}}{J_m(b/c,a)}\int_{-\infty}^{+\infty} y^m\left(\frac{y}{e^{cy}-1}\right)^{a+1} e^{by}dy \\
&= \frac{a}{d}\frac{c^{a+m+1}}{J_m(b/c,a)}\frac{1}{c^{a+m+2}}\int_0^{\infty} u^{b/c-1}[\log(u)]^m\left(\frac{\log(u)}{u-1}\right)^{a+1}du \\
&= \frac{a}{cd}\frac{J_m(b/c,a+1)}{J_m(b/c,a)}.
\end{aligned}$$

The second moment of X can be obtained with the same reasoning

$$E(X^2) = \int_0^{+\infty}\int_{-\infty}^{+\infty} x^2 f_{X,Y}(x,y)dxdy = \frac{a^2}{(cd)^2}\frac{J_m(b/c,a+1)}{J_m(b/c,a)}.$$

The variance of X is then

$$\begin{aligned}
Var(X) &= E(X^2) - E^2(X) \\
&= \frac{a^2}{(cd)^2}\left(\frac{J_m(b/c,a)J_m(b/c,a+2) - J_m^2(b/c,a+1)}{J_m^2(b/c,a)}\right).
\end{aligned}$$

Thus we obtain the standard deviation of X as in (iv).

Proof of (v): The expectation of XY is

$$
\begin{aligned}
E(XY) &= \int_0^{+\infty}\int_{-\infty}^{+\infty} xy f_{X,Y}(x,y)dxdy \\
&= \int_0^{+\infty}\int_{-\infty}^{+\infty} \frac{d^a c^{a+m+1}}{\Gamma(a)J_m(b/c,a)} x^a y^{m+1}\exp\left\{by - dx\left(e^{cy}-1\right)/y\right\}dxdy \\
&= \frac{d^a c^{a+m+1}}{\Gamma(a)J_m(b/c,a)} \int_{-\infty}^{+\infty} y^{m+1}e^{by}\left(\int_0^{+\infty} x^a \exp\left\{-dx\left(e^{cy}-1\right)/y\right\}dx\right)dy \\
&= \frac{d^a c^{a+m+1}}{\Gamma(a)J_m(b/c,a)} \int_{-\infty}^{+\infty} y^{m+1}e^{by}\frac{\Gamma(a+1)}{[d\,(e^{cy}-1)/y]^{a+1}}dy \\
&= \frac{a}{d}\frac{c^{a+m+1}}{J_m(b/c,a)} \int_{-\infty}^{+\infty} y^{m+1}\left(\frac{y}{e^{cy}-1}\right)^{a+1} e^{by}dy \\
&= \frac{a}{d}\frac{c^{a+m+1}}{J_m(b/c,a)}\frac{1}{c^{a+m+3}} \int_0^{\infty} u^{b/c-1}[\log(u)]^{m+1}\left(\frac{\log(u)}{u-1}\right)^{a+1} du \\
&= \frac{a}{c^2 d}\frac{J_{m+1}(b/c,a+1)}{J_m(b/c,a)}.
\end{aligned}
$$

Hence the covariance between X and Y is

$$
\begin{aligned}
Cov(X,Y) &= E(XY) - E(X)E(Y) \\
&= \frac{a}{c^2 d}\frac{J_m(b/c,a)J_{m+1}(b/c,a+1) - J_m(b/c,a+1)J_{m+1}(b/c,a)}{J_m^2(b/c,a)}.
\end{aligned}
$$

□

Mode

Let's consider the G-M-G distribution with parameter $(a,b,c,d,0)$. If $a > 1$ the density has an unique mode (x_{mod}, y_{mod}) where

$$
x_{mod} = \frac{a-1}{d}\frac{y}{e^{cy}-1}
$$

and y_{mod} is the unique solution of the equation

$$
y - \frac{(a-1)(e^{cy}-1)}{((a-1)c-b)e^{cy}+b} = 0.
$$

This equation can be solved numerically by Newton-Raphson method as following

$$
\begin{aligned}
f(y) &= y - \frac{(a-1)(e^{cy}-1)}{((a-1)c-b)e^{cy}+b}, \\
f'(y) &= 1 - \left(\frac{(a-1)ce^{cy/2}}{((a-1)c-b)e^{cy}+b}\right)^2.
\end{aligned}
$$

The initial value for this iteration can be chosen as $y_0 = (2/c)\log(b/((a-1)c-b))$.

Graphical illustration

The figure 4.2 displays 3-D plots of density functions of G-M-G distribution with parameters. Wet set $m = 0$ and $a > 1$ so the densities have unique mode. Here we keep a, c, d being constant and vary b from c to $(a-1)c$.

4.4.3 Conjugate priors

We now consider the Bayesian inference for the G-M-G with conjugate priors. Let $\underline{t} = (t_1, \ldots, t_n)$ be a realization of the ELP with intensity $\lambda(t) = \alpha e^{\beta t}$, $\alpha > 0$, $\beta \in \mathbb{R}$ in a time window $[0, C]$. Denote $s_n = \sum_{i=1}^{n} t_i$. The following theorem shows that G-M-G distribution is a natural conjugate prior for the ELP.

Theorem 4.4.3. – *Let $\underline{t} = (t_1, \ldots, t_n)$ be the jump dates of a ELP with intensity $\alpha e^{\beta t}$ observed in a time window $[0, C]$. Denote $\tilde{\alpha}, \tilde{\beta}$ the Bayesian estimates of α, β respectively. Corresponding to each scenario of prior information we have associated conjugate prior as following:*

(i) – With two prior guesses, the G-M-G distribution with parameters $(a, b, \tau, 1, 0)$ is a conjugate prior for the ELP and the posterior distribution is a G-M-G distribution with parameters $(a+n, b+s_n, \tau, 2, 0)$.
The Bayesian estimators are:

$$\tilde{\beta} = \frac{1}{\tau} \frac{J_1((b+s_n)/\tau, a+n)}{I((b+s_n)/\tau, a+n)},$$
$$\tilde{\alpha} = \left(\frac{a+n}{\tau}\right)\left(\frac{I((b+s_n)/\tau, a+n+1)}{I((b+s_n)/\tau, a+n)}\right).$$

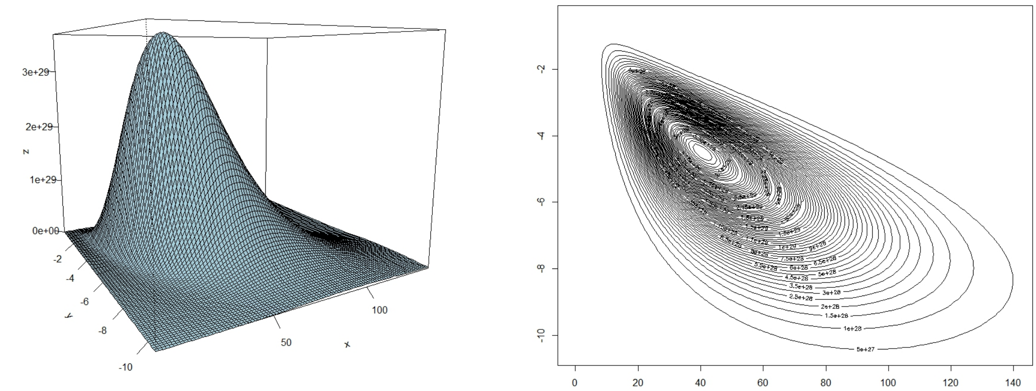
(ii) – With three prior guesses, the G-M-G distribution with parameters $(a, b, \tau, d, 0)$ is a conjugate prior for the ELP and the posterior distribution is a G-M-G distribution with parameters $(a+n, b+s_n, \tau, d+1, 0)$. The Bayesian estimators are:

$$\tilde{\beta} = \frac{1}{\tau} \frac{J_1((b+s_n)/\tau, a+n)}{I((b+s_n)/\tau, a+n)},$$
$$\tilde{\alpha} = \left(\frac{a+n}{\tau(d+1)}\right)\left(\frac{I((b+s_n)/\tau, a+n+1)}{I((b+s_n)/\tau, a+n)}\right).$$

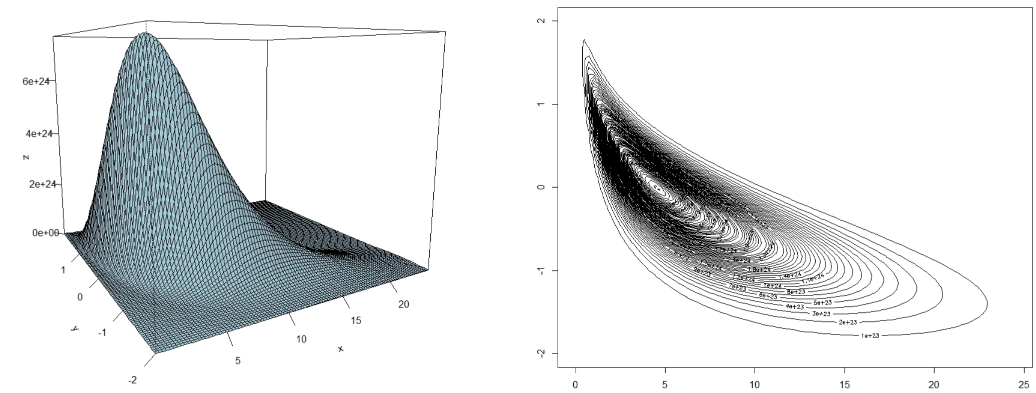
(iii) – With four prior guesses, the G-M-G distribution with parameters (a, b, τ, d, m) is a conjugate prior for the ELP and the posterior distribution is the G-M-G distribution with parameters $(a+n, b+s_n, \tau, d+1, m)$.
The Bayesian estimators are:

$$\tilde{\beta} = \frac{1}{\tau} \frac{J_{m+1}((b+s_n)/\tau, a+n)}{J_m((b+s_n)/\tau, a+n)},$$
$$\tilde{\alpha} = \left(\frac{a+n}{\tau(d+1)}\right)\left(\frac{J_m((b+s_n)/\tau, a+n+1)}{J_m((b+s_n)/\tau, a+n)}\right).$$

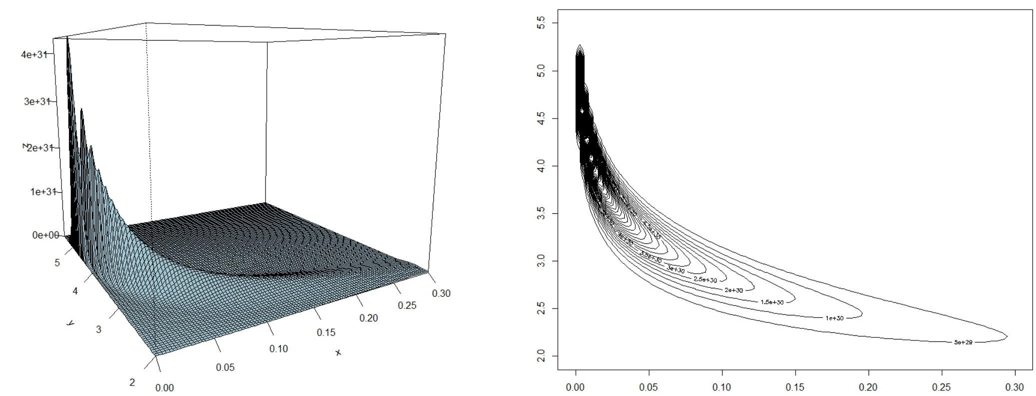
Figure 4.2: *P.d.f and contour plot of G-M-G distributions with different parameters*



Density plot and contour plot of G-M-G$(10,2,2,1,0)$



Density plot and contour plot of G-M-G$(10,9,2,1,0)$



Density plot and contour plot of G-M-G$(10,17,2,1,0)$

Proof 12. Recall that the probability density of an observation of the ELP is

$$f(\underline{t} \mid \alpha, \beta) = \alpha^n \exp\left\{\beta s_n - \alpha\left(e^{\beta\tau} - 1\right)/\beta\right\}$$

where $s_n = \sum_{i=1}^n t_i$, $\tau = t_n$ for event truncation and $\tau = T$ for time truncation.
Proof of (i): Let's take G-M-G$(a, b, \tau, 1, 0)$ as the joint prior distribution:

$$\pi(\alpha, \beta) \propto \alpha^{a-1} \exp\left\{b\beta - \alpha\left(e^{\beta\tau} - 1\right)/\beta\right\}.$$

Applying the Bayes' theorem, the posterior distribution is:

$$\pi(\alpha, \beta \mid \underline{t}) \propto \alpha^{a+n-1} \exp\left\{(b + s_n)\beta - 2\alpha\left(e^{\beta\tau} - 1\right)/\beta\right\}.$$

That is to say a G-M-G distribution with parameters $(a + n, b + s_n, \tau, 2, 0)$.
Proof of (ii): Now we take G-M-G$(a, b, \tau, d, 0)$ as the joint prior distribution:

$$\pi(\alpha, \beta) \propto \alpha^{a-1} \exp\left\{b\beta - d\alpha\left(e^{\beta\tau} - 1\right)/\beta\right\}.$$

Applying the Bayes' theorem, the posterior distribution is:

$$\pi(\alpha, \beta \mid \underline{t}) \propto \alpha^{a+n-1} \exp\left\{(b + s_n)\beta - (d+1)\alpha\left(e^{\beta\tau} - 1\right)/\beta\right\}.$$

That is a G-M-G distribution with parameters $(a + n, b + s_n, \tau, d + 1, 0)$.
Proof of (iii): Finally, we take G-M-G(a, b, τ, d, m) as the joint prior distribution:

$$\pi(\alpha, \beta) \propto \alpha^{a-1} \beta^m \exp\left\{b\beta - d\alpha\left(e^{\beta\tau} - 1\right)/\beta\right\}.$$

Applying the Bayes' theorem, the posterior distribution is:

$$\pi(\alpha, \beta \mid \underline{t}) \propto \alpha^{a+n-1} \beta^m \exp\left\{(b + s_n)\beta - (d+1)\alpha\left(e^{\beta\tau} - 1\right)/\beta\right\}.$$

That is a G-M-G distribution with parameters $(a + n, b + s_n, \tau, d + 1, m)$. Assuming a quadratic loss, the Bayes estimators are the expectation of the posterior distributions. Since the expectations of a G-M-G distribution has given in the theorem 4.4.2, we can easily obtain the Bayesian estimates of α, β as in $(i), (ii), (iii)$.

□

4.4.4 Prior elicitation

We suggest some strategies to elicitate the prior parameters according to provided prior guesses. Each scenario can have different prior elicitation strategies. Since G-M-G distributions do not closed-form expressions for their components, it requires trials and errors procedure to obtain values for prior parameters.

Scenario 1

In this scenario, we are at disposal of two prior guesses. In the first case, practitioners have a guess on the value of β and a guess on the confidence of the first guess (strong confidence, moderate confidence or weak confidence). In the other case, practitioners are provided a guess on the value of β and a guess on the value of α. Let's consider H-B$(a, b, \tau, 1, 0)$ as conjugate prior for the PLP. The values for a, b need to be provided.

Elicitation strategy 1:

Suppose that the practitioner has a guess $g_{\beta,1}$ at the value of β and a guess $g_{\beta,2}$ at the standard deviation associated with $g_{\beta,1}$. The value for a and b are obtained by solving the system of two equations:

$$\begin{cases} \dfrac{1}{\tau}\dfrac{J_1(b/\tau,a)}{J_0(b/\tau,a)} = g_{\beta,1}, \\ \dfrac{1}{\tau}\sqrt{\dfrac{J_0(b/\tau,a)J_2(b/\tau,a) - J_1^2(b/\tau,a)}{J_0^2(b/\tau,a)}} = g_{\beta,2}. \end{cases}$$

One can start a value for a at $a = n$ for example. With the chosen value for a, the value for b need trials and errors procedure such that:

$$\begin{cases} \dfrac{J_1(b/\tau,a)}{J_0(b/\tau,a)} = \tau g_{\beta,1}, \\ \dfrac{J_0(b/\tau,a)J_2(b/\tau,a) - J_1^2(b/\tau,a)}{J_0^2(b/\tau,a)} = (\tau g_{\beta,2})^2. \end{cases}$$

Elicitation strategy 2:

Suppose that the practitioner has a guess g_β at the value of β and a guess g_α at the value of α. The value for a, k can be obtained by solving the system of two equations:

$$\begin{cases} \dfrac{1}{\tau}\dfrac{J_1(b/\tau,a)}{J_0(b/\tau,a)} = g_\beta, \\ \dfrac{a}{\tau}\dfrac{I(b/\tau,a+1)}{I(b/\tau,a)} = g_\alpha. \end{cases}$$

One can start a value for a at $a = n$ for example. With the chosen value for a, the value for b need trials and errors procedure such that:

$$\begin{cases} \dfrac{J_1(b/\tau,a)}{J_0(b/\tau,a)} = y g_\beta, \\ \dfrac{I(b/\tau,a+1)}{I(b/\tau,a)} = \tau g_\alpha / a. \end{cases}$$

Scenario 2

In this scenario, we are at disposal of three prior guesses: a guess $g_{\beta,1}$ at the value of β, a guess $g_{\beta,2}$ at the standard deviation associated with $g_{\beta,1}$, and a guess g_α at the value of α. We now employ G-M-G$(a,b,\tau,d,0)$ as conjugate prior for the ELP.

The values for a,b are obtained as in scenario 1, strategy 1. Since the prior expectation of α is

$$E^\pi(\alpha) = \frac{a}{\tau g_\alpha}\frac{I(b/\tau,a+1)}{I(b/\tau,a)}.$$

The value for d is then obtained easily as

$$d = \frac{a}{\tau g_\alpha}\frac{I(b/\tau,a+1)}{I(b/\tau,a)}.$$

Scenario 3

When we have four prior guesses, we need four free prior parameters to be elicitated. Suppose that we have two guesses $g_{\beta,1},g_{\alpha,1}$ at the values of β,α, two guesses $g_{\beta,2},g_{\alpha,2}$ at the standard deviation associated with $g_{\beta,1},g_{\alpha,1}$. A conjugate prior of G-M-G distribution with five parameters (a,b,τ,d,m) would be suitable for the ELP. Four prior parameters satisfy the following system of equations:

$$\begin{cases} \dfrac{1}{\tau}\dfrac{J_{m+1}(b/\tau,a)}{J_m(b/\tau,a)} = g_{\beta,1}, \\ \dfrac{1}{\tau}\sqrt{\dfrac{J_m(b/\tau,a)J_{m+2}(b/\tau,a)-J^2_{m+1}(b/\tau,a)}{J^2_m(b/\tau,a)}} = g_{\beta,2}, \\ \dfrac{a}{\tau d}\dfrac{J_m(b/\tau,a+1)}{J_m(b/\tau,a)} = g_{\alpha,1}, \\ \dfrac{a}{\tau d}\sqrt{\dfrac{J_m(b/\tau,a)J_m(b/\tau,a+2)-J^2_m(b/\tau,a+1)}{J^2_m(b/\tau,a)}} = g_{\alpha,2}. \end{cases}$$

Those equations can not be solved explicitly but require trials and errors. One can start at $m=0, a=n$ then vary b to fulfill the first two equations. Finally, the value for d is obtained from the third equation.

Scenario 4

In this scenario, we are at disposal of two prior guesses: a guess g_β at the value of β and a guess g_α at the value of α. We want to employ a prior G-M-G distribution with parameters $(a,b,\tau,d,0)$ as conjugate prior for the ELP. The parameter c is already fixed equaling to τ, that is t_n or C corresponding to event truncation or time truncation scheme of the observation. We suggest a elicitation strategy integrating prior guesses to elicitate three hyperparameters a, b, d.

Recall that

$$\hat{\alpha} = \frac{n\hat{\beta}}{e^{\hat{\beta}} - 1}$$

The conditional expectation of α given $\beta = \hat{\beta}$ is

$$E^{\pi}(\alpha \mid \beta = \hat{\beta}) = \frac{a}{d}\frac{\hat{\beta}}{e^{\tau\hat{\beta}} - 1} = \frac{a}{dn}\hat{\alpha}$$

One can interpret $a/(dn)$ is a correction factor between maximum likelihood estimate of α and its prior conditional expectation. By setting $a/(dn) = 1$ we obtain $d = a/n$. Now, taking the guesses of α and β to the following equations:

$$\frac{n}{\tau}\frac{I(b/\tau, a+1)}{I(b/\tau, a)} = g_{\alpha},$$
$$\frac{1}{\tau}\frac{J_1(b/\tau, a)}{I(b/\tau, a)} = g_{\beta}.$$

The values of a, b can be deduced from those above equations by numerical approximation. Start from $a = n$ then varies b and calculate integrals $I(b/\tau, a+1)$, and $I(b/\tau, a)$ such that $I(b/\tau, a+1)/I(b/\tau, a) = \tau g_{\alpha}/n$ then calculate $J_1(b/\tau, a)$ and check if $J_1(b/\tau, a)/I(b/\tau, a) = \tau g_{\beta}$.

4.4.5 Application

Simulated data

A simulation study is conducted to compare Bayesian estimation with G-M-G prior and maximum likelihood estimation. The ELP(α, β) with input values (1.0, 0.01) is simulated in three scenarios of sample size: small sample size ($n = 50$), medium sample size ($n = 100$), and big sample size ($n = 1000$). Three scenarios of prior guesses for β are considered: **1.** under-guess $g_{\beta,1} = 0.005$, **2.** precise-guess $g_{\beta,1} = 0.009$; and **3.** over-guess $g_{\beta,1} = 0.015$. For each given prior guess $g_{\beta,1}$, computations are carried out using three incertitude values of variability $g_{\beta,2}$ according to the scheme: $g_{\beta,2} = \rho g_{\beta,1}$, where $\rho = 0.3, 0.6, 0.9$ are the coefficient of variation.

The mean Bayesian estimates and maximum likelihood estimates are shown in table 4.1. We use the elicitation strategy 1 of scenario 4 for choosing the hyper-parameters of the conjugate prior. It requires trials and errors method to obtain parameters of G-M-G distribution. With large sample size, the Bayesian estimates with conjugate prior G-M-G distribution (BaE) are very similar to the maximum likelihood estimates (MLE). Both of the BaEs and the MLEs are close to the input values of the two parameters whatever the prior guess for β. With medium sample size, the BaEs are more accurate than the MLEs in most of the case. For instance, a under-guess associated

Table 4.1: *Mean of the Bayesian estimates with strategy 1 for simulated data from ELP(α,β) with input values (1.0, 0.01)*

Sample-size	Prior guess		Bayes estimates	
n	$g_{\beta,1}$	$g_{\beta,2}$	$\tilde{\beta}$	$\tilde{\alpha}$
50	0.005	0.0015	0.0253 (0.006607)	0.5210 (0.057246)
		0.0030	0.0753 (0.043955)	0.8134 (0.037773)
		0.0045	0.0748 (0.092842)	0.8327 (0.032956)
	0.009	0.0027	0.0141 (0.001514)	0.4351 (0.074324)
		0.0054	0.0553 (0.034864)	0.7252 (0.022375)
		0.0081	0.0475 (0.032764)	0.9226 (0.032945)
	0.015	0.0045	0.0426 (0.014202)	0.3521 (0.045248)
		0.0090	0.0332 (0.042648)	0.87523 (0.052759)
		0.0135	0.0573 (0.028964)	0.63572 (0.025943)
	MLE		0.0269 (0.002689)	0.8177 (0.053247)
100	0.005	0.0015	0.0125 (0.002676)	0.8122 (0.027245)
		0.0030	0.0235 (0.004394)	0.7651 (0.033736)
		0.0045	0.0785 (0.092884)	0.8327 (0.029567)
	0.009	0.0027	0.0412 (0.015142)	0.6355 (0.074732)
		0.0054	0.0535 (0.034856)	0.7265 (0.025237)
		0.0081	0.0752 (0.032786)	0.9252 (0.032894)
	0.015	0.0045	0.0402 (0.004203)	0.5213 (0.045824)
		0.0090	0.0302 (0.004264)	0.8527 (0.052752)
		0.0135	0.0536 (0.002896)	0.8325 (0.059482)
	MLE		0.0119 (0.000415)	0.8164 (0.033695)
1000	0.005	0.0015	0.0102 (0.006603)	0.9205 (0.005724)
		0.0030	0.0107 (0.004395)	0.9136 (0.003777)
		0.0045	0.0108 (0.002845)	0.9373 (0.003295)
	0.009	0.0027	0.0104 (0.001515)	0.9325 (0.000743)
		0.0054	0.0103 (0.003486)	0.9825 (0.000223)
		0.0081	0.0107 (0.003276)	0.9226 (0.000329)
	0.015	0.0045	0.0106 (0.001204)	0.9518 (0.0004524)
		0.0090	0.0132 (0.002683)	0.9752 (0.0005275)
		0.0135	0.0107 (0.000896)	0.9857 (0.000295)
	MLE		0.01009 (0.000635)	0.9995 (0.000005)

Table 4.2: *Test-fix-test for failure times of a repairable system within* $C = 400$ *hours*

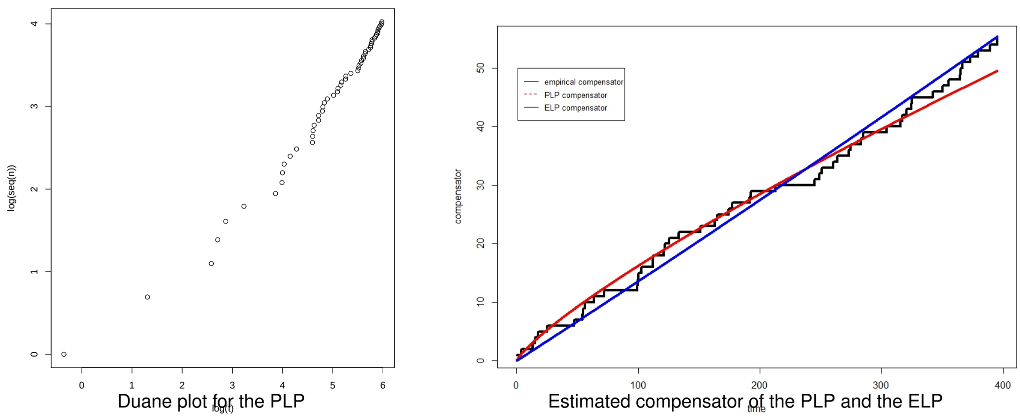
0.7	3.7	13.2	15	17.6	25.3	47.5	54	54.5	56.4
63.6	72.2	99.2	99.6	100.3	102.5	112	112.2	120.9	121.9
125.5	133.4	151	163	164.7	174.5	177.4	191.6	192.7	213
244.8	249	250.8	260.1	263.5	273.1	274.7	282.8	285	304
315.4	317.1	320.6	324.5	324.9	342	350.2	355.2	364.6	364.9
366.3	373	379.4	389	394.9	395.2				

with moderate confidence or weak confidence on the value of β give the BaEs better than the MLEs. With small sample size, the BaEs outperform the MLEs in some cases. With a precise-guess for β, the BaEs are always more accurate than the MLEs.

Real data

We use the data that are introduced by Crow ([15]). The system is tested in a fix time-window $C = 400$ hours with the 56 failure times given in the table 4.2. The first failure was recorded at 0.7 hours into the test and the last failure occurred at 395.2 hours into the test. If we fit the data with the exponential-law process, the MLE for α is $\hat{\alpha} = 0.1347$ and the MLE for β is $\hat{\beta} = 0.0002$. If we fit the data with the power-law process, the MLE for μ is $\hat{\mu} = 0.2397$ and the MLE for β is $\hat{\beta} = 0.9103$. The AIC score of the ELP is 169.9364 and the AIC score of the PLP is 169.373 so both models fit well with the data. The figure 4.3 illustrates the residual test of the two models.

Figure 4.3: *Residual test for the PLP and the ELP with data in the table 4.2*



Duane plot for the PLP

Estimated compensator of the PLP and the ELP

4.5 Concluding Remarks

In this chapter, we have studied the exponential-law process (ELP) which can be considered as an alternative to the power-law process (PLP). However, beside the tractable likelihood, this model has no convenient graphical test and no closed forms for the maximum likelihood estimators (MLE). The maximum likelihood estimation procedure requires numerical approximation and we have employed a Newton-Raphson method. As for the PLP, we have constructed a natural conjugate prior that we have called the *G-m-G distribution*. This bivariate distribution has similar shape as the H-B distribution but there is no closed forms for its marginal distribution. Therefore, Bayesian analysis for the ELP requires trial and error procedure for prior elicitation and numerical integration for posterior inference. A simulation study has been deducted to compare maximum likelihood estimates and Bayesian estimates.

Despite the fact that we still need some numerical integrals and some trials and errors procedures, the G-m-G distribution as conjugate prior for the ELP lessens the calculation of Bayesian estimates. Simulation studies support the well-known result that Bayesian estimates are better than maximum likelihood estimates when the sample size is small. More investigations concerning the properties of this distribution need to be carried out. In particular a better understanding of its properties will be helpful to elicit prior parameters.

Chapter 5

Self-Exciting Point Processes

In this chapter we study self-exciting point process and we introduce the power-law covariate self-exciting point process. Inference for this latter is driven and we compare maximum likelihood and Bayes approaches.

5.1 Introduction

In the previous chapter, we deal with point processes where the intensity was only depending on time and has a rather simple and flexible expression. We are now going to consider the case where the intensity is not only a function of time but also the point process itself. A process with such intensity is called a *self-exciting point process* (SEPP).

We will use the notation suggested by Snyder and Miller [56],

$$\lambda^*(t) = \lambda(t, N(t); \underline{w}_{N(t)})$$

for the intensity where $\underline{w}_{N(t)}$ is the set of occurrences times $w_1 < \cdots < w_{N(t)}$ in the time window $[0, C]$.

A basic example of such intensity is studied by Mino [39] who considers the following intensity expression:

$$\lambda^*(t) = \mu\big(1 + \alpha e^{-\beta(t - w_{N(t)})}\big),$$

where $\mu > 0$, $\alpha \geqslant -1$ and $\beta > 0$.

When $\alpha = 0$, $\lambda^*(t) = \mu$, the process is a homogeneous Poisson process.

When $\alpha > 0$, $\lambda^*(t)$ decreases after a jump until the next jump where it returns to the value $\mu(1+\alpha)$; the process is said to be **excited**.

When $-1 \leqslant \alpha < 0$, $\lambda^*(t)$ increases just after a jump to raise a value less than μ before returning to the value $\mu(1+\alpha)$; the process is said to be **inhibited**.

Excitation or inhibition can be viewed as events trigger by the occurrence of jump.

The figures 5.1 and 5.2 displayed a representation of the intensity in the cases of excited process and inhibited process.

Figure 5.1: *Intensity of a Mino process (excited):* $\mu = 200$, $\alpha = 0.5$, $\beta = 250$ *et* $T = 0.1$ *ms*

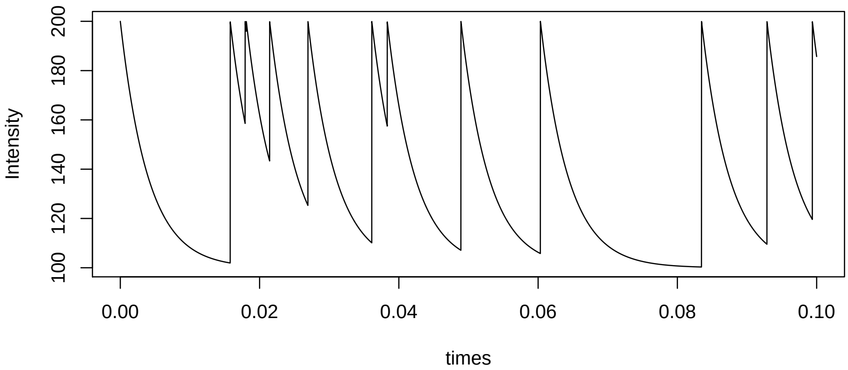
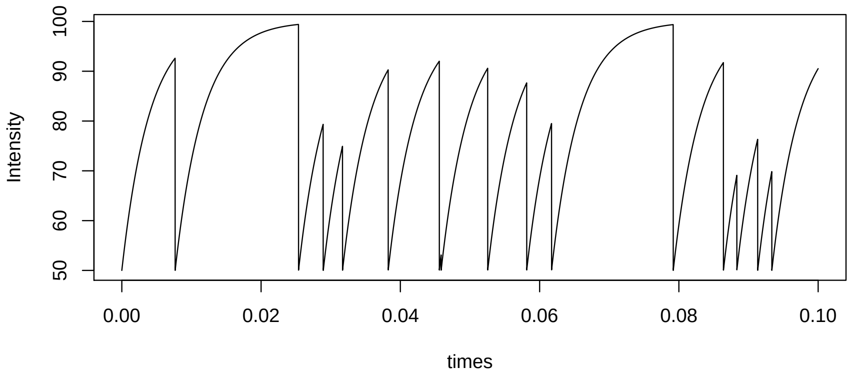


Figure 5.2: *Intensity of a Mino process (inhibited) :* $\mu = 100$, $\alpha = -0.5$, $\beta = 250$ *et* $T = 0.1$ *ms*



Gouno and Rabih (2015) [22] investigate the Mino process in more details and shows that it can be interpreted as a renewal process.

A very classical and well-known example of SEPP is the Hawkes process introduce by Hawkes in 1971. Hawkes [24] considers a process with intensity:

$$\lambda^*(t) = \mu + \int_0^t g(t-s)dN(s),$$

where $\underline{t}_{N(t)} = (t_1, \ldots, t_{N(t)})$ and g is a positive function such that

$$m = \int_0^{+\infty} g(u)du < +\infty.$$

From this expression, on can see that a SEPP has a HPP component with intensity μ that generate *main jumps* also called *background events* and another component depending on the function g which is called the *response function*. This last component corresponds to triggered events.

Hawkes and Oakes (1974) shows that all stationary self-exciting point processes can be represented as a Poisson cluster process which is an age-dependent immigration-birth process. If $m < 1$, the mean cluster size is $c = 1/(1-m)$ and the rate of the process is $\mu/(1-m)$.

The behaviour of the SEPP will strongly depend on the nature of g.

Many models for g can be found in the literature.

Hawkes suggests the case of exponential decay:

$$g(t) = \sum_{j=1}^{k} \alpha_j e^{-\beta_j t}, \ t \in \mathbb{R}^+ \text{ avec } \sum_{j=1}^{k} \alpha_j/\beta_j < 1.$$

The case $k = 1$ leads to the formula $g(t) = \alpha e^{-\beta t}$ with $\alpha < \beta$ which is often used in sismology and named *Lomnitz formula*. In this case, $m = \alpha/\beta$ and is called the *branching ratio*.

In this case, the expression of the intensity is:

$$\lambda^*(t) = \mu + \alpha \int_{-\infty}^{t} e^{-\beta(t-s)} dN(s) = \mu + \alpha \sum_{t_i < t} e^{-\beta(t-t_i)}.$$

The figures 5.1 and 5.1 represent Hawkes process with exponential decay for different parameters values. For the first representation the size of the cluster is 5 and it is 2 for the second representation.

Figure 5.3: *Representation of the intensity for $\mu = 0.5$, $\alpha = 4$ and $\beta = 5$.*

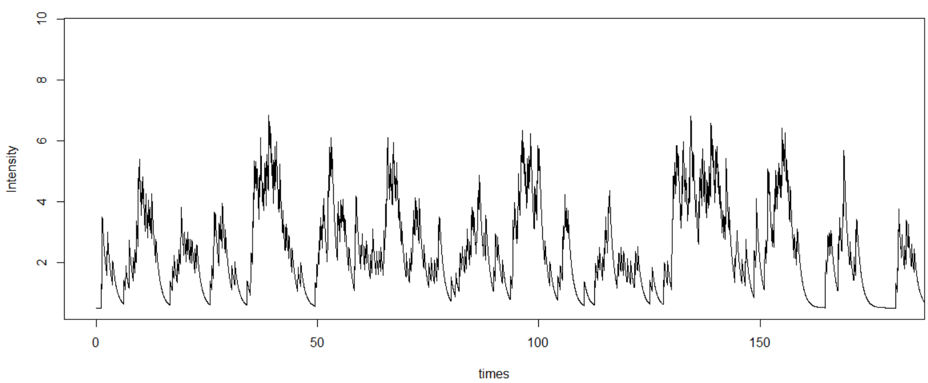
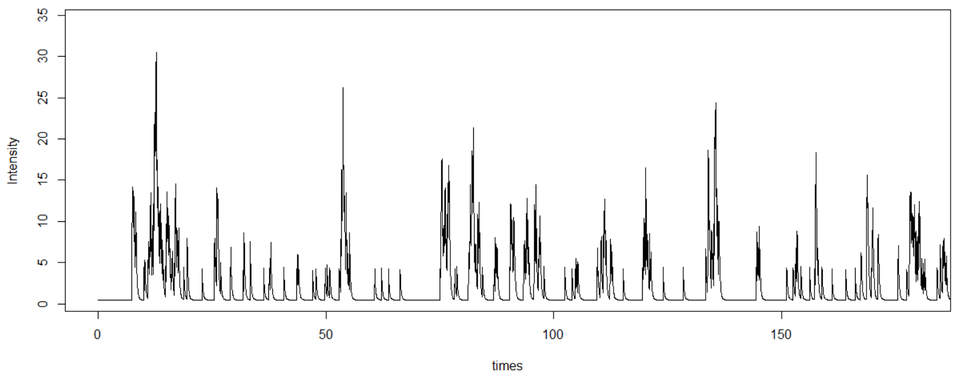


Figure 5.4: *Representation of the intensity for* $\mu = 0.8$, $\alpha = 0.5$ *and* $\beta = 1$.



Hawkes & Oakes [25] prove the asymptotic normality of the counting process under certain conditions:

Theorem 5.1.1. *If* $\int_0^{+\infty} ug(u)du < \infty$, *then*

$$\frac{N[0,t) - \mu t/(1-m)}{\sqrt{\mu t/(1-m)^3}} \sim \mathcal{N}(0,1) \text{ as } t \to \infty.$$

Hawkes models are commonly used in seismology, where they are sometimes called *epidemic-type aftershock sequence* (ETAS) models. An earthquake is described by *main shocks* which are follow by smaller earthquakes named *aftershocks*. Those aftershocks can be considered as response to main shocks.

Another form of function g that is used in modeling earthquake aftershocks relies the Omori's law introduced by Utsu [60].

$$g(t) = \frac{K}{(t+c)^p},$$

where the parameter K depends on the lower bound of the magnitude of aftershocks, the parameter c is in days and the paramter p has no dimension. This model corresponds to a power-law decay in the clustering behavior over time.

Ogata and Akaike [44] suggest the Laguerre type polynomial response function:

$$g(t) = \sum_{k=0}^{m} a_k t^k e^{-bt}.$$

In finance, Lorenzen [35] considers the expression of a Weibull p.d.f. as a response function.

In many situations, it can be interesting to consider that the evolution of the process depends on some covariates. For example, the occurrence of aftershocks may depend of the magnitude of the main shock and magnitude of the successive aftershocks can have an influence on each other. For thunderstorm, we adopt this kind of reasoning supposing that the charge associated to a thunderstorm impact will have an effect on the occurrence time (and charge) of the next one.

Covariate intensity models are devoted to reflect this phenomena. The idea is to consider response functions which depend on covariates. In the sequel a self-exciting point process with an intensity depending on covariates will be called a *covariate self-exciting point process* (C-SEPP).

In the field of sismology, Ogata [43] propose a model with five parameters $(\mu, \alpha, \beta, c, p)$ for modeling earthquakes data:

$$\lambda^*(t) = \mu + \sum_{t_i<t} e^{-\beta(m_i - M)} \frac{\alpha}{(t - t_i + c)^p}.$$

where m_i denotes the magnitude associated to the i^{th} event and M denotes the minimum magnitude.

Peruggia and Santner [48] propose a model with four parameters (μ, α, β, K)

$$\lambda^*(t) = \mu + \sum_{t_i<t} e^{-\beta(m_i - M)} K e^{-\alpha(t-t_i)}.$$

One can suggest the following general definition for a covariate self-exciting point process:

$$\lambda^*(t) = \mu + \sum_{t_i<t} \psi(z_i)\, g(t - t_i).$$

where ψ is called the *covariate function*.

This covariate function is a function that describes the effect of a jump on the environment; the environment being characterized by a random variable z.

For models (5.1) and (5.1), z is m, the magnitude of the earthquake and the covariate function is of the form: $\psi(z) = e^{-\beta(z-M)}$.

Our suggestion for thunderstorms is to consider the form of a power-law model for the covariate function defined as:

$$\psi(z_i) = \begin{cases} \left(\dfrac{z_i}{z_0}\right)^\eta & \text{if } z_i > z_0, \\ 1 & \text{if } z_i \leqslant z_0. \end{cases}$$

where z_i denote the amplitude in volt of the lightning strike occurring at time t_i and z_0 is a fixed threshold. All lightning strike that has an amplitude charge lower than the fixed constant z_0, will have no effect on the intensity. The full expression of the intensity is then:

$$\lambda^*(t) = \mu + \alpha \sum_{t_i<t} \left(\frac{z_i}{z_0}\right)^\eta e^{-\beta(t-t_i)}$$

where the response function has an exponential-law form.

A SEPP with such intensity will be called a *power-law covariate self-exciting point process* PLC-SEPP.

The idea behind using this model is that lightning strikes of a thunderstorm trigger the atmosphere to conditions for more impacts nearby in time and space. If z_i is greater than z_0, the impact excites the process. The bigger the amplitude is, the greater influence it makes to the atmosphere. If z_i is smaller than z_0 then the impact has no effect on the process.

5.2 Maximum Likelihood

The general expression of log-likelihood for a stochastic process observed in a time window [0, C] (see [16], p. 23) is given by:

$$\log L(\theta) = \int_0^C \log \lambda^*(t;\theta)dt - \int_0^C \log \lambda^*(t,\theta)dN(t) \tag{5.1}$$

We suggest in appendix E a scheme of construction.

Applying (5.1), the log-likelihood function considering an ordered sequence of dates of jumps $t_1, t_2, \dots t_n$ from a Hawkes process with an intensity defined as $\lambda^*(t) = \mu + \int_0^t g(t-s)dN(s)$ is given by

$$-\Lambda(C) + \sum_{t_i<C} \log\Big(\mu + \sum_{t_j<t_i} g(t_i - t_j)\Big),$$

When the response function is $g(t) = \alpha e^{-\beta t}$ (exponential decay), we obtain the following expression for the log-likelihood function:

$$\begin{aligned}\log L(\mu,\alpha,\beta) &= -\mu C - \frac{\alpha}{\beta}\sum_{j=1}^{n-1}\Big[e^{-\beta(C-t_j)} - 1\Big] + \sum_{i=1}^{n}\log\Big(\mu + \sum_{j=1}^{i-1}\alpha e^{-\beta(t_i-t_j)}\Big)\\ &+ \log\Big(\mu + \sum_{j=1}^{n}\alpha e^{-\beta(C-t_j)}\Big)\end{aligned}$$

Details of the computation are displayed in appendix F.

When the time window of the observation is undefined, we set $C = t_n$, and the log-likelihood is:

$$\log L(\mu,\alpha,\beta) = -\mu t_n - \frac{\alpha}{\beta}\sum_{j=1}^{n-1}\Big[e^{-\beta(t_n-t_j)} - 1\Big] + \sum_{i=1}^{n}\log\Big(\mu + \sum_{j=1}^{i}\alpha e^{-\beta(t_i-t_j)}\Big)$$

This last result is given by Ozaki [46] who obtains maximum likelihood estimates for μ, α and β using a Newton-Raphson method.

In the following two sections, we consider parameters estimation issues for the classical Hawkes process and for the power-law covariate self-exciting point process. We introduce what we called a *reduced maximum likelihood procedure* (RMLE).

5.2.1 Reduced maximum likelihood procedure for the Hawkes process

Let us denote:

$$S_n(\beta) = \frac{1}{\beta}\sum_{i=1}^{n}\left[1 - e^{-\beta(t_n - t_i)}\right]$$

and

$$A_k(\beta) = \sum_{i=1}^{k-1} e^{-\beta(t_k - t_i)} \text{ for} k > 1,\ A_1 = 0.$$

The expression of the log-likelihood is then:

$$\log L(\mu,\alpha,\beta) = -\mu\tau - \alpha S_n(\beta) + \sum_{i=1}^{n}\log\left(\mu + \alpha A_i(\beta)\right).$$

Let us recall the expression of the compensator of the Hawkes process, we have:

$$\Lambda^*(\tau) = \int_0^{t_n} \lambda^*(u)du = \mu\tau + \alpha S_n(\beta).$$

This quantity is the expected number of events in the interval $[0,\tau]$; it can be approximated by n. Therefore we can consider the problem: maximizing the likelihood (5.2.1) subject to the constraint : $\mu\tau + \alpha S_n(\beta) = n$ and maximizing (5.2.1) becomes equivalent to maximizing

$$-n + \sum_{i=1}^{n}\log\left(\frac{1}{\tau}\left[n - \alpha S_n(\beta)\right] + \alpha A_i(\beta)\right).$$

Let us denote $\overline{\mu} = n/\tau$ and $\overline{S}_n = S_n(\beta)/\tau$.

The log-likelihood is now:

$$L^*(\alpha,\beta) = -n + \sum_{i=1}^{n}\log\left[\overline{\mu} + \alpha(A_i - \overline{S}_n)\right].$$

The maximization of the log-likelihood of the HaP has been reduced by one dimension; it does not depend on μ. We have now a bivariate function that can be investigated graphically through its 3-dimension surface and its contour representation. Then the existence and unicity of a unique maximum can be easily checked and we can pick starting points up for the Newton -Raphson algorithm. The Newton algorithm is described in appendix H.

Numerical results

We test our reduced maximum likelihood procedure with simulated data. We generated 500 jumps of a Hawkes process with two sets of input parameters values. The first one is $\mu = 0.5$, $\alpha = 4$ and $\beta = 5$. The second one is $\mu = 0.5$, $\alpha = 0.8$ and $\beta = 1$. They have the same baseline intensity and the same branching ratio ($\alpha/\beta = 0.8$) but a different decay rate.

The figures (5.5) and (5.6) display a representation of the cumulative intensity (empirical compensator) with the two sets of parameters above.

One can see that in our examples, each objective function has only one maximum. The maximum likelihood estimates are shown in table 5.1. The objective functions for the RMLE procedure are shown in figure (5.7).

We consider now the problem of estimating the parameters of a PLC-SEPP with the ML procedure.

Figure 5.5: *Empirical compensator for the Hawkes process with parameters* $\mu = 0.5$, $\alpha = 4$, $\beta = 5$

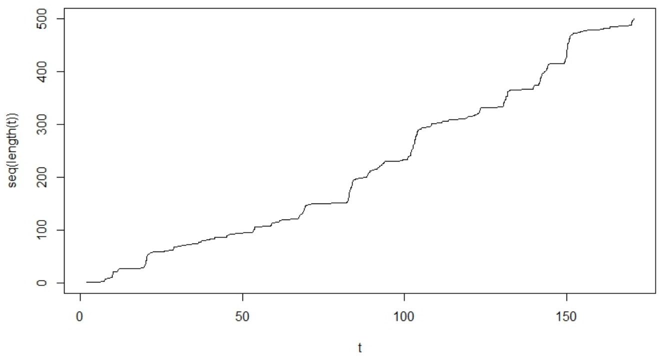


Figure 5.6: *Empirical compensator for the Hawkes process with parameters* $\mu = 0.5$, $\alpha = 0.8$, $\beta = 1$

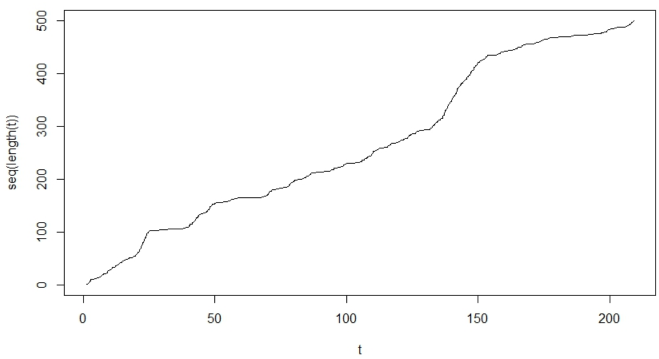
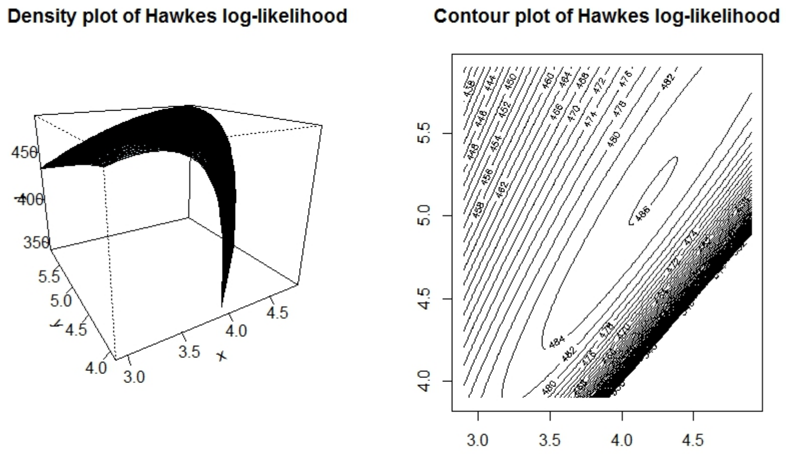
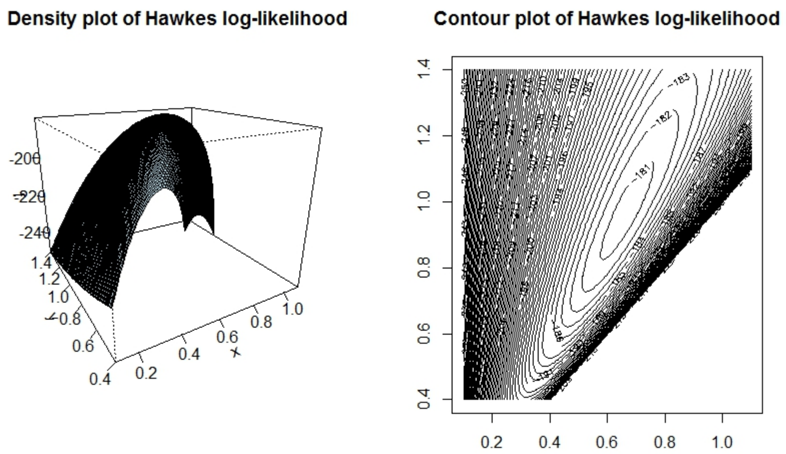


Figure 5.7: *Objective functions for the RMLE procedure*



HaP(0.5, 4.0, 5.0)



HaP(0.5, 0.8, 1.0)

Table 5.1: Mean of maximum likelihood estimates for the parameters of the HaP

Parameter	μ	α	β
Input	0.5	4.0	5.0
MLE	0.657	4.216	5.147
mse	0.0953	0.2357	0.1542
Input	0.5	0.8	1.0
MLE	0.498	0.793	0.963
mse	0.0854	0.3526	0.2154

5.2.2 Inference for the PLC-SEPP

Let us now consider an intensity of the form (5.1). Assume that z is a r.v. following a distribution f_z. z and the date of jump are independent r.v.'s. Since z and the dates are independent r.v's, the likelihood is the following product ([48]):

$$\prod_{i=1}^{n} f_z(z_i) \times \exp\{-\Lambda^*(\tau)\} \prod_{i=1}^{n} \lambda^*(t_i)$$

where

$$\Lambda^*(\tau) = \mu\tau + \sum_{i=1}^{n} \psi(z_i) \left[1 - e^{-\beta(\tau - t_i)}\right].$$

Remark that the last term of the product can be expressed as:

$$exp\left(-\Lambda^*(\tau) + \sum_{t_i<t} log\Big(\mu + \sum_{t_i<t} \psi(z_i) g(t-t_i)\Big)\right).$$

Let us consider an example. Let $\{(t_1, z_1), \ldots, (t_n, z_n)\}$ be a sample of observations where (t_i, z_i) are respectively the date and the value of the covariate associated with the i^{th} jump of a covariate self-exciting point process with intensity:

$$\lambda^*(t) = \mu + \sum_{t_i<t} \left(\frac{z_i}{z_0}\right)^{\eta} \alpha e^{-\beta(t-t_i)}.$$

The compensator is:

$$\Lambda^*(\tau) = \mu\tau + \alpha \sum_{i=1}^{n} \left(\frac{z_i}{z_0}\right)^{\eta} \frac{1}{\beta} \left[1 - e^{-\beta(\tau - t_i)}\right].$$

Let us assume that z follows a log-normal distribution with parameters (ω, σ^2). Therefore $\mu, \alpha, \beta, \eta, \omega, \sigma^2$ are six parameters to be estimated.

The likelihood is split in two parts

$$L(\mu, \alpha, \beta, \eta, \omega, \sigma) = L_1(\mu, \alpha, \beta, \eta) \times L_2(\omega, \sigma)$$

where

$$L_2(\omega,\sigma) = \prod_{i=1}^{n} f_z(z_i) = (2\pi)^{-n/2}\left(\prod_{i=1}^{n} z_i\right)^{-1}\sigma^{-n}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(z_i-\omega)^2\right\}$$

and

$$L_1(\mu,\alpha,\beta,\eta) = \left\{\prod_{i=1}^{n}\left[\mu+\alpha A(i)\right]\right\}\exp\left\{-\mu\tau-\frac{\alpha}{\beta}\sum_{i=1}^{n}\left(\frac{z_i}{z_0}\right)^{\eta}\left[1-e^{-\beta(\tau-t_i)}\right]\right\}$$

where $A(1)=0$ and for $k\geqslant 2$

$$A(k) = \sum_{i=1}^{k-1}\left(\frac{z_i}{z_0}\right)^{\eta} e^{-\beta(t_k-t_i)}.$$

The maximum likelihood estimators for parameters ω,σ of the log-normal distribution can be easily obtained:

$$\hat{\omega} = \frac{1}{n}\sum_{i=1}^{n}\log(z_i), \tag{5.2}$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left[\log(z_i)-\hat{\omega}\right]^2. \tag{5.3}$$

Therefore the main issue of maximizing the likelihood of the covariate SEPP, is to maximize the first part $L_1(\mu,\alpha,\beta,\eta)$. Detailed computations are given in appendix G. The optimization problem requires numerical methods since there are no closed-forms estimators.

Numerical results

We present an example of application of the PLC-SEPP with simulated data. We generate four simulated datasets of PLC-SEPP with different input values to verify the maximum likelihood procedure, with the same sample size $n = 5000$ (event truncation scheme).

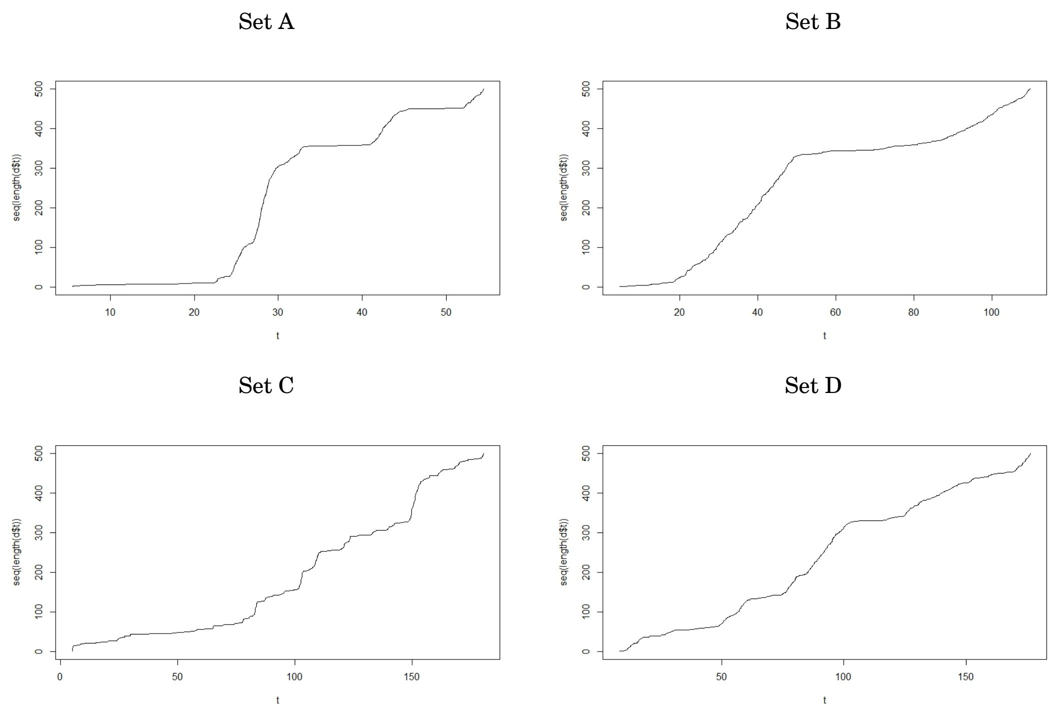
The four sets A, B, C and D of parameters input values are presented in table 5.2. All datasets have the same baseline intensity ($\mu = 0.5$), same branching ratio ($\alpha/\beta = 0.8$) but a different decay rate. The covariate-threshold is set to be $z_0 = 1/10^6$. The marks z follows log-normal distribution with parameters $(2.5, 0.4)$.

The empirical compensators are shown in the figure 5.8. The maximum likelihood estimates are shown in the table 5.3.

Table 5.2: Parameters input values for simulation of the PLC-SEPP

Set	Parameter	μ	α	β	η	σ
A	0.5	4.0	5.0	0.01	2.5	0.4
B	0.5	0.8	1.0	0.01	2.5	0.4
C	0.5	4.0	5.0	0.001	2.5	0.4
D	0.5	0.8	1.0	0.001	2.5	0.4

Figure 5.8: *Empirical compensators of the PLC-SEPP for the different set of parameters values from table 5.2*



5.3 Bayesian approach

To our knowledge no Bayesian strategies have been developed for the classical Hawkes process with exponential decay response function. However for covariate models, some papers can found in the literature.

Table 5.3: *Mean of MLE of the four sets A, B, C, D*

Parameter	μ	α	β	η	ω	σ
Set A	0.5	4.0	5.0	0.001	2.5	0.4
MLE	0.5142	3.5032	4.9810	0.0086	2.5016	0.3996
MSE	(0.0003)	(0.2467)	(0.0001)	(0.000006)	(0.000002)	(0.000001)
Set B	0.5	0.8	1.0	0.001	2.5	0.4
MLE	0.5717	0.8244	1.0587	0.000953	2.4934	0.3953
MSE	(0.0051)	(0.0006)	(0.0034)	(0.00002)	(0.00004)	(0.00002)
Set C	0.5	4.0	5.0	0.01	2.5	0.4
MLE	0.5036	3.5000	4.4999	0.0121	2.4962	0.3960
MSE	(0.0004)	(0.2504)	(0.0053)	(0.0001)	(0.00006)	(0.0000001)
Set D	0.5	0.8	1.0	0.01	2.5	0.4
MLE	0.4251	0.8613	1.0523	0.0116	2.5005	0.4129
MSE	(0.0085)	(0.0063)	(0.0075)	(0.0002)	(0.00001)	(0.0001)

Peruggia & Santner [48] suggest a bayesian methodology to analyze the time evolution of earthquake activity. They consider the epidemic model [43]

$$\lambda^*(t) = \mu + \sum_{t_i<t} e^{\omega(m_i - M_r)} \beta e^{-\alpha(t-t_i)},$$

where t_i, $i = 1,\dots,n$ are the occurrence times, m_i, $i = 1,\dots,n$ are the magnitude of events, and M_r is a structural parameter given by experts, a threshold (prespecified).

The parameters (ω,α,β) characterize the aftershock and are the parameter to be estimated with μ which characterizes the main shock.

The authors choose gamma distributions for prior on the parameters assuming independence except for ω for which they consider a gamma distribution given α, β. To compute the posterior distributions, they develop Markov Chain Monte Carlo (MCMC) algorithms to obtain the posterior.

Another example of Bayesian strategy for SEPP is presented by Ruggeri & Soyer [52] in the context of software reliability.

They introduce a SEPP model where the intensity increase each time a bug is attempted to be fixed. The maintenance introduces new bugs and so on. The repair is imperfect. They suggest the following expression for the intensity:

$$\lambda^*(t) = \mu(t) + \sum_{j=1}^{N(t^-)} Z_j\, g_j(t - t_j),$$

Occurrences of bugs are basically described as a power-law process process with intensity $\mu(t) = M\beta t^{\beta-1}$. Z_j is a Bernoulli r.v. with parameter p_j. $Z_j = 1$ if the repair of the the j^{th} failure (bug) introduced a new bug and 0 otherwise. The response function $g_j(t)$ is supposed to be positive.

Suppose we observe a sequence of jumps $t_1,\dots,t_n$ in a time window $[0,C]$ and $Z_1,\dots,Z_n$. The distribution of the observation is:

$$\begin{aligned} f(\underline{t}\mid\underline{Z};M,\alpha,\beta) &= \prod_{i=1}^{n}\Big[\mu(t_i)+\sum_{j=1}^{i-1}Z_j g(t_i-t_j)\Big] \\ &\times \exp\Big\{-\int_0^C\mu(t)dt-\sum_{j=1}^{N(t^-)}Z_j\int_O^{C-t_j}g_j(t)dt\Big\} \\ &= M^n\beta^n\prod_{i=1}^{n}A_i(\beta,\underline{Z}_{i-1})\,exp\Big\{-M\,B(\beta,\underline{Z}_n)\Big\}, \end{aligned}$$

where $\underline{Z}_i=(Z_1,\dots,Z_i)$, $A_i(\beta,\underline{Z}_{i-1})=t_i^{\beta-1}+\sum_{j=1}^{i-1}Z_j(t_i-t_j)$ and $B(\beta,\underline{Z}_n)=C^\beta+\sum_{j=1}^{N(C^-)}Z_j(C-t_j)^\beta$.

Ruggeri & Soyer [52] propose the following choices for the prior distributions: $M\sim\mathscr{G}(\alpha,\delta)$, $\beta\sim\mathscr{G}(\rho,\lambda)$, and $p_j\sim\mathscr{B}eta(\mu_j,\sigma_j)$, $j=1,\dots,n$.

And the following conditional posteriors distributions are obtained:

- $M\mid\beta,\underline{Z},\underline{p})\sim\mathscr{G}(\alpha+n,\delta+B(\beta,\underline{Z}),$
- $\beta\mid M,\underline{Z},\underline{p})\propto\beta^{\rho+n}\prod_{i=1}^{n}A_i(\beta,\underline{Z}_{i-1})\,e^{-MB(\beta,\underline{Z})-\lambda\beta},$
- $p_j\mid M,\beta,\underline{Z},p_{-j})\sim\mathscr{B}eta(\mu_j+Z_j,\sigma_j+(1-Z_j)),\forall j,$

MCMC methods are used to make inference on the parameters.

Bayesian inference for the PLC-SEPP

To develop techniques for a Bayesian approach of inference on covariate SEPP, we need to specify the functional form of the prior distribution and propose a strategy to elicit values for the hyperparameters.

The full expression of the distribution of the observation is:

$$\prod_{i=1}^{n}\left(\mu+\sum_{t_j<t_i}\left(\frac{z_j}{z_0}\right)^\eta\alpha e^{-\beta(t_i-t_j)}\right).$$
$$\exp\left\{-\left(\mu\tau+\frac{\alpha}{\beta}\sum_{i=1}^{n}\left(\frac{z_i}{z_0}\right)^\eta\left[1-e^{-\beta(\tau-t_i)}\right]\right)\right\}.$$

A prior distribution for $\theta=(\mu,\eta,\alpha,\beta)$ needs to be proposed. Supposing that the parameters are independent, a product of gamma distributions can be considered. Simulations of marginal posterior distributions can then be obtained using a MCMC techniques.

Chapter 6

Application: Thunderstorms

In this chapter we will study thunderstorm data. Our purpose is to model thunderstorm with SEPP. Jumps are impacts. Firstly we describe the available information, then we consider applying the methods described in the previous chapter.

6.1 The Dataset

The data set covers information on thunder lightning impacts in Ardèche and Drome (France) from 1990 to 2010. Date, time, location, electrical charge and some other attributes are recorded for each impact.

The table 6.1 is an extraction of the available database. A total of 676

Table 6.1: *Extraction from the thunderstorm data*

Date	time	latitude in m	longitude in m	amplitude in kA	1/2 major axis in km	1/2 minor axis in km	angle	nb of arcs associed	n^o
09/12/2000	01:06:16	2031878	803282	-10.9	3.0	0.4	126.7	1	1
09/12/2000	01:07:10	2030986	805221	-17.6	0.4	0.2	111.7	2	1
09/12/2000	01:07:10	2033606	804073	-14.7	0.3	0.3	93.2	2	2
09/12/2000	01:07:42	2034443	804815	-24.2	0.3	0.2	114.0	6	1
09/12/2000	01:07:42	2034240	804398	-13.3	0.5	0.3	114.0	6	2
09/12/2000	01:07:42	2034602	804920	-31.1	0.2	0.2	105.6	6	3
09/12/2000	01:07:42	2034435	804902	-36.9	0.2	0.2	111.7	6	4
09/12/2000	01:07:42	2035994	802605	-12.3	3.0	0.4	126.3	6	5
09/12/2000	01:07:43	2034298	804467	-15.2	0.4	0.3	111.9	6	6
09/12/2000	01:08:46	2025072	798175	-17.2	0.5	0.3	112.3	2	1
09/12/2000	01:08:46	2025027	795353	-12.9	0.5	0.4	56.2	2	2
09/12/2000	01:08:47	2025007	795063	-16.0	3.2	0.4	80.2	1	1
09/12/2000	01:09:15	2030484	805842	-17.6	0.4	0.2	111.9	2	1
09/12/2000	01:09:15	2030765	806603	-16.2	0.3	0.2	107.2	2	2
09/12/2000	01:09:29	2031399	805262	-19.8	0.5	0.4	118.5	2	1
09/12/2000	01:09:29	2031192	804381	-15.0	1.5	0.5	78.3	2	2

097 dates of impacts are available. Remark that more than one impact can occur at the same time but in different locations. This is the multiplicity of impacts.

Considering sequences of impacts, two objects can be defined: the process of impacts associated with a given thunderstorm and the global process of thunderstorms (number of thunderstorm in a given year). We define a thun-

Table 6.2: *Numbers of impacts in Ardèche and Drôme for differents periods*

Area	Beginning	end	Impact
Ardèche 1990 – 2001	29/01/1990	11/11/2001	129 852
Ardèche 2002 – 2010	15/03/2002	23/11/2010	190 711
Drôme 1990 – 2001	28/01/1990	18/11/2001	143 875
Drôme 2002 – 2010	24/01/2002	02/12/2010	211 659
		Total	**676 097**

derstorm as a group of lightning impacts that occurs within a period of time in a given area. Therefore to define a thunderstorm we need to consider a time threshold. In a given area (to be defined) when the distance in time between two impacts is less than this threshold, we are going to consider that they belong to the same thunderstorm.

The table 6.3 displays numbers of thunderstorms, their mean duration and mean number of impacts for 2008 in Drôme with respect to different values of the threshold. The table 6.5 presents the number of thunderstorm

Table 6.3: *Some thunderstorms characteristics for the Drôme district in 2008 (only thunderstorms with at least five impacts are retained)*

Threshold	Number of thunderstorms	Mean duration (sd)	Mean number of impacts per thunderstorm
5 h	55	6.4887 (5.1151)	497 (1379)
2 h 30 $'$	64	4.5604 (3.7368)	513 (1293)
1 h	77	3.1392 (3.3232)	426 (1188)
30$'$	98	2.1132 (2.4990)	334 (947)
10$'$	154	1.0049 (1.5180)	211 (732)

per year for a five hours threshold in each French department. The figure 6.1

Table 6.4: Impact multiplicity for thunderstorm on 31 May 2001

Order	1	2	3	4	5	6	7	13	*Total impacts*
Number	325	99	40	20	7	9	3	1	**504**

displays the cumulative intensity (empirical compensator) for thunderstorms process in Drôme from 2002 to 2009. Here the thunderstorms are defined

Table 6.5: *Number of thunderstorms (Threshold: five hours) per year in each department*

	1990	1991	1992	1993	1994	1995	1996
Ardèche	87	62	72	83	70	70	65
Drôme	99	68	90	88	72	85	86

	1997	1998	1999	2000	2001	2002	2003
Ardèche	72	60	83	74	64	77	76
Drôme	73	76	77	75	74	87	86

	2004	2005	2006	2007	2008	2009	2010
Ardèche	73	53	88	68	75	85	67
Drôme	78	67	88	71	76	94	74

with a threshold of 5 hours and characterized by their first jump.

We can consider the thunderstorm that occurs in Drôme on 31 may 2001. This thunderstorm begins at 12:56:17 and end at 17:12:46. It has a total duration of 4 hrs 16 min. 26 sec. and a total number of impacts equals to 846. The multiplicity orders of the impacts are given in table 6.4. With a thresold of $30'$, we define 98 thunderstorms occurring in Drôme department during the year 2008.

The table 6.6 displays 6 examples among these thunderstorms and the figure 6.2 displays the thunderstorm localisation of table 6.6.

6.2 Inference on Some Specific Thunderstorms

In this section, we select some specific thunderstorms to employ the PLC-SEPP model. The data set includes jumps and their associated amplitude of charges. The associated amplitudes are independent to the jumps and follow

Figure 6.1: *Cumulative number of thunderstorms in Drôme per year from 2002 to 2009*

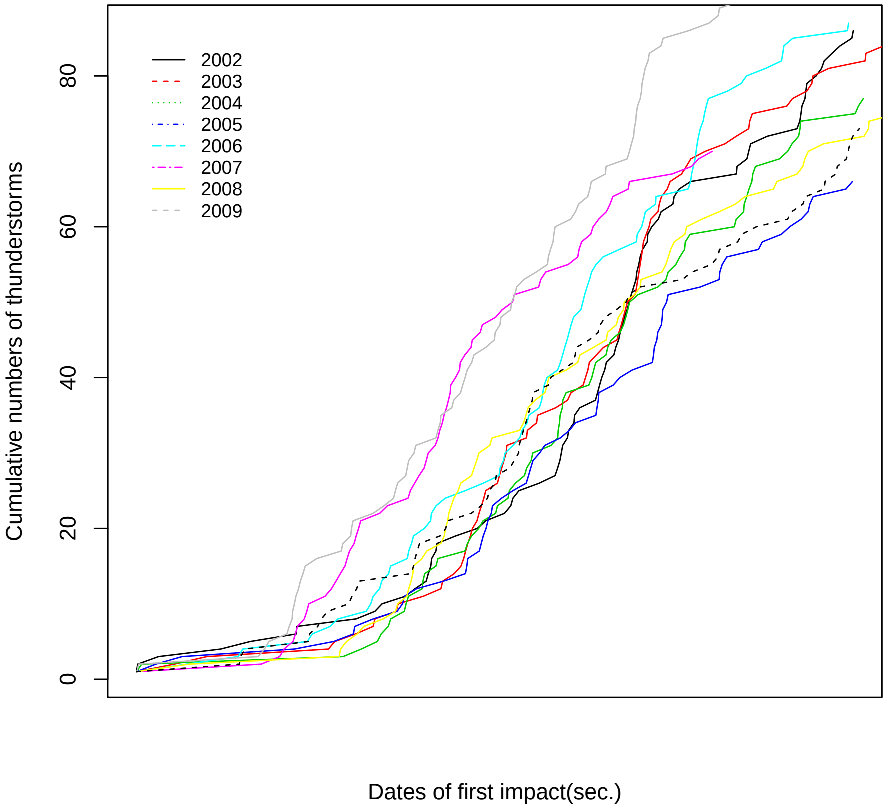


Table 6.6: *Some selected thunderstorms (threshold 30′) in the Drôme department during the year 2008*

	First impact		Last impact		Number	Duration
	Date	time	Date	time	of impacts	of the thunderstorm
4	10/04/2008	20:50:39	10/04/2008	22:27:29	430	01 h 36 ′ 51 ″
23	28/05/2008	21:00:50	28/05/2008	23:32:04	75	02 h 31 ′ 15 ″
27	02/06/2008	12:52:28	02/06/2008	18:08:12	755	05 h 15 ′ 45 ″
42	03/07/2008	05:24:28	03/07/2008	16:10:17	3657	10 h 45 ′ 49 ″
66	12/08/2008	14:59:57	12/08/2008	20:50:35	2557	05 h 50 ′ 38 ″
89	03/10/2008	19:40:04	03/10/2008	20:11:41	19	0 0 h 31 ′ 38 ″

the log-normal distribution.

6.2.1 How to define a thunderstorm?

There are different ways defining a thunderstorm. It is reasonable to consider that a thunderstorm should be a group of impacts that occur in a short period of time around a place. Related to temporal factor, it requires a time-threshold. That is a maximum time-distance of two consecutive impacts to be in the same group. For spatial factor, we localize each grouped impacts on the map to see how they are clustered. One can also take into account the amplitudes of the impacts to discover some pattern in order to group impacts.

At this step, we try our way of defining thunderstorms then fitting them with our PLC-SEPP model described in the previous chapter.

We define a thunderstorm as a group of lightning impacts that occur within a period of time in a given area. We propose a three-step procedure to group impacts into different thunderstorms:

(1) Grouping step.
At the first step, we classify impacts into groups by consider the date of jumps. With a given time-threshold, if two consecutive impacts occur in a period less than this time-threshold then we put them in the same group.

Time-thresholds are varied to see how the recorded impacts spread out in the time-line from 1990 to 2010. The larger time-threshold might give bigger groups with less impacts. We specify a time-threshold of thirty minutes and another time-threshold of five hours.

(2) Clustering step.
At the second step, for each grouped impacts, we localize them on the map taking into account their latitudes and longitudes to see how they cluster in space.

This step gives us different clusters. They might be spread out on the map or might be concentrate at some point.

One can also conduct a clustering analysis to decide the number clusters and then find the center point for each cluster.

(3) Thinning step.
At the last step, we choose some specific center point and consider points nearby within a specific radius. We will consider a cluster with center point as a thunderstorm.

6.2.2 Grouping step: groups of impacts

At the first step, the time-threshold T is varied every thirty minutes from 0 hour to 24 hours to see how the many groups detected with respect to time-threshold (in hour). The table 6.7 shows number of groups in each department. Note that when $T = 0$ the total number of 320,563 impacts in Ardeche

Table 6.7: Number of groups in Ardeche and Drome by time-thresholds

T	Ardeche	Drome	T	Ardeche	Drome
0.0	193574	215539	12.0	1220	1395
0.5	3576	3927	12.5	1206	1373
1.0	2544	2763	13.0	1190	1354
1.5	2122	2340	13.5	1170	1340
2.0	1943	2125	14.0	1159	1324
2.5	1804	1982	14.5	1148	1308
3.0	1713	1899	15.0	1137	1291
3.5	1648	1827	15.5	1118	1275
4.0	1593	1765	16.0	1097	1252
4.5	1558	1729	16.5	1085	1233
5.0	1523	1681	17.0	1064	1216
5.5	1487	1643	17.5	1045	1193
6.0	1455	1620	18.0	1025	1164
6.5	1429	1585	18.5	1008	1135
7.0	1404	1567	19.0	990	1112
7.5	1388	1548	19.5	977	1088
8.0	1364	1529	20.0	964	1059
8.5	1349	1514	20.5	947	1047
9.0	1335	1492	21.0	924	1016
9.5	1316	1470	21.5	908	994
10.0	1295	1454	22.0	894	974
10.5	1276	1441	22.5	883	955
11.0	1261	1418	23.0	866	939
11.5	1240	1405	23.5	853	923

are classified into 193,574 groups and 355,534 impacts of Drome are classified into 215,539 groups. That means more than one impacts can occur at the same time but in different places. The larger time-threshold gives bigger groups (in number of impacts) but less number of groups. With threshold of five hours ($T = 5$) we have 1,523 groups in Ardeche and 1,681 groups in Drome whereas with threshold of thirty minutes ($T = 0.5$) we have 3,576 groups in Ardeche and 3,927 groups in Drome.

We describe here some specific groups in the two departments Ardeche and Drome. Each group is specified by their typical features: Date and time begin, date and time end, number of impacts without replicates, number of impacts without replicates, duration. The table 6.8 describes four groups (threshold five hours) in Ardeche. Group 1 is the first group which occurred

Table 6.8: *Some groups (threshold five hours) in the Ardeche department*

Group	First impact		Last impact		Impact		Duration
	Date	time	Date	time	non-rep	rep	
1	28/01/1990	11:13:57	28/01/1990	16:16:04	171	200	05 h 02 ′ 07 ″
1058	17/08/2004	01:56:22	18/08/2004	00:21:37	5596	13796	22 h 25 ′ 05 ″
999	28/08/2003	20:51:20	29/08/2003	10:22:20	4631	11469	13 h 31 ′ 00 ″
1354	04/09/2008	15:54:33	05/09/2008	00:15:41	2410	5972	08 h 21 ′ 08 ″

at the end of January, 1990 and it lasted more than four hours within a day. Group 1058 is the biggest group (in number of impacts) and it lasted from day 17 to day 18 August, 2004. Group 999 and group 1354 are the second and the third biggest group, respectively. Both of them started and ended in different days.

The figure 6.3 displays the groups localization of table 6.8. One can see that group 1 has little number of impacts but it is very scattered. Group 1058, due to its large number of impacts, scattered everywhere in Ardeche but more concentrate in the strip from south-west to north-east. Group 999 is also large-scattered and it is more dense in the north-east of the department. Group 1354, despite the large number of impacts, is quite concentrate in a small part (north-east) of the department.

The table 6.9 describes four groups (threshold five hours) in Drôme and the figure 6.4 displays the localization of those groups. The description of those group in Drome is similar the those in Ardeche.

Table 6.9: *Some groups (threshold five hours) in the Drôme department*

Group	First impact		Last impact		Impact		Duration
	Date	time	Date	time	non-rep	rep	
1	28/01/1990	13:40:22	28/01/1990	17:54:14	35	43	04 h 13 ′ 52 ″
1494	04/09/2008	07:49:07	04/09/2008	02:07:04	3492	8360	18 h 17 ′ 57 ″
1113	28/08/2003	21:19:02	29/08/2003	14:00:34	3379	7760	16 h 41 ′ 32 ″
1105	17/08/2003	14:21:36	17/08/2003	18:07:40	2154	6951	03 h 46 ′ 04 ″

6.2.3 Clustering step: clusters of impacts

At this step, we choose some specific groups from the first step that have some clustering pattern. We choose the year 2008 to pickup some groups in Drome. The table 6.10 displays six clusters in Drome (threshold 30′) and the figure 6.5 displays the localization of groups in the table 6.10. This time we specify six towns (Valence, Montelimar, Nyons, Mevouillon, Saillans, and Die) in Drome department to better describe how each group cluster then we choose some specific clusters around a town.

Cluster 4 and cluster 23 is spread out and have no clustering pattern. Group 42, due to the numerous impact, is also spread out but is more clustered in the south-west of the department. One can see that cluster 66 mostly concentrates in two village Montelimar and Nyons. cluster 27 has a medium size and mostly concentrate in the south-east of the department. Cluster 89 is small and occurred in a small area of the north of the department.

Table 6.10: *Some clusters (threshold 30′) in the Drôme department in 2008*

Cluster	First impact		Last impact		Impact	Duration
	Date	time	Date	time		
4	10/04/2008	20:50:39	10/04/2008	22:27:29	430	01 h 36 ′ 51 ″
23	28/05/2008	21:00:50	28/05/2008	23:32:04	75	02 h 31 ′ 15 ″
27	02/06/2008	12:52:28	02/06/2008	18:08:12	755	05 h 15 ′ 45 ″
42	03/07/2008	05:24:28	03/07/2008	16:10:17	3657	10 h 45 ′ 49 ″
66	12/08/2008	14:59:57	12/08/2008	20:50:35	2557	05 h 50 ′ 38 ″
89	03/10/2008	19:40:04	03/10/2008	20:11:41	19	00 h 31 ′ 38 ″

6.2.4 Thinning step: thunderstorms

We choose five center points at Montelimar, Nyons, Saillans, Vallence in the cluster 66 and consider only the impacts nearby within the radius of 20 kilometers. This step give us six clusters with specific center point that we call thunderstorms.

The table 6.11 displays five thunderstorms in Drome and the figure 6.6 displays the localization of those groups.

6.2.5 Inference on the selected thunderstorms

The figure 6.7 shows some selected thunderstorms with date of impacts and the associated charges on the time line. The figure 6.8 shows the multiplicity of the selected thunderstorms. For the selected thunderstorms, one of two models of self-exciting point process will be a candidate: some thunderstorms are fitted to the Hawkes process and some others are fitted to the power-law covariate self-exciting point process.

Table 6.11: *Six thunderstorms (threshold 30′, radius 20 km) in the Drôme department in 2008*

Thundertorm	First impact		Last impact		Impact	Duration
	Date	time	Date	time		
Montelimar	12/08/2008	16:19:12	12/08/2008	19:00:03	850	02 h 40 ′ 51 ″
Nyons	12/08/2008	18:50:23	12/08/2008	20:25:32	913	01 h 35 ′ 09 ″
Saillans	12/08/2008	17:19:30	12/08/2008	19:37:09	139	02 h 17 ′ 39 ″
Die	12/08/2008	17:37:20	12/08/2008	19:43:09	57	02 h 05 ′ 49 ″
Valence	12/08/2008	14:59:57	12/08/2008	18:08:06	23	03 h 08 ′ 09 ″
Mevouillon	12/08/2008	19:41:41	12/08/2008	20:50:35	184	01 h 08 ′ 54 ″

6.2.6 Fitting thunderstorms to the Hawkes process

Considering a thunderstorm as a temporal point process, that means we consider only the date of jumps of the thunderstorm to fit it with Hawkes process. For a multiple impact we keep only the unique jump. Since we do not know the starting point of the process, the first impact of the thunderstorm is consider as starting point of the self-exciting point process. The figure 6.9 shows the empirical compensators of the selected thunderstorms. The table 6.12 gives estimates of parameters of Hawkes process for some thunderstorms in Drôme.

Table 6.12: *Parameter estimated of Hawkes model for 3 thunderstorms in table 6.11*

Thundertorm	n	$\hat{\mu}$	$\hat{\alpha}$	$\hat{\beta}$
Montelimar	338	0.0018	-0.0087	0.0091
Nyons	325	0.0569	1689964	1687129
Saillans	139	0.005	-0.2931	0.9796

6.2.7 Fitting thunderstorms to the the power-law covariate self-exciting point process

We now fit some other selected thunderstorms to the PLC-SEPP model taking into account the associated amplitudes. Since charges could be positive or negative, we take their absolute values. For a multiple impact, we keep only the unique jump so it needs to average of the absolute value of the associated charges. The figure 6.10 shows the histograms of the charges of the selected thunderstorms.

Table 6.13: *Parameters estimated of log-normal distributions of charges of the selected thunderstorms*

Thundertorm	n	$\hat{\omega}$	$\hat{\sigma}$
Montelimar	850	2.91	0.40
Nyons	913	2.92	0.46
Saillans	139	2.94	0.59
Die	57	2.94	0.56
Valence	32	2.94	0.44
Mevouillon	184	3.12	0.57

Table 6.14: *Estimates of parameters of PLC-SEPP model for five thunderstorms in table 6.11*

Thundertorm	n	$\hat{\mu}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\eta}$
Montelimar	338	0.0349	3.1510	48.4244	-990.9540
Nyons	325	0.0568	3.8040	48.0980	-961.0424
Saillans	62	0.0075	16.9808	41.5096	-179.6893
Die	27	0.0036	18.7389	40.6305	-78.2097
Valence	15	0.0008	19.3402	40.3299	-37.7755
Mevouillon	75	0.0181	16.3259	41.8370	-230.9308

6.3 Concluding Remarks

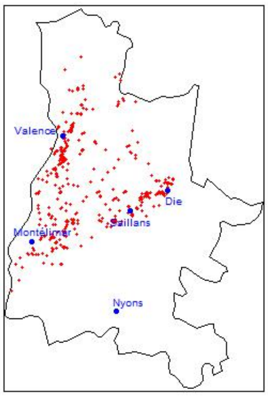
Providing the dataset of thunderstorms collected in the two French regions, we have proposed a way to define a thunderstorm. At the first step we separate the impacts into different groups comparing the time-length between two consecutive impacts to a time-threshold. We find that some groups are have a large number of impacts and some have only one impact. Then we visualize the groups on the maps of the two regions to see the clusters. We see that some groups are very clustered in space and some are very spread-out. Finally, we localize the clusters nearby specific center points and obtain thunderstorms. We discover that some thunderstorms are big while some others are very small (in duration and number of impacts).

Some thunderstorms have been chosen to be analyzed. We consider a thunderstorm as a stochastic point process where date of jumps are associated by the charges occurring nearby in time and space. The multiplicity impacts are also presented.

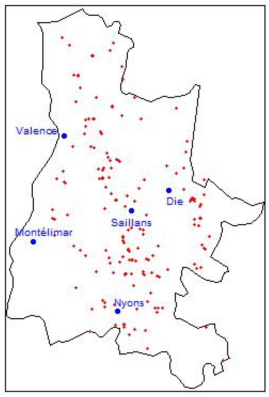
Considering date of jumps of a thunderstorm as a temporal point process, the analysis show that some thunderstorms follow to the Hawkes process. Then taking into account both date of jumps and the associated charges, some thunderstorms are fitted to the power-law covariate self-exciting process. While the charges can be negative or positive, we consider only their absolute values. It is interesting to find that the charges of thunderstorms follow log-normal distribution.

The spatio-temporal point process might be another candidate model for this dataset. We will work on this direction in the future.

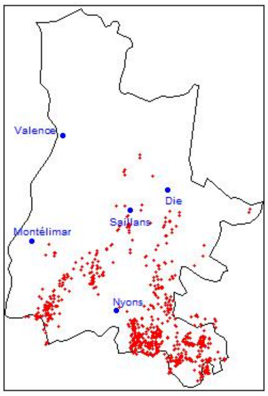
Figure 6.2: *Impacts localization for thunderstorms from table 6.6 (Drôme)*



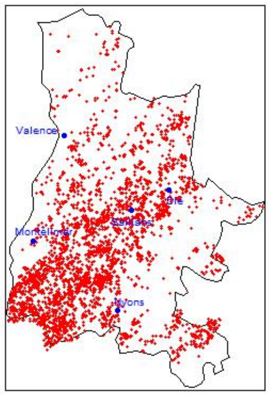
10 April 2008
430 impacts - Duration: 1h 36′51″



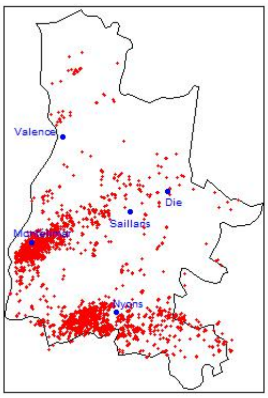
28 May 2008
168 impacts - Duration: 4h 14′11″



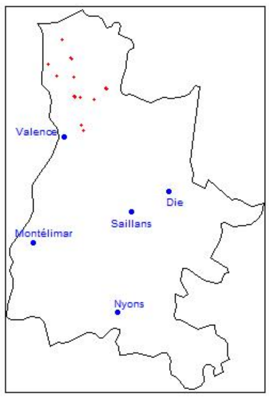
2 June 2008
755 impacts - Duration: 5h 15′45″



3 July 2008
3657 impacts - Duration: 10h 45′49″



12 August 2008
2557 impacts - Duration: 5h 50′38″



31 October 2008
19 impacts - Duration: 31′38″

Figure 6.3: *Impacts localization for groups in table 6.8*

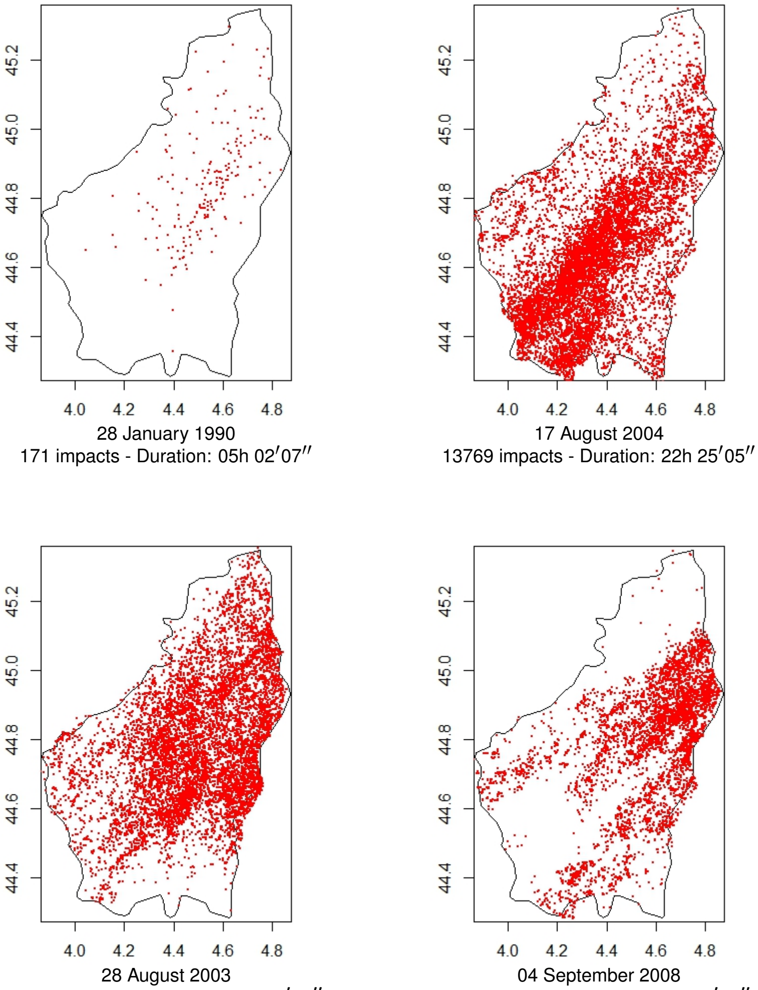


Figure 6.4: *Impacts localization for groups in table 6.8*

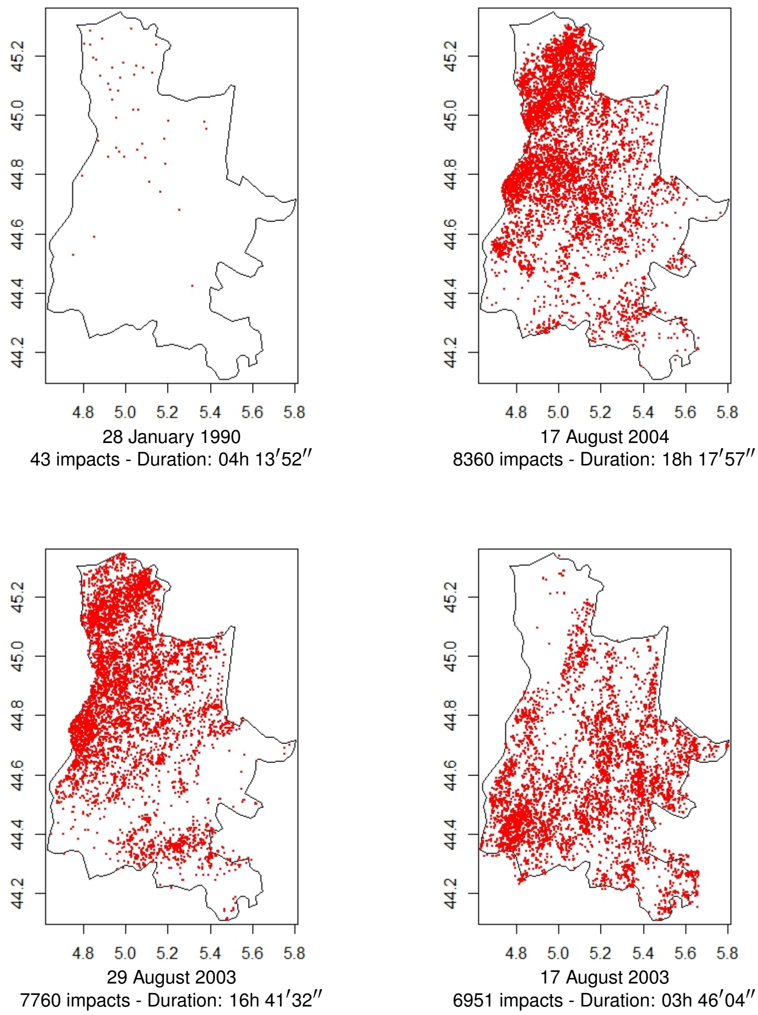
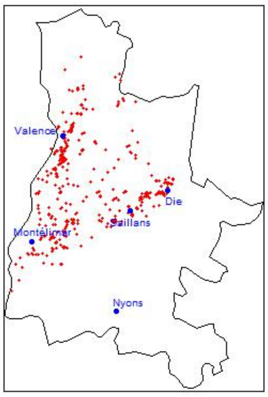
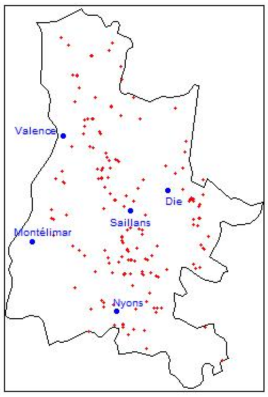


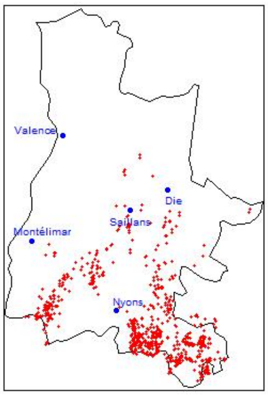
Figure 6.5: *Impacts localization for clusters of table 6.10 (Drôme)*



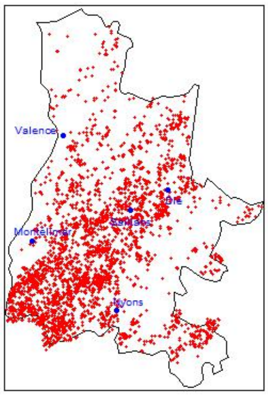
10 April 2008
430 impacts - Duration: 1h 36′51″



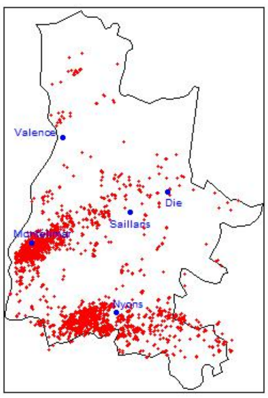
28 May 2008
168 impacts - Duration: 4h 14′11″



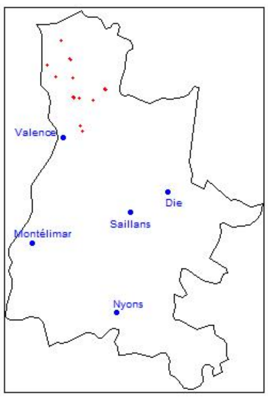
2 June 2008
755 impacts - Duration: 5h 15′45″



3 July 2008
3657 impacts - Duration: 10h 45′49″

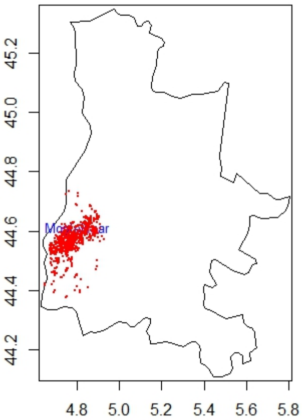


12 August 2008
2557 impacts - Duration: 5h 50′38″

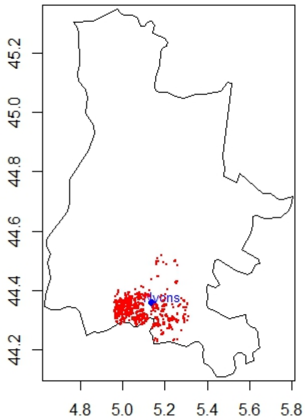


31 October 2008
19 impacts - Duration: 31′38″

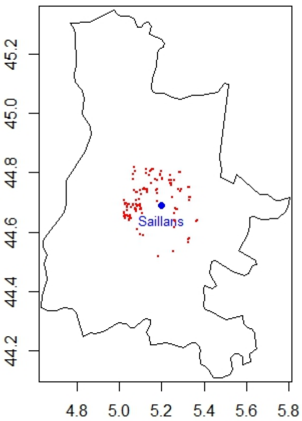
Figure 6.6: *Impacts localization for thunderstorms from table 6.11 (Drôme)*



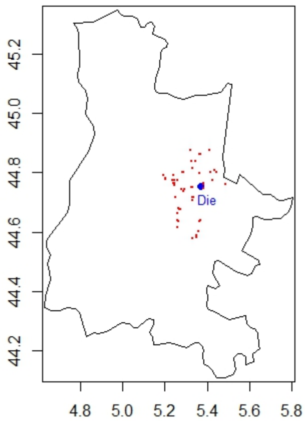
Montelimar: 12 August 2008
850 impacts - Duration: 03h 19′09″



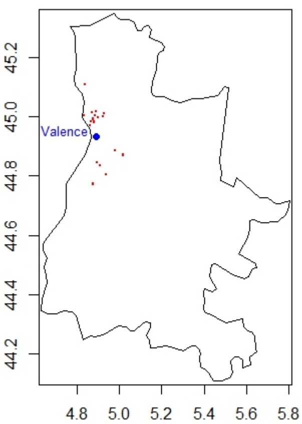
Nyons: 12 August 2008
913 impacts - Duration: 01h 35′09″



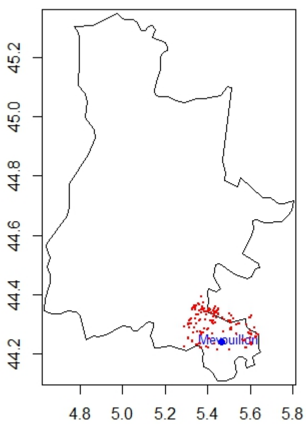
Saillans: 12 August 2008
139 impacts - Duration: 02h 17′39″



Die: 12 August 2008
57 impacts - Duration: 02h 05′49″

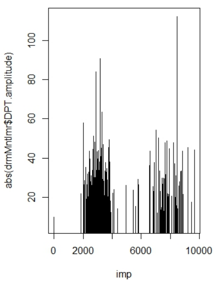


Valence: 12 August 2008
23 impacts - Duration: 03h 08′09″

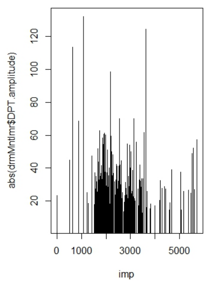


Mevouillon: 12 August 2008
184 impacts - Duration: 01h 08′54″

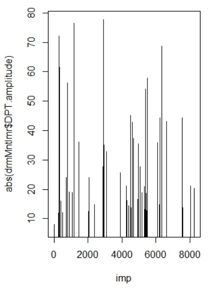
Figure 6.7: *Some selected thunderstorms with date of impacts and the associated charges*



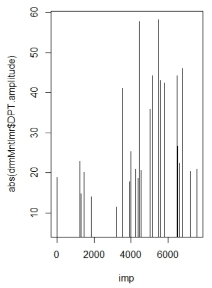
Montelimar: 12 August 2008



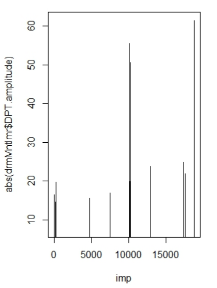
Nyons: 12 August 2008



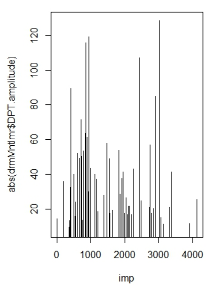
Saillans: 12 August 2008



Die: 12 August 2008

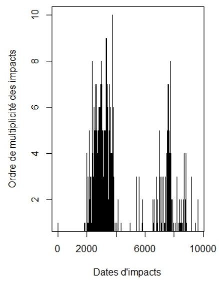


Valence: 12 August 2008

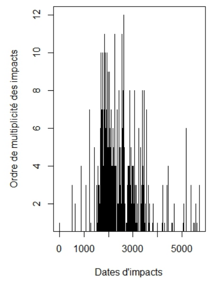


Mevouillon: 12 August 2008

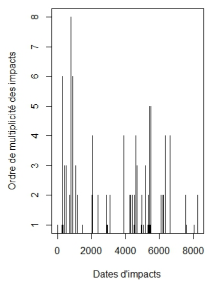
Figure 6.8: *Multiplicity of the selected thunderstorms*



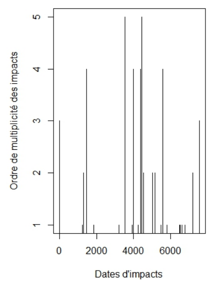
Montelimar: 12 August 2008



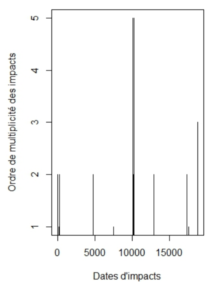
Nyons: 12 August 2008



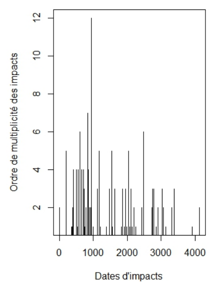
Saillans: 12 August 2008



Die: 12 August 2008

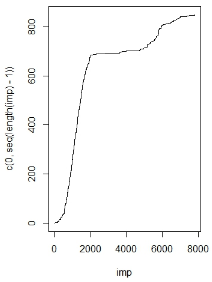


Valence: 12 August 2008

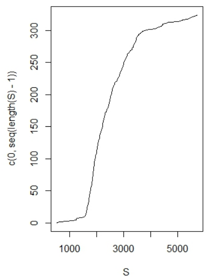


Mevouillon: 12 August 2008

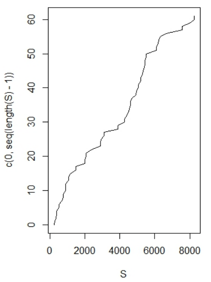
Figure 6.9: *Empirical compensators of the selected thunderstorms*



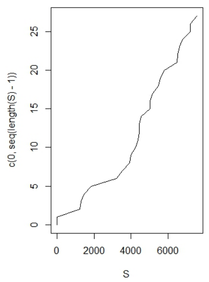
Montelimar: 12 August 2008



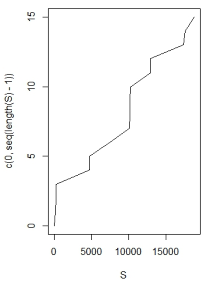
Nyons: 12 August 2008



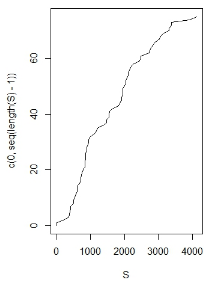
Saillans: 12 August 2008



Die: 12 August 2008

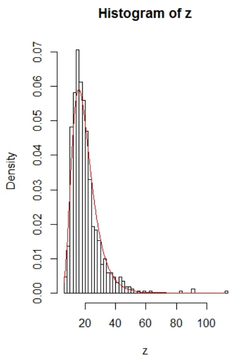


Valence: 12 August 2008

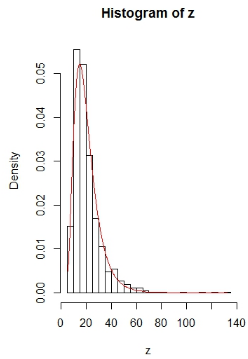


Mevouillon: 12 August 2008

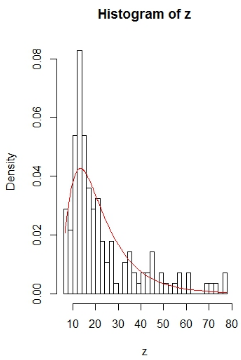
Figure 6.10: *Histograms of the charges of the selected thunderstorms*



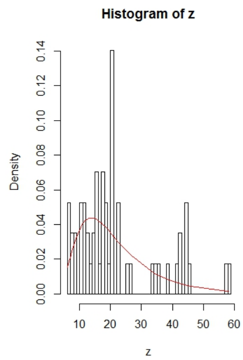
Montelimar: 12 August 2008



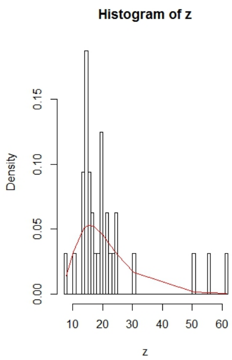
Nyons: 12 August 2008



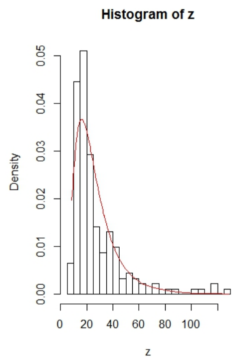
Saillans: 12 August 2008



Die: 12 August 2008



Valence: 12 August 2008



Mevouillon: 12 August 2008

Chapter 7

General Conclusion

Throughout this thesis, we have studied some stochastic processes and develop some models to investigate the thunderstorm data. The data set is rich so there are many ways to define a thunderstorm and many available models that could fit in it. We consider a thunderstorm as a group of impacts that occur in a short period of time and cluster near by a place. Each impact is associated with a charge that can be positive or negative.

Like the phenomenon of earthquake that including main shocks and aftershocks, the phenomenon of thunderstorm including big impacts and small impacts. That is, a big impact might trigger more impacts right after and nearby. Intuitively, based on the physical nature of the phenomenon of thunderstorms, we think of the self-exciting point process that have been used widely to model the clustering point patterns. While considering the impacts with charges that exceed a fixed amplitude-threshold, their jumps might occur more frequently or less frequently depending on the condition of the atmosphere. Therefore, we think of the power-law process and the exponential-law process that have been widely used in modeling improving or aging repairable systems. Among many approaches to deal with stochastic processes, we use the notion of conditional intensity function as the main tool for our work. This concept is convenient for both classical inference and Bayesian inference for stochastic processes. Inference issues lead us to optimization problems and we have suggested some efficient strategies to deal with.

In chapter 3, the power-law process has been studied in details. The intensity of this process depends only on the present time with power-law form. This convenient form of the intensity function allows us to investigate some typical aspects of stochastic processes analysis such as graphical test for the adequacy, distribution of inter-events, distributions of maximum likelihood estimators, conjugate prior for Bayesian analysis.

The literature on this process is abundant. We have done some more detailed work that could be more practical for practitioners. We have presented some convenient probabilistic properties of maximum likelihood estimators for time truncation scheme. For Bayesian approach we have considered three different parametrizations. We started from a non-informative prior to an independent conjugate prior and obtain a joint conjugate prior (that is, with dependency of the two parameters). The joint conjugate prior belongs to a family of bivariate distributions that we name H-B. The properties of H-B distribution are given with explicit expressions. Considering schemes of prior information, we have proposed different elicitation strategies to obtain values for the hyper-parameters. A simulation study has been conducted using our proposed prior with associate strategies and the results show the advantages of our method. Bayesian estimates assuming quadratic loss are shown to be the convex combination of the MLE and the prior expectation.

Our contribution in this chapter has been to propose and study the H-B distribution as the natural conjugate prior for the PLP. This work completes the results presented by Huang & Bier [27] who consider the same problem. Their proposed distribution is not a conjugate prior since the posterior distribution and the prior distribution do not belong to the same family.

In chapter 4, we have studied the exponential-law process. Although this process has an intensity function as simple as the power-law process, the form of the compensator makes it difficult to maximize the likelihood analytically. That is, there are no closed-forms for the maximum likelihood estimators. A maximum likelihood procedure has been proposed and investigated by a simulation study.

Similarly to the previous chapter, we have developed a sequence of priors for Bayesian approach, from a non informative prior to a joint conjugate prior the process. We consider the Jeffrey's rule for non informative prior by computing the Fisher information matrix. The associated posterior leads us to a new univariate distribution that we name *modified-Gumbel* (m-G) distribution. By a new parametrization relying on the value of the compensator at the end time, we obtain independent conjugate priors as a product of two distributions: a *gamma distribution* and a *m-G distribution*. Another joint distribution that we have called *Gamma-modified-Gumbel* (G-m-G) distribution is obtained considering dependency between the two parameters. We have given some properties of this bivariate distribution to prepare the elicitation while employing the it as joint conjugate prior. Then we have suggested appropriate elicitation strategies that need numerical approximation with trials and errors methods. This study on the exponential-law process gives us some clues to deal with a process that requires numerical approximation to maximize the likelihood.

A simple graphical test as Duane plot is not available for the ELP. Howerver, a residual test can be conducted for goodness-of-fit. While the power-law process is easily simulated thanks to its explicit simulation equation, the exponential-law process requires again a method of approximation because of its implicit simulation equation.

Our contribution in this chapter are: introducing the maximum likelihood procedure for the ELP, proposing the m-G distribution as an independent conjugate prior for the ELP, proposing and studying the G-m-G distribution as the natural conjugate prior for the PLP. Huang and Bier ([28]) also propose a prior distribution for the ELP but it is not a conjugate prior. Moreover, they consider only the ELP with $\beta > 0$ to model aging systems. We consider a more general model of ELP when β can be positive or negative to model both improving and aging systems.

In chapter 5, we have investigated the self-exciting point process. The Hawkes model of the self-exciting point process has been reconsidered with a modified method to maximize the likelihood. We have introduced some covariate models of the self-exciting point process and focused on the power-law covariate model that would be fitted to the thunderstorm data in the next chapter. The model has been investigated by maximum likelihood procedure. A simulation procedure for the model has been built to examine the maximum likelihood estimated.

In the last chapter, we have conducted an application of our proposed model to the thunderstorm data. Firstly, the data has been described with some typical features. We have suggested a three-step strategy to define a thunderstorm. This strategy allows us to divide impacts into groups providing a time-threshold. Each group has been presented on the maps of the two French regions (Ardèche and Drôme) to investigate how it clusters in space. A thunderstorm has been defined as a cluster on the map with a fixed radius nearby a fixed center point. Thunderstorms have been classified by their typical features such as the duration, the number of impacts.

Some selected thunderstorms have been fitted to the Hawkes model of the self-exciting point process considering their dates of jumps. Other selected thunderstorms have been fitted to the power-law covariate model the self-exciting point process taking into account the associated charges. We take the absolute values of the charges then remove the multiplicity by averaging the charges of a multiple impact. The model showed that in most of the case, the charges of the thunderstorms follow the log-normal distribution. Estimated parameters of the model have been given for some selected thunderstorms.

Perspectives

We are developing a Shiny package to present several ways of defining a thunderstorm. More thunderstorms need to be analyzed employing our proposed model then we can conduct a classification to those thunderstorms based on their estimated parameters.

The existence and the uniqueness of maximum likelihood estimators of the power-law covariate model of the self-exciting point process need to be investigated by considering the Hessian matrix to assure out method. We are working on this problem at the moment. A strategy to choose a reasonable initial values for the maximum likelihood procedure of the model is also an important point to develop.

In the future work, we will consider the Bayesian inference for the power-law covariate model of the self-exciting point process. It requires Monte-Carlo Markov Chain method to make inference on the posterior.

We are thinking of employing the Omori-form covariate model of the self-exciting point process to make a comparison with the power-law covariate model. We consider also the spatio-temporal stochastic process for the data of thunderstorms. This process allow us to model the spatial factor of thunderstorms.

We are also thinking of applying the functional data analysis for classifying thunderstorms.

We can say that our work implies many different topics. It has been a great challenge and experience of learning for us. We have settle through this work the basis for further research. A lot of points need to be investigated more closely and that could be finalized with research papers.

Appendices

Appendix A

Goodness-of-fit test for the power-law process

([20]) For a set of data $(x_i, y_i), i \in \{1, \ldots, n\}$, the linear regression problem is to determine a linear relationship between x_i and y_i, and to study the strength of this relationship. The simple linear regression model is

$$y_i = a_1 x_i + a_0 + \epsilon_i, \quad i \in \{1, \ldots, n\}. \tag{A.1}$$

a_1 and a_0 are the slope and intercept of the regression line, respectively; ϵ_i is an error term or residual.
An important aspect of regression analysis is to study the GOF of the model. For linear-regression, this is commonly measured by

$$R^2 = \frac{\left[\frac{1}{n}\sum_{i=1}^{n}(x_i \cdot y_i) - \bar{x}_n \cdot \bar{y}_n\right]^2}{\left[\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}_n^2\right] \cdot \left[\frac{1}{n}\sum_{i=1}^{n} y_i^2 - \bar{y}_n^2\right]}$$

$$\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \bar{y}_n = \frac{1}{n}\sum_{i=1}^{n} y_i.$$

A value of R^2 close to 1 indicates that the hypothesis of a linear relationship between x_i and y_i is acceptable. To build a GOD-test based on R^2, its distribution under the H_0 has to be known. This is the case, for example, for the Gaussian linear model.
The following theorem shows the independence of the distribution of R^2 on model parameters. For a Duane plot, the T_i are plotted versus i on a log-log scale; thus in (A.1),

$$y_i = \ln(i), \quad x_i = \ln(T_i).$$

Because the α and β only change, respectively, the scale and slope of the Duane plot, they should not affect the fitting. This theorem validates that fact.

Theorem A.0.1. *For a Duane plot:*

$$
\begin{aligned}
R^2 &= \frac{\Psi_N^2}{\Psi_{D_1} \cdot \Psi_{D_2}};\\
\Psi_N &= \sum_{i=1}^{n} (\ln(T_i) \cdot \ln(i)) - \frac{1}{n} \cdot \left(\sum_{i=1}^{n} \ln(T_i)\right) \cdot \left(\sum_{i=1}^{n} \ln(i)\right);\\
\Psi_{D_1} &= \sum_{i=1}^{n} (\ln(T_i))^2 - \frac{1}{n} \cdot \left(\sum_{i=1}^{n} \ln(T_i)\right)^2;\\
\Psi_{D_2} &= \sum_{i=1}^{n} (\ln(i))^2 - \frac{1}{n} \cdot \left(\sum_{i=1}^{n} \ln(i)\right)^2;
\end{aligned}
\tag{A.2}
$$

and distribution of R^2 under the PLP hypothesis is independent of α and β.

Proof 13. Under H_0 (the underlying process is a PLP), the vector

$$
\left(\beta \ln\left(\frac{T_n}{T_{n-1}}\right), \beta \ln\left(\frac{T_n}{T_{n-2}}\right), \ldots, \beta \ln\left(\frac{T_n}{T_1}\right)\right)
$$

is distributed as the order statistics of a sample of size $n-1$ from the exponential distribution with parameter 1.
Then, for all $i \in \{1, \ldots, n\}$, let

$$
\begin{aligned}
U_i^* &= \beta \ln\left(\frac{T_n}{T_{n-1}}\right) = -\beta \ln\left(\frac{T_{n-1}}{T_n}\right).\\
U_n^* &= 0.
\end{aligned}
$$

The Ψ_N in (A.2) can be written as follows:

$$
\begin{aligned}
&\sum_{i=1}^{n}(\ln(T_i)\ln(i)) - \frac{1}{n}\left(\sum_{i=1}^{n}\ln(T_i)\right)\left(\sum_{i=1}^{n}\ln(i)\right)\\
=&\sum_{i=1}^{n}(\ln(T_i)\ln(i)) - \ln(T_n)\sum_{i=1}^{n}\ln(T_i) + \ln(T_n)\sum_{i=1}^{n}\ln(T_i)\\
&-\frac{1}{n}\left(\sum_{i=1}^{n}\ln(T_i)\right)\left(\sum_{i=1}^{n}\ln(i)\right)\\
=&\sum_{i=1}^{n}(\ln(T_i)-\ln(T_n))\ln(i) + \left(\sum_{i=1}^{n}\ln(i)\right)\left(\ln(T_n)-\frac{1}{n}\sum_{i=1}^{n}\ln(T_i)\right)\\
=&\sum_{i=1}^{n}\ln\left(\frac{T_i}{T_n}\right)\ln(i) + \frac{1}{n}\left[\sum_{i=1}^{n}\ln(i)\right]\left[\sum_{i=1}^{n}\ln\left(\frac{T_i}{T_n}\right)\right]\\
=&\frac{1}{\beta}\left[-\sum_{i=1}^{n}U_{n-i}^{*}\ln(i) + \frac{1}{n}\left(\sum_{i=1}^{n}\ln(i)\right)\left(\sum_{i=1}^{n}U_{n-i}^{*}\right)\right]\\
=&-\frac{1}{\beta}\left[\sum_{i=1}^{n-1}U_{n-i}^{*}\ln(i) - \frac{1}{n}\left(\sum_{i=1}^{n}\ln(i)\right)\left(\sum_{i=1}^{n}U_{n-i}^{*}\right)\right].
\end{aligned}
$$

Furthermore,

$$
\frac{1}{\beta^2}\sum_{i=1}^{n-1}(U_{n-i}^{*})^2 - \frac{1}{n}\left(\sum_{i=1}^{n-1}\sum_{i=1}^{n-1}U_{n-i}\right)^2
$$

$\square$

Appendix B

Non-informative prior for PLP

We consider two types of non-informative vague priors for α and β. We motivate our choice of such priors by just considering the time T_1 to the first failure. This has a Weibull distribution with scale parameter α and shape parameter β. The use of the distribution of T_1 allows us to use the location-scale model and is justified by the fact that our prior knowledge is not affected by whether we observe T_1 or $T_1, T_2, \ldots, T_n$. Indeed, the distribution of $\log T_1$ can be written in the form of a location-scale distribution, that is, the distribution of $\log T_1$ is called Gumbell distribution with location parameter $\mu = \log \alpha$ and scale parameter $\sigma = \beta^{-1}$.

Lemma B.0.1. *If the random variable X has the Weibull distribution with parameters (α, β) then the random variable $Y = \log(X)$ has Gumbell distribution with parameters (μ, σ) where $\mu = \log(\alpha), \quad \sigma = 1/\beta$.*

Recall that if X follows the distribution Weibull(α, β) then its survival function is

$$S_X(t) = Pr(X > t) = \exp\left\{-(t/\alpha)^\beta\right\}.$$

The survival function of $Y = \log X$ is then

$$\begin{aligned} S_Y(t) = Pr(Y > t) = Pr(\log(X) > t) = Pr(X > e^t) \\ = \exp\left\{-\left(e^t/\alpha\right)^\beta\right\} = \exp\left\{-\exp\left(\frac{t - \log(\alpha)}{1/\beta}\right)\right\}. \end{aligned}$$

Hence, $\log(X)$ follows the distribution Gumbell(μ, σ) with $\mu = \log(\alpha)$ and $\sigma = \beta^{-1}$.

Following Jeffrey's rule for non-informative priors in the location-scale situation (see Box & Tiao [4], pp.56-57), we consider two cases:

(a) μ and σ are known to be independent a priori,

(b) The prior independence assumption is ignored.

For the case (a), the non-informative prior of (μ,σ) will be proportional to σ^{-1}, it means $\pi(\mu,\sigma) \propto \sigma^{-1}$.
For the case (b), the non-informative prior of (μ,σ) will be proportional to σ^{-2}, it means $\pi(\mu,\sigma) \propto \sigma^{-2}$.
By changing variables $\mu = \log(\alpha)$, $\quad \sigma = \beta^{-1}$ the Jacobien is $J = \alpha^{-1}\beta^{-2}$. Applying the rule of substitution we have $\pi(\alpha,\beta) \propto (\alpha\beta)^{-1}$ for the case (a) and $\pi(\alpha,\beta) \propto \alpha^{-1}$ for the case (b). For notational convenience we consider the prior $\pi(\alpha,\beta) \propto (\alpha\beta^{\gamma})^{-1}$ and get the following theorem

Theorem B.0.1. *Denote $\underline{t} = (t_1,\ldots,t_n)$ a realization of the PLP that we observe in the time window $[0,C]$. Using Jeffrey non-informative prior with density $\pi(\alpha,\beta) \propto (\alpha\beta^{\gamma})^{-1}$ we have the following statements hold:*

(i) The posterior distribution of α,β is

$$p(\alpha,\beta \mid \underline{t}) = c_{\gamma}(\underline{t})\alpha^{-(n\beta+1)}\beta^{n-\gamma}\left(\prod_{i=1}^{n} t_i\right)^{\beta} \exp\left\{-(y/\alpha)^{\beta}\right\}$$

where

$$c_{\gamma}(\underline{t}) = \frac{\left(\sum_{i=1}^{n}\log(y/t_i)\right)^{n-\gamma}}{\Gamma(n)\Gamma(n-\gamma)}.$$

(ii) The marginal posterior density of β is Gamma distribution with parameters $(n-\gamma, l_n)$ with $l_n = \sum_{i=1}^{n}\log(y/t_i)$.

(iii) The marginal posterior density of α is

$$p(\alpha \mid \underline{t}_n) \propto \alpha^{-1}\int_{0}^{+\infty} \beta^{n-1}\left(\prod_{i=1}^{n} t_i/\alpha\right)^{\beta} \exp\left\{-(y/\alpha)^{\beta}\right\} d\beta \qquad \text{(B.1)}$$

Proof 14. The probability of a realization is

$$f(\underline{t} \mid \alpha,\beta) = \alpha^{-n\beta}\beta^{n}\left(\prod_{i=1}^{n} t_i\right)^{\beta-1} \exp\left\{-(y/\alpha)^{\beta}\right\}$$

Applying Bayes' formula we have

$$p(\alpha,\beta \mid \underline{t}) \propto \pi(\alpha,\beta)f(\underline{t}_n \mid \alpha,\beta) = \alpha^{-(n\beta+1)}\beta^{n-\gamma}\left(\prod_{i=1}^{n} t_i\right)^{\beta} \exp\left\{-(y/\alpha)^{\beta}\right\}. \qquad \text{(B.2)}$$

Therefore

$$p(\alpha,\beta \mid \underline{t}_n) = c_{\gamma}(\underline{t}_n)\alpha^{-(n\beta+1)}\beta^{n-\gamma}\left(\prod_{i=1}^{n} t_i\right)^{\beta} \exp\left\{-(y/\alpha)^{\beta}\right\}$$

where

$$c_\gamma(\underline{t})^{-1} = \int_0^\infty \int_0^\infty \alpha^{-(n\beta+1)} \beta^{n-\gamma} \left(\prod_{i=1}^n t_i \right)^\beta \exp\left\{ -(y/\alpha)^\beta \right\}.$$

Applying the Fubini's theorem we have

$$I = \int_0^\infty \int_0^\infty \beta^{n-\gamma} \left(\prod_{i=1}^n t_i \right)^\beta \alpha^{-(n\beta+1)} \exp\left\{ -(y/\alpha)^\beta \right\} d\alpha d\beta \tag{B.3}$$

$$= \int_0^\infty \beta^{n-\gamma} \left(\prod_{i=1}^n t_i \right)^\beta \left\{ \int_0^\infty \alpha^{-(n\beta+1)} \exp\left\{ -(y/\alpha)^\beta \right\} d\alpha \right\} d\beta \tag{B.4}$$

Let's consider the integral

$$I(\beta) = \int_0^\infty \alpha^{-(n\beta+1)} \exp\left\{ -(y/\alpha)^\beta \right\} d\alpha.$$

Use a changing variable as $(y/\alpha)^\beta = u$ then

$$I(\beta) = \left(y^{-n\beta}/\beta \right) \int_0^\infty u^{n-1} e^{-u} du = \Gamma(n) y^{-n\beta}/\beta$$

Hence

$$I = \int_0^\infty \beta^{n-\gamma} \left(\prod_{i=1}^n t_i \right)^\beta I(\beta) d\beta \tag{B.5}$$

$$= \Gamma(n) \int_0^\infty \beta^{n-\gamma-1} \exp\left\{ -\beta \sum_{i=1}^n \log(y/t_i) \right\} d\beta \tag{B.6}$$

$$= \frac{\Gamma(n)\Gamma(n-\gamma)}{\left(\sum_{i=1}^n \log(y/t_i) \right)^{n-\gamma}}. \tag{B.7}$$

This gives us

$$c_\gamma(\underline{t}) = I^{-1} = \frac{\left(\sum_{i=1}^n \log(y/t_i) \right)^{n-\gamma}}{\Gamma(n)\Gamma(n-\gamma)}.$$

The marginal posterior density for β can be obtained by integrating out α from the joint posterior density of (α, β)

$$\begin{aligned} p(\beta \mid \underline{t}) &= \int_0^{+\infty} p(\alpha, \beta \mid \underline{t}_n) d\alpha \\ &\propto \int_0^{+\infty} \beta^{n-1} \alpha^{-(n\beta+1)} \prod_{i=1}^n t_i^{\beta-1} \exp\left\{ -(y/\alpha)^\beta \right\} d\alpha \\ &\propto \beta^{n-1} \prod_{i=1}^n t_i^{\beta-1} \int_0^{+\infty} \alpha^{-(n\beta+1)} \exp\left\{ -(y/\alpha)^\beta \right\} d\alpha \\ &\propto \beta^{n-1} \prod_{i=1}^n t_i^{\beta-1} \Gamma(n) y^{-n\beta}/\beta \\ &\propto \beta^{n-2} \exp\left\{ -\beta \sum_{i=1}^n \log(y/t_i) \right\} \end{aligned}$$

That means marginal posterior density of β is Gamma distribution with parameters $(n-1, l_n)$ with $l_n = \sum_{i=1}^{n} \log(y/t_i)$.
The posterior density of α is obtained by integrating out β from the joint posterior density of (α, β)

$$p(\alpha \mid \underline{t}) = \int_0^\infty p(\alpha, \beta \mid \underline{t}_n) d\beta \tag{B.8}$$

$$\propto \int_0^\infty \beta^{n-1} \alpha^{-(n\beta+1)} \prod_{i=1}^{n} t_i^{\beta-1} \exp\left\{-(y/\alpha)^\beta\right\} d\beta \tag{B.9}$$

$$\propto \alpha^{-1} \int_0^\infty \beta^{n-1} \left(\prod_{i=1}^{n} t_i/\alpha \right)^\beta \exp\left\{-(y/\alpha)^\beta\right\} d\beta \tag{B.10}$$

This integral can not be computed explicitly but by numerically approximation.

Appendix C

Distribution of the PLP shape parameter β

Proof of theoremB.0.1

Event truncation

The joint distribution of $\underline{T} = (T_1, \ldots, T_n)$ is

$$f_{\underline{T}}(\underline{t}) = \beta^n \alpha^{-n\beta} \prod_{i=1}^{n} t_i^{\beta-1} \exp\left\{-(t_n/\alpha)^\beta\right\}.$$

Let denote $U_i = \varphi_i(\underline{T}) = \log(T_n/T_i), i = 1, \ldots n-1$ and $U_n = \varphi_n(\underline{T}) = (T_n/\alpha)^\beta$. We need to find the distribution of $U = \sum_{i=1}^{n-1} U_i$.

Let us calculate the joint distribution of $\underline{U} = (U_1, \ldots, U_n)$. We have

$$f_{\underline{U}}(\underline{u}) = f_{\underline{T}}(\varphi_1^{-1}(\underline{u}), \ldots, \varphi_n^{-1}(\underline{u})) \mid det(J(\underline{u} \mid \underline{t}) \mid^{-1}.$$

where $J(\underline{u} \mid \underline{t})$ is the Jacobian matrix. The term (i,j) of the matrix is:

$$J(\underline{u} \mid \underline{t}) \mid_{i,j} = \frac{\partial \varphi_i(\underline{t})}{\partial t_j}.$$

We have:

$$J(\underline{u} \mid \underline{t}) = \begin{pmatrix} -\frac{1}{t_1} & 0 & \ldots & 0 & \frac{1}{t_n} \\ 0 & -\frac{1}{t_2} & \ldots & 0 & \frac{1}{t_n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & -\frac{1}{t_{n-1}} & \frac{1}{t_n} \\ 0 & 0 & \ldots & 0 & \frac{\beta}{\alpha^\beta} t_n^{\beta-1} \end{pmatrix}.$$

and

$$|\, det(J(\underline{u} \mid \underline{t}) \,| = \frac{\beta}{\alpha^\beta} t_n^{\beta-1} \prod_{i=1}^{n-1} t_i^{-1}.$$

Note that $t_n = \alpha u_n^{1/\beta}$ and $t_i = \alpha u_n^{1/\beta} e^{-u_i}$, $i = 1, \ldots, n-1$. Therefore

$$\begin{aligned}
f_{\underline{U}}(\underline{u}) &= f_{\underline{T}}(\underline{t}) \,|\, det(J_{\underline{u}|\underline{t}}(\underline{t}) \,|^{-1} \\
&= \beta^n \alpha^{-n\beta} \prod_{i=1}^{n-1} t_i^{\beta-1}\, t_n^{\beta-1} \exp\left\{-(t_n/\alpha)^\beta\right\} \times \frac{\alpha^\beta}{\beta} \frac{1}{t_n^{\beta-1}} \prod_{i=1}^{n-1} t_i \\
&= \beta^{n-1} \alpha^{-(n-1)\beta} \prod_{i=1}^{n-1} t_i^{\beta-1} \exp\left\{-(t_n/\alpha)^\beta\right\} \\
&= \beta^{n-1} \exp\left\{-\beta \sum_{i=1}^{n-1} u_i\right\} u_n^{n-1} \exp\{-u_n\}.
\end{aligned}$$

The joint distribution of $(U_1, \ldots, U_{n-1})$ can be found by integrating out the variable u_n.

$$\begin{aligned}
f_{U_1,\ldots,U_{n-1}}(u_1, \ldots, u_{n-1}) &= \int_0^\infty f_{\underline{U}}(u_1, \ldots, u_n) du_n \\
&= \beta^{n-1} \exp\left\{-\beta \sum_{i=1}^{n-1} u_i\right\} \int_0^\infty u_n^{n-1} \exp\{-u_n\} du_n \\
&= \Gamma(n) \beta^{n-1} \exp\left\{-\beta \sum_{i=1}^{n-1} u_i\right\} = (n-1)! \prod_{i=1}^{n-1} \beta e^{-\beta u_i}.
\end{aligned}$$

We deduce from a well-known property of order statistics that $U_1, \ldots, U_{n-1}$ are i.i.d random variables of exponential distribution with parameter β. Sum of $n-1$ i.i.d random variables of exponential distribution with parameter β has a gamma distribution with parameters $(n-1, \beta)$. Hence $\sum_{i=1}^{n-1} U_i = \sum_{i=1}^{n-1} \log(T_n/T_i)$ has a gamma distribution with parameters $(n-1, \beta)$. Recall that: if X has a gamma distribution with parameters (a, b) then λX has a gamma distribution with parameters $(a, b/\lambda)$ and a chi-square distribution with n degrees of freedom is a gamma distribution with parameters $(n/2, 1/2)$. Thus $2n\beta/\hat{\beta} = 2\beta \sum_{i=1}^{n-1} U_i$ has a chi-square distribution with $2(n-1)$ degrees of freedom.

Time truncation

The conditional joint distribution of $(T_1 = t_1, T_2 = t_2, \ldots, T_n = t_n)$, given $N(C) = n$, is

$$f^*_{\underline{C}}(\underline{t}) = n! \prod_{i=1}^{n} \left[\frac{\lambda(t_i)}{\Lambda(C)}\right] = n! C^{-n\beta} \beta^n \left(\prod_{i=1}^{n} t_i\right)^{\beta-1}.$$

Let denote $U_i = \log(C/T_i), i = 1,\dots n$ then $2n\beta/\hat{\beta} = 2\beta\sum_{i=1}^n U_i$. The Jacobian matrix is

$$J_{\underline{t}}(\underline{u}) = \begin{pmatrix} -Ce^{-u_1} & 0 & \dots & 0 & 0 \\ 0 & -Ce^{-u_2} & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & -Ce^{-u_{n-1}} & 0 \\ 0 & 0 & \dots & 0 & -Ce^{-u_n}. \end{pmatrix}$$

Hence $| det(J_{\underline{t}}(\underline{u})) |= (-1)^n C^n e^{-\sum_{i=1}^n u_i}$. Thus the joint density of $(U_1,\dots,U_n)$ is

$$f^*_{\underline{U}}(\underline{u}) = n!C^{-n\beta}\beta^n \left(\prod_{i=1}^n Ce^{-u_i}\right)^{\beta-1} C^n e^{-\sum_{i=1}^n u_i} = n!\beta^n e^{-\sum_{i=1}^n u_i}.$$

That means, conditional on $N(C) = n$, random vector $(U_n < \dots < U_1)$ is distributed as order statistics of n i.i.d random variables of exponential distribution with parameter β. Sum of n i.i.d random variables of exponential distribution with parameter β has gamma distribution with parameters (n,β). Therefore $2n\beta/\hat{\beta} = 2\beta\sum_{i=1}^n U_i \sim Gamma(2n/2, 1/2) \equiv \mathcal{X}^2(2n)$.

Appendix D

Convergence of the integrals for the exponential-law process

Let consider the integral

$$J_m(\alpha,\beta) = \int_0^\infty x^{\alpha-1} \log^m(x) \left(\frac{\log(x)}{x-1}\right)^\beta dx. \tag{D.1}$$

We will prove that the integral converges for $1 \leq \alpha \leq \beta - 1$ and for all $m \in \mathbb{N}$. Note that $J_0(\alpha,\beta) = I(\alpha,\beta)$. Since $x^{\alpha-1} \leq x^{\beta-2}$, it is sufficient to prove the following integral converges

$$H_n(\beta) = \int_0^\infty x^{\beta-2} \log^n(x) \left(\frac{\log(x)}{x-1}\right)^\beta dx. \tag{D.2}$$

Decompose $H_n(\beta)$ into three integrals $H_n(\beta) = H_1 + H_2 + H_3$ where

$$H_1 = \int_0^{1/2} x^{\beta-2} \log^n(x) \left(\frac{\log(x)}{x-1}\right)^\beta dx, \tag{D.3}$$

$$H_2 = \int_{1/2}^{2} x^{\beta-2} \log^n(x) \left(\frac{\log(x)}{x-1}\right)^\beta dx, \tag{D.4}$$

$$H_3 = \int_2^\infty x^{\beta-2} \log^n(x) \left(\frac{\log(x)}{x-1}\right)^\beta dx. \tag{D.5}$$

Let's consider H_1. By changing variable $x = 1/u$ we have

$$H_1 = \int_2^\infty \frac{\log(u)^{\beta+n}}{(u-1)^\beta} du = \int_1^\infty \frac{\log(u+1)^{\beta+n}}{u^\beta} du.$$

Recall that

$$\int_1^\infty \frac{1}{x^m} dx$$

converges for all $m > 1$ and

$$\lim_{x\to+\infty} \frac{\log^a(x)}{x^b} = 0, \quad b > 0, a \in \mathbb{R}$$

Therefore

$$\int_1^\infty \frac{\log^a(x)}{x^m} dx$$

converges for all $m > 1$ and $a \in \mathbb{R}$. H_1 is finite since $\beta \geq 2$.

The integral H_2 is finite since $\lim_{x\to 1} \log(x)/(x-1) = 1$. The function $f(x) = \log^(x)/(x-1)$ is defined for all $x \in [1/2, 2]$ with $f(1) = 1$.

For the integral H_3, rewrite it as

$$H_3 = \int_2^\infty \frac{x^{\beta-2}}{(x-1)^\beta} \log(x)^{\beta+n} dx = \int_1^\infty \frac{(x+1)^{\beta-2}}{x^\beta} \log(x+1)^{\beta+n} dx$$

We have $(x+1)^{\beta-2}/x^\beta$ is equivalent to $1/x^2$ when x tends to infinity and the fact that

$$\int_1^\infty \frac{\log(x+1)^{\beta+n}}{x^2} dx$$

converges, so H_3 is finite.

Appendix E

Likelikood function for a stochastic process

Let us consider a stochastic process with intensity $\lambda^*(t \mid \theta)$, observed in a time window $[0,\tau]$. Denote $\underline{t} = (t_1, t_2, \ldots, t_n)$, the dates of jumps of the process assuming that n jumps occurs. For $i = 1, \ldots, n$,

$$P(T_i > t \mid t_{i-1}, \ldots, t_1) = P(N(t_{i-1}, t) = 0 \ ; \ \theta).$$

Taking the derivative, we can expressed the distribution of T_i, for $i = 1, \ldots, n$

$$f(t \mid t_{i-1}, \ldots, t_1) = \frac{d}{dt} P\big(N(t_{i-1}, t) = 0 \ ; \ \theta\big).$$

Therefore the likelihood is:

$$\prod_{i=1}^{n} f(t_i \mid t_{i-1}, \ldots, t_1) = \prod_{i=1}^{n} \frac{d}{dt} P\big(N(t_{i-1}, t) = 0 \ ; \ \theta\big)\Big|_{t=t_i}.$$

But

$$P\big(N(t_{i-1}, t) = 0; \theta\big) = \exp\left\{-\int_{t_{i-1}}^{t} \lambda^*(s \mid \theta) ds\right\},$$

Thus

$$\begin{aligned} \frac{d}{dt} P(N(t_{i-1}, t) = 0; \theta) &= \frac{d}{dt} \exp\left\{-\int_{t_{i-1}}^{t} \lambda^*(s \mid \theta) ds\right\} \\ &= \lambda^*(t \mid \theta) \exp\left\{-\int_{t_{i-1}}^{t} \lambda^*(s \mid \theta) ds\right\} \end{aligned}$$

Since there are no jumps between t_n et τ,

$$P\big(N(t_n, \tau) = 0; \theta\big) = \exp\left\{-\int_{t_n}^{\tau} \lambda^*(s \mid \theta) ds\right\}.$$

The likelihood is:

$$
\begin{aligned}
L(\theta) &= \prod_{i=1}^{n} \lambda(t_i \mid \theta) \exp\left\{-\int_{t_{i-1}}^{t_i} \lambda(s \mid \theta) ds\right\} \times \exp\left\{-\int_{t_n}^{\tau} \lambda(s \mid \theta) ds\right\} \\
&= \left(\prod_{i=1}^{n} \lambda(t_i \mid \theta)\right) \times \exp\left\{-\int_{0}^{\tau} \lambda(s \mid \theta) ds\right\}
\end{aligned}
$$

And the log-likelihood is:

$$\log L(\theta) = \sum_{i=1}^{n} \log \lambda(t_i \mid \theta) - \int_0^{\tau} \lambda(s \mid \theta) ds$$

or

$$\log L(\theta) = \int_0^{\tau} \log \lambda(s \mid \theta) dN(s) - \int_0^{\tau} \lambda(s \mid \theta) ds$$

Appendix F

Likelihood for the Hawkes process

The intensity of a Hawkes process with an exponential decay response function assuming n jumps $t_1, t_2, \ldots, t_n$ occurring in a time window $[0,T]$ is:

$$
\begin{aligned}
\lambda(t \mid \alpha, \beta) &= \mu + \int_{-\infty}^{t} g(t-s \mid \alpha,\beta) dN(s) \\
&= \mu + \sum_{t_i < t} \alpha e^{-\beta(t-t_i)}
\end{aligned}
$$

We have:

- $\lambda(t \mid \alpha,\beta) = \mu$, for all $t \in [0, t_1[$,
- $\lambda(t \mid \alpha,\beta) = \mu + \alpha \sum_{j=1}^{i-1} e^{-\beta(t-t_j)}$ for all $t \in [t_{i-1}, t_i[$, $i = 2, \ldots, n$,
- $\lambda(t \mid \alpha,\beta) = \mu + \alpha \sum_{j=1}^{n} e^{-\beta(t-t_j)}$ for all $t \in [t_n, T]$.

The general expression of the log-likelihood is, setting $\theta = (\alpha, \beta)$:

$$
\log L(t_1, \ldots, t_n \mid \theta) = -\int_0^T \lambda(t \mid \theta) dt + \int_0^T \log \lambda(t \mid \theta) dN(t)
$$

▶

$$
\int_0^T \lambda(t \mid \theta) dt = \int_0^{t_1} \lambda(t \mid \theta) dt + \sum_{i=2}^{n} \int_{t_{i-1}}^{t_i} \lambda(s \mid \theta) ds + \int_{t_n}^{T} \lambda(t \mid \theta) dt.
$$

- $\int_0^{t_1} \Lambda(t \mid \theta) dt = \mu t_1$

- $$\begin{aligned}\sum_{i=2}^{n}\int_{t_{i-1}}^{t_i}\lambda(s\mid\theta)ds &= \sum_{i=2}^{n}\int_{t_{i-1}}^{t_i}\mu ds+\alpha\sum_{i=2}^{n}\sum_{j=1}^{i-1}\int_{t_{i-1}}^{t_i}e^{-\beta(s-t_j)}ds\\ &= \mu\sum_{i=2}^{n}(t_i-t_{i-1})+\frac{\alpha}{\beta}\sum_{i=2}^{n}\sum_{j=1}^{i-1}\left[e^{-\beta(s-t_j)}\right]_{t_{i-1}}^{t_i}\end{aligned}$$

$$\begin{aligned}\sum_{i=2}^{n}\sum_{j=1}^{i-1}\left[e^{-\beta(s-t_j)}\right]_{t_{i-1}}^{t_i} &= \sum_{i=2}^{n}\sum_{j=1}^{i-1}\left[e^{-\beta(t_i-t_j)}-e^{-\beta(t_{i-1}-t_j)}\right]\\ &= \sum_{i=2}^{n}\sum_{j=1}^{i-1}e^{-\beta(t_i-t_j)}-\sum_{i=2}^{n}\sum_{j=1}^{i-1}e^{-\beta(t_{i-1}-t_j)}\end{aligned}$$

Let set $U_i=\sum_{j=1}^{i-1}e^{-\beta(t_i-t_j)}$, for $i=2,\dots,n$.
We have: $U_i=U_{i-1}+1$ with $U_1=0$, $i=2,\dots,n$.

$$\begin{aligned}\sum_{i=2}^{n}\sum_{j=1}^{i-1}e^{-\beta(t_i-t_j)}-\sum_{i=2}^{n}\sum_{j=1}^{i-1}e^{-\beta(t_{i-1}-t_j)} &= \sum_{i=2}^{n}U_i-\sum_{i=2}^{n}(U_{i-1}+1)\\ &= U_n-(n-1)\\ &= \sum_{j=1}^{n-1}e^{-\beta(t_n-t_j)}+n-1\\ &= \sum_{j=1}^{n-1}\left[e^{-\beta(t_n-t_j)}-1\right]\end{aligned}$$

And,

$$\sum_{i=2}^{n}\int_{t_{i-1}}^{t_i}\lambda(s\mid\theta)ds=\mu(t_n-t_1)+\frac{\alpha}{\beta}\sum_{j=1}^{n-1}\left[e^{-\beta(t_n-t_j)}-1\right].$$

- $$\begin{aligned}\int_{t_n}^{T}\lambda(t\mid\theta)dt &= \mu(T-t_n)+\alpha\sum_{j=1}^{n-1}\int_{t_n}^{T}e^{-\beta(t-t_j)}dt\\ &= \mu(T-t_n)+\frac{\alpha}{\beta}\sum_{j=1}^{n-1}\left[e^{-\beta(T-t_j)}-e^{-\beta(t_n-t_j)}\right].\end{aligned}$$

Then
$$\int_0^T\lambda(t\mid\theta)dt=\mu T+\frac{\alpha}{\beta}\sum_{j=1}^{n-1}\left[e^{-\beta(T-t_j)}-1\right].$$

►

$$\int_0^T\log\lambda(t\mid\theta)dN(t)=\sum_{i=1}^{n}\log\left[\mu+\sum_{j=1}^{i-1}\alpha e^{-\beta(t_i-t_j)}\right]+\log\left(\mu+\sum_{j=1}^{n}\alpha e^{-\beta(T-t_j)}\right)$$

The full expression for the likelihood is then:

$$
\begin{aligned}
\log L(\underline{t} \mid \alpha, \beta) &= -\mu T - \frac{\alpha}{\beta} \sum_{j=1}^{n-1} \left[e^{-\beta(T-t_j)} - 1 \right] + \sum_{i=1}^{n} \log\Big(\mu + \sum_{j=1}^{i-1} \alpha e^{-\beta(t_i - t_j)} \Big) \\
&+ \log\Big(\mu + \sum_{j=1}^{n} \alpha e^{-\beta(T-t_j)} \Big)
\end{aligned}
$$

Appendix G

Likelihood for the PLP-SEPP

Taking logarithm of this function gives us

$$\ell(\mu,\alpha,\beta,\eta) = -\mu t_n - \alpha \sum_{i=1}^{n} \left(\frac{z_i}{z_0}\right)^{\eta} \frac{1}{\beta}\left[1-e^{-\beta(t_n-t_i)}\right] + \sum_{i=1}^{n} \log\left[\mu+\alpha A(i)\right]. \quad (G.1)$$

The gradients are

$$\frac{\partial}{\partial\mu}\ell(\mu,\alpha,\beta,\eta) = -t_n + \sum_{i=1}^{n} \frac{1}{\mu+\alpha A(i)}, \quad (G.2)$$

$$\frac{\partial}{\partial\alpha}\ell(\mu,\alpha,\beta,\eta) = -\sum_{i=1}^{n} \left(\frac{z_i}{z_0}\right)^{\eta} \frac{1}{\beta}\left[1-e^{-\beta(t_n-t_i)}\right] + \sum_{i=1}^{n} \frac{A(i)}{\mu+\alpha A(i)}. \quad (G.3)$$

$$\begin{aligned}\frac{\partial}{\partial\beta}\ell(\mu,\alpha,\beta,\eta) = &-\alpha \sum_{i=1}^{n} \left(\frac{z_i}{z_0}\right)^{\eta} \frac{1}{\beta}(t_n-t_i)e^{-\beta(t_n-t_i)} - \alpha \sum_{i=1}^{n} \frac{1}{\beta^2}\left[1-e^{-\beta(t_n-t_i)}\right] \\ &-\alpha \sum_{i=1}^{n} \frac{B(i)}{\mu+\alpha A(i)}, \quad (G.4)\end{aligned}$$

where $B(1)=0$ and for $k \geq 2$

$$B(k) = \sum_{i=1}^{k-1} \left(\frac{z_i}{z_0}\right)^{\eta} (t_k-t_i)e^{-\beta(t_k-t_i)}. \quad (G.5)$$

$$\frac{\partial}{\partial\eta}\ell(\mu,\alpha,\beta,\eta) = -\alpha \sum_{i=1}^{n} \left(\frac{z_i}{z_0}\right)^{\eta} \log\left(\frac{z_i}{z_0}\right) \frac{1}{\beta}\left[1-e^{-\beta(t_n-t_i)}\right] + \alpha \sum_{i=1}^{n} \frac{D(i)}{\mu+\alpha A(i)}, \quad (G.6)$$

where $D(1)=0$ and for $k \geq 2$

$$D(k) = \sum_{i=1}^{k-1} \left(\frac{z_i}{z_0}\right)^{\eta} \log\left(\frac{z_i}{z_0}\right) e^{-\beta(t_k-t_i)}. \quad (G.7)$$

The Hessian matrix can also be computed explicitly

$$\frac{\partial^2}{\partial \mu^2}\ell(\mu,\alpha,\beta,\eta) = -\sum_{i=1}^{n} \frac{1}{[\mu+\alpha A(i)]^2}, \tag{G.8}$$

$$\frac{\partial^2}{\partial \alpha^2}\ell(\mu,\alpha,\beta,\eta) = -\sum_{i=1}^{n} \left[\frac{A(i)}{\mu+\alpha A(i)}\right]^2, \tag{G.9}$$

$$\begin{aligned}\frac{\partial^2}{\partial \beta^2}\ell(\mu,\alpha,\beta,\eta) =& \alpha\sum_{i=1}^{n}\left(\frac{z_i}{z_0}\right)^{\eta}\frac{1}{\beta}(t_n-t_i)^2 e^{-\beta(t_n-t_i)} \\ &+2\alpha\sum_{i=1}^{n}\left(\frac{z_i}{z_0}\right)^{\eta}\frac{1}{\beta^2}(t_n-t_i)e^{-\beta(t_n-t_i)} \\ &-2\alpha\sum_{i=1}^{n}\left(\frac{z_i}{z_0}\right)^{\eta}\frac{1}{\beta^3}\left[1-e^{-\beta(t_n-t_i)}\right] \\ &+\alpha\sum_{i=1}^{n}\frac{C(i)}{\mu+\alpha A(i)}-\alpha^2\sum_{i=1}^{n}\left[\frac{B(i)}{\mu+\alpha A(i)}\right]^2,\end{aligned} \tag{G.10}$$

where $C(1) = 0$ and for $k \geq 2$

$$C(k) = \sum_{i=1}^{k-1}\left(\frac{z_i}{z_0}\right)^{\eta}(t_k-t_i)^2 e^{-\beta(t_k-t_i)}. \tag{G.11}$$

$$\begin{aligned}\frac{\partial^2}{\partial \eta^2}\ell(\mu,\alpha,\beta,\eta) =& -\alpha\sum_{i=1}^{n}\left(\frac{z_i}{z_0}\right)^{\eta}\left[\log\left(\frac{z_i}{z_0}\right)\right]^2\frac{1}{\beta}\left[1-e^{-\beta(t_n-t_i)}\right] \\ &+\alpha\sum_{i=1}^{n}\frac{E(i)}{\mu+\alpha A(i)}+\alpha^2\sum_{i=1}^{n}\frac{B(i)C(i)}{[\mu+\alpha A(i)]^2},\end{aligned} \tag{G.12}$$

where $E(1) = 0$ and for $k \geq 2$

$$E(k) = \sum_{i=1}^{k-1}\left(\frac{z_i}{z_0}\right)^{\eta}\left[\log\left(\frac{z_i}{z_0}\right)\right]^2 e^{-\beta(t_k-t_i)}. \tag{G.13}$$

$$\frac{\partial^2}{\partial \mu \partial \alpha}\ell(\mu,\alpha,\beta,\eta) = -\sum_{i=1}^{n}\frac{A(i)}{[\mu+\alpha A(i)]^2}, \tag{G.14}$$

$$\frac{\partial^2}{\partial \mu \partial \beta}\ell(\mu,\alpha,\beta,\eta) = \alpha\sum_{i=1}^{n}\frac{B(i)}{[\mu+\alpha A(i)]^2}, \tag{G.15}$$

$$\frac{\partial^2}{\partial \mu \partial \eta}\ell(\mu,\alpha,\beta,\eta) = \alpha\sum_{i=1}^{n}\frac{D(i)}{[\mu+\alpha A(i)]^2}. \tag{G.16}$$

$$\begin{aligned}\frac{\partial^2}{\partial\alpha\partial\beta}\ell(\mu,\alpha,\beta,\eta) = &- \sum_{i=1}^{n}\left(\frac{z_i}{z_0}\right)^{\eta}\frac{1}{\beta}(t_n - t_i)e^{-\beta(t_n-t_i)}\\ &+ \sum_{i=1}^{n}\left(\frac{z_i}{z_0}\right)^{\eta}\frac{1}{\beta^2}\left[1 - e^{-\beta(t_n-t_i)}\right]\\ &- \sum_{i=1}^{n}\frac{B(i)}{\mu+\alpha A(i)} + \sum_{i=1}^{n}\frac{A(i)B(i)}{\left[\mu+\alpha A(i)\right]^2},\end{aligned} \tag{G.17}$$

$$\begin{aligned}\frac{\partial^2}{\partial\alpha\partial\eta}\ell(\mu,\alpha,\beta,\eta) = &- \sum_{i=1}^{n}\left(\frac{z_i}{z_0}\right)^{\eta}\log\left(\frac{z_i}{z_0}\right)\left[1 - e^{-\beta(t_n-t_i)}\right]\\ &\sum_{i=1}^{n}\frac{D(i)}{\mu+\alpha A(i)} - \alpha\sum_{i=1}^{n}\frac{A(i)D(i)}{\left[\mu+\alpha A(i)\right]^2},\end{aligned} \tag{G.18}$$

$$\begin{aligned}\frac{\partial^2}{\partial\beta\partial\eta}\ell(\mu,\alpha,\beta,\eta) = &\,\alpha\sum_{i=1}^{n}\left(\frac{z_i}{z_0}\right)^{\eta}\log\left(\frac{z_i}{z_0}\right)\frac{1}{\beta}(t_n - t_i)^2 e^{-\beta(t_n-t_i)}\\ &+ 2\alpha\sum_{i=1}^{n}\left(\frac{z_i}{z_0}\right)^{\eta}\log\left(\frac{z_i}{z_0}\right)\frac{1}{\beta^2}(t_n - t_i)e^{-\beta(t_n-t_i)}\\ &- 2\alpha\sum_{i=1}^{n}\left(\frac{z_i}{z_0}\right)^{\eta}\log\left(\frac{z_i}{z_0}\right)\frac{1}{\beta^3}\left[1 - e^{-\beta(t_n-t_i)}\right]\\ &- \alpha\sum_{i=1}^{n}\frac{F(i)}{\mu+\alpha A(i)} + \alpha^2\sum_{i=1}^{n}\frac{B(i)D(i)}{\left[\mu+\alpha A(i)\right]^2},\end{aligned} \tag{G.19}$$

where $F(1) = 0$ and for $k \geq 2$

$$F(k) = \sum_{i=1}^{k-1}\left(\frac{z_i}{z_0}\right)^{\eta}\log\left(\frac{z_i}{z_0}\right)(t_k - t_i)e^{-\beta(t_k-t_i)}. \tag{G.20}$$

Appendix H

A reduced Newton-Raphson algorithm for the Hawkes process

We compute:

$$\overline{S}'_n(\beta) = -\frac{1}{\beta}\Big[\overline{S}_n(\beta) - \frac{1}{\beta y}A'_n$$

Denote $\overline{S}_n(\beta)'' = \frac{\partial^2 \overline{S}_n(\beta)}{\partial\beta^2}$ then

$$\overline{S}_n(\beta)'' = \frac{2}{\beta^2}\overline{S}_n(\beta) + \frac{2}{\beta^2 y}A'_n - \frac{1}{\beta y}A''_n \tag{H.1}$$

The gradients are

$$\frac{\partial}{\partial\alpha}f(\alpha,\beta) = \sum_{i=1}^{n}\frac{A_i - \overline{S}_n(\beta)}{\bar{\mu} + \alpha(A_i - \overline{S}_n(\beta))}, \tag{H.2}$$

$$\frac{\partial}{\partial\beta}f(\alpha,\beta) = \alpha\sum_{i=1}^{n}\frac{A'_i - \overline{S}_n(\beta)'}{\bar{\mu} + \alpha(A_i - \overline{S}_n(\beta))}. \tag{H.3}$$

The Hessian matrix is

$$\frac{\partial^2}{\partial\alpha^2}f(\alpha,\beta) = -\sum_{i=1}^{n}\left[\frac{A_i - \overline{S}_n(\beta)}{\bar{\mu} + \alpha(A_i - \overline{S}_n(\beta))}\right]^2, \tag{H.4}$$

$$\frac{\partial^2}{\partial\beta^2}f(\alpha,\beta) = \alpha\sum_{i=1}^{n}\left[\frac{A''_i - \overline{S}_n(\beta)''}{\bar{\mu} + \alpha(A_i - \overline{S}_n(\beta))}\right] - \alpha^2\sum_{i=1}^{n}\left[\frac{A'_i - \overline{S}_n(\beta)'}{\bar{\mu} + \alpha(A_i - S_n)}\right]^2, \tag{H.5}$$

$$\frac{\partial^2}{\partial\alpha\partial\beta}f(\alpha,\beta) = \sum_{i=1}^{n}\left[\frac{A'_i - \overline{S}_n(\beta)'}{\bar{\mu} + \alpha(A_i - \overline{S}_n(\beta))}\right] - \alpha\sum_{i=1}^{n}\frac{(A_i - \overline{S}_n(\beta))(A'_i - \overline{S}_n(\beta)')}{\left[\bar{\mu} + \alpha(A_i - \overline{S}_n(\beta))\right]^2}. \tag{H.6}$$

(H.7)

Appendix I

Simulation Agorithms for Some Stochastic Processes

I.1 Simulation Algorithms for the Homogeneous Poisson Process

Denote HPP(μ) the homogeneous Poisson process with intensity $\lambda(t) = \mu$ and the compensator $\Lambda(t) = \mu t$.The simulation equation is

$$\log(U) + \mu(u - t_k) = 0. \tag{I.1}$$

This equation can be easily solved giving us

$$u = t_k - \frac{\log(U)}{\mu}.$$

Event truncation

Simulation algorithm for HPP(μ) for a fix number of jump n is

(1) Generate a uniform random variable $U \sim \text{Unif}(0,1)$.

(2) Let $t_1 = -\log(U)/\mu$.

(3) Generate a uniform variable $U \sim \text{Unif}(0,1)$.

(4) Solve the simulation equation (I.1) to get the solution

$$u = t_k - \log(U)/\mu$$

.

(5) Let $t_{k+1} = u$. If $k+1 = n$ stop, otherwise $k := k+1$ and return to step (3).

Time truncation

Simulation algorithm for HPP(μ) in the time window $(0,C)$ is

(1) Generate a uniform random variable $U \sim \text{Unif}(0,1)$.

(2) Let $t_1 = -\log(U)/\mu$.

(3) Generate a uniform variable $U \sim \text{Unif}(0,1)$.

(4) Solve the simulation equation (I.1) to get the solution

$$u = t_k - \log(U)/\mu$$

.

(5) While $u < C$ let $t_{k+1} = u$ and return to step (3).

Remark 1. *Recall that, interevents of the process HPP(μ) are independent identical distributed as exponential law with parameter μ so one can also simulate the process as following*

(1) Generate a uniform random variable $U \sim Unif(0,1)$.

(2) Let $t_1 = -\log(U)/\mu$.

(3) Generate a uniform variable $U \sim Unif(0,1)$.

(4) Solve the simulation equation (I.1) to get the solution

$$u = t_k - \log(U)/\mu$$

.

(5) Let $t_{k+1} = u$. If $k+1 = n$ stop, otherwise $k := k+1$ and return to step (3).

I.2 Simulation Algorithms for the Power-Law Process

Denote PLP(α,β) the power-law process with intensity $\lambda(t) = \left(\beta/\alpha\right)(t/\alpha)^{\beta-1}$ and the compensator $\Lambda(t) = (t/\alpha)^{\beta}$. The simulation equation is

$$\log(U) + \left[\left(\frac{u}{\alpha}\right)^{\beta} - \left(\frac{t_k}{\alpha}\right)^{\beta}\right] = 0. \tag{I.2}$$

This equation has an unique solution

$$u = \left[t_k^{\beta} - \alpha^{\beta}\log(U)\right]^{1/\beta}.$$

Event truncation

Simulation algorithm for PLP(α, β) for a fix number of jump n is

(1) Generate a uniform random variable $U \sim \text{Unif}(0,1)$.

(2) Let $t_1 = \alpha[-\log(U)]^{1/\beta}$.

(3) Generate a uniform variable $U \sim \text{Unif}(0,1)$.

(4) Solve the simulation equation (I.1) to get the solution

$$u = \left[t_k^\beta - \alpha^\beta \log(U)\right]^{1/\beta}$$

.

(5) Let $t_{k+1} = u$. If $k+1 = n$ stop, otherwise $k := k+1$ and return to step (3).

Time truncation

Simulation algorithm for PLP(α, β) in the time window $(0, C)$ is

(1) Generate a uniform random variable $U \sim \text{Unif}(0,1)$.

(2) Let $t_1 = \alpha[-\log(U)]^{1/\beta}$.

(3) Generate a uniform variable $U \sim \text{Unif}(0,1)$.

(4) Solve the simulation equation (I.1) to get the solution

$$u = \left[t_k^\beta - \alpha^\beta \log(U)\right]^{1/\beta}$$

.

(5) While $u < C$ let $t_{k+1} = u$ and return to step (3).

Remark 2. *From the equation (I.2), if we use reparametrization with $\mu = \alpha^{-\beta}$ then*

$$u_{PLP} = \left[t_k^\beta - \log(U)/\mu\right]^{1/\beta}.$$

Recall that

$$u_{HPP} = t_k - \log(U)/\mu.$$

One can see that when $\beta = 1$ the two processes are identical, that is PLP($\mu, \beta = 1$) $\equiv$ HPP(μ). That means, the PLP degenerates to the HPP when the growth rate β equals to 1.

I.3 Simulation Algorithms for the Exponential-Law Process

Denote ELP(α,β) the exponential-law process with intensity $\lambda(t)=\alpha e^{\beta t}$ and the compensator $\Lambda(t)=\alpha\left(e^{\beta t}-1\right)/\beta$. The simulation equation is

$$\log(U)+\alpha\frac{e^{\beta u}-e^{\beta t_k}}{\beta}=0. \tag{I.3}$$

When $\beta>0$ this equation has a unique solution

$$u=\frac{1}{\beta}\log\left[e^{\beta t_k}-\frac{\beta}{\alpha}\log(U)\right].$$

When $\beta<0$ this equation requires a condition $U>\exp\left\{\alpha e^{\beta t_k}/\beta\right\}$ then the unique solution is

$$u=\frac{1}{\beta}\log\left[e^{\beta t_k}-\frac{\beta}{\alpha}\log(U)\right].$$

Event truncation

Simulation algorithm for ELP(α,β) for a fix number of jump n is

(1) If $\beta>0$ then generate a uniform random variable $U\sim\text{Unif}(0,1)$.
If $\beta<0$ then generate a uniform random variable $U\sim\text{Unif}(e^{\alpha/\beta},1)$.

(2) Let $t_1=\log\left[1-\beta\log(U)/\alpha\right]/\beta$.

(3) If $\beta>0$ then generate a uniform random variable $U\sim\text{Unif}(0,1)$.
If $\beta<0$ then generate a uniform random variable

$$U\sim\text{Unif}(\exp\left\{\alpha e^{\beta t_k}/\beta\right\},1)$$

.

(4) Solve the simulation equation (I.3) to get the solution

$$u=\log\left[e^{\beta t_k}-\beta\log(U)/\alpha\right]/\beta$$

.

(5) Let $t_{k+1}=u$. If $k+1=n$ stop, otherwise $k:=k+1$ and return to step (3).

Time truncation

Simulation algorithm for ELP(α,β) in the time window $(0,C)$ is

(1) If $\beta>0$ then generate a uniform random variable $U\sim\text{Unif}(0,1)$.
If $\beta<0$ then generate a uniform random variable $U\sim\text{Unif}(e^{\alpha/\beta},1)$.

(2) Let $t_1 = \log\left[1-\beta\log(U)/\alpha\right]/\beta$.

(3) If $\beta > 0$ then generate a uniform random variable $U \sim \text{Unif}(0,1)$.
If $\beta < 0$ then generate a uniform random variable

$$U \sim \text{Unif}(\exp\left\{\alpha e^{\beta t_k}/\beta\right\}, 1)$$

.

(4) Solve the simulation equation (I.3) to get the solution

$$u = \log\left[e^{\beta t_k} - \beta\log(U)/\alpha\right]/\beta$$

.

(5) While $u < C$ let $t_{k+1} = u$ and return to step (3).

Remark 3. *From the equation (I.3)*

$$\log(U) + \alpha\frac{e^{\beta u} - e^{\beta t_k}}{\beta} = 0.$$

When $\beta \to 0$ then $\left(e^{\beta u} - e^{\beta t_k}\right)/\beta \to (u - t_k)$ so that equation becomes

$$\log(U) + \alpha(u - t_k) = 0$$

and the solution is then

$$u_{ELP} = t_k - \log(U)/\alpha.$$

Recall that for HPP(μ)

$$u_{HPP} = t_k - \log(U)/\mu.$$

One can see that when $\beta = 0$ the two processes are identical, that is ELP($\alpha, \beta = 1$) $\equiv$ HPP(α). That means, the ELP degenerates to the HPP when the growth rate β equals to 0.

I.4 Simulation Algorithms for the Hawkes Process

Denote HaP(μ, α, β) the Hawkes process with conditional intensity of the form

$$\lambda^*(t) = \mu + \sum_{t_i < t} \alpha e^{-\beta(t - t_i)}$$

The compensator is then

$$\Lambda^*(t_k, u) = \mu(u - t_k) + \frac{\alpha}{\beta}\left[1 - e^{-\beta(u - t_k)}\right] S_k.$$

where

$$S(1) = 1$$
$$S(k) = 1 + S(k-1)e^{-\beta(t_k - t_{k-1})}, k > 1.$$

The simulation equation is

$$\log(U) + \mu(u - t_k) + \frac{\alpha}{\beta}\left[1 - e^{-\beta(u-t_k)}\right] S_k. \tag{I.4}$$

This equation can not be solved explicitly but by Newton-Raphson algorithm as following

$$u_{i+1} = u_i - \frac{f(u_i)}{f'(u_i)},$$

where

$$f(u) = \log(U) + \mu(u - t_k) + \frac{\alpha}{\beta}\left[1 - e^{-\beta(u-t_k)}\right] S_k = 0,$$
$$f'(u) = \mu + \alpha S_k e^{-\beta(u-t_k)}.$$

An initial value can be chosen as $u_0 = t_k - \log(U)/\mu$.

Event truncation

Simulation algorithm for HaP(μ, α, β) for a fix number of jump n is

(1) Generate a uniform random variable $U \sim$ Unif$(0,1)$.

(2) Let $t_1 = -\log(U)/\mu$.

(3) Generate a uniform variable $U \sim$ Unif$(0,1)$.

(4) Solve the simulation equation (I.4) with respect to u.

(5) Let $t_{k+1} = u$ and compute

$$S_{k+1} = e^{-\beta(t_{k+1} - t_k} S_k + 1$$

where $S_1 = 0$.
If $k+1 = n$ stop, otherwise $k := k+1$ and return to step (3).

Time truncation

Simulation algorithm for HaP(μ, α, β) in the time window $(0, C)$ is

(1) Generate a uniform random variable $U \sim$ Unif$(0,1)$.

(2) Let $t_1 = -\log(U)/\mu$.

(3) Generate a uniform variable $U \sim$ Unif$(0,1)$.

(4) Solve the simulation equation (I.4) with respect to u.

(5) While $u < C$ let $t_{k+1} = u$ and compute

$$S_{k+1} = e^{-\beta(t_{k+1}-t_k}S_k + 1$$

where $S_1 = 0$.
Then return to step (3).

I.5 Simulation Algorithms for the Power-law Covariate Self-Exciting Point Process

Denote PLC-SEPP(μ, α, β, η) the power-law covariate self-exciting point process with conditional intensity of the form

$$\lambda^*(t) = \mu + \sum_{t_i<t} \left(\frac{z_i}{z_0}\right)^\eta \alpha e^{-\beta(t-t_i)}$$

The compensator is

$$\Lambda^*(t_k, u) = \mu(u - t_k) + \frac{\alpha}{\beta}\left[1 - e^{-\beta(u-t_k)}\right] S_k.$$

where

$$S(1) = (z_1/z_0)^\eta$$
$$S(k) = (z_1/z_0)^\eta + S(k-1)e^{-\beta(t_k-t_{k-1})}, k > 1.$$

The simulation equation is

$$\log(U) + \mu(u - t_k) + \frac{\alpha}{\beta}\left[1 - e^{-\beta(u-t_k)}\right] S_k. \tag{I.5}$$

This equation can not be solved explicitly but by Newton-Raphson algorithm as following

$$u_{i+1} = u_i - \frac{f(u_i)}{f'(u_i)},$$

where

$$f(u) = \log(U) + \mu(u - t_k) + \frac{\alpha}{\beta}\left[1 - e^{-\beta(u-t_k)}\right] S_k = 0,$$
$$f'(u) = \mu + \alpha S_k e^{-\beta(u-t_k)}.$$

An initial value can be chosen as $u_0 = t_k - \log(U)/\mu$.

Event truncation

Simulation algorithm for PLC-SEPP(μ, α, β, η) for a fix number of jump n is

(1) Generate a uniform random variable $U \sim$ Unif(0, 1).

(2) Let $t_1 = -\log(U)/\mu$.

(3) Generate a log-normal random variable $z_1 \sim$ LN(m, τ^2).

(4) Generate a uniform variable $U \sim$ Unif(0, 1).

(5) Solve the simulation equation (I.5) with respect to u.

(6) Let $t_{k+1} = u$ and compute

$$S_{k+1} = e^{-\beta(t_{k+1}-t_k} S_k + (z_1/z_0)^\eta$$

where $S_1 = 0$.
If $k+1 = n$ stop, otherwise $k := k+1$ and return to step (3).

Time truncation

Simulation algorithm for PLC-SEPP(μ, α, β, η) in the time window $(0, C)$ is

(1) Generate a uniform random variable $U \sim$ Unif(0, 1).

(2) Let $t_1 = -\log(U)/\mu$.

(3) Generate a log-normal random variable $z_1 \sim$ LN(m, τ^2).

(4) Generate a uniform variable $U \sim$ Unif(0, 1).

(5) Solve the simulation equation (I.5) with respect to u.

(6) While $u < C$ let $t_{k+1} = u$ and compute

$$S_{k+1} = e^{-\beta(t_{k+1}-t_k} S_k + 1$$

where $S_1 = 0$.
Then return to step (3).

Appendix J

Event Truncation and Time Truncation

There are two schemes of collecting data for a stochastic point process: one can observe the process in a fixed time window, that we call time truncation; one also can observe the process until a fixed number of jumps, that we call failure truncation.
Let $N(t), t \geq 0$ the Poisson process with intensity $\lambda(t)$ and n first jumps of the process and likelihood in each scheme.

$$\Lambda(t) = \int_0^t \lambda(u) du$$

is the compensator of the process. The following two theorems give the distribution of

Theorem J.0.1. *Suppose that the Poisson process $N(t), t \geq 0$ with intensity $\lambda(t)$ is observed until the $n-th$ jump ($n \geq 1$) and we obtain $(t_1, \ldots, t_n)$ as n first jumps of the process. Denote $\underline{T} = (T_1, \ldots, T_n)$ a random vector of n first events of the process and $\underline{t} = (t_1, \ldots, t_n)$ a realization of that random vector. The joint density of $(T_1, \ldots, T_n)$ can be computed as the following theorem.*

$$f_{\underline{T}}(\underline{t}) = \left[\prod_{i=1}^{n} \lambda(t_i)\right] \exp\{-\Lambda(t_n)\} \tag{J.1}$$

and the likelihood of a realization $(t_1, \ldots, t_n)$ is

$$L(\theta) = f_{\underline{T}}(\underline{t} \mid \theta) = \left[\prod_{i=1}^{n} \lambda(t_i)\right] \exp\{-\Lambda(t_n)\}. \tag{J.2}$$

Proof 15. *Recall that for the Poisson process, $N(0, t_1), N(t_1, t_1 + dt_1), \ldots, N(t_{n-1} + dt_{n-1}, t_n), N(t_n, t_n + dt_n)$ are independent and are distributed by the law of*

Poisson with parameter $\Lambda(0,t_1), \Lambda(t_1,t_1+dt_1), \dots, \Lambda(t_{n-1}+dt_{n-1},t_n), \Lambda(t_n,t_n+dt_n)$ *respectively. For* $n \geq 1$ *we have*

$$\begin{aligned}
&P(t_1 < T_1 \leq t_1+dt_1, \dots, t_n < T_n \leq t_n+dt_n) \\
&= P[N(0,t_1)=0, N(t_1,t_1+dt_1)=1, \dots, \\
&\quad N(t_{n-1}+dt_{n-1},t_n)=0, N(t_n,t_n+dt_n)=1] \\
&= P[N(0,t_1)=0]P[N(t_1,t_1+dt_1)=1]\dots \\
&\quad P[N(t_{n-1}+dt_{n-1},t_n)=0)P(N(t_n,t_n+dt_n)=1] \\
&= \exp\{-\Lambda(0,t_1)\}\Lambda(t_1,t_1+dt_1)\exp\{-\Lambda(t_1,t_1+dt_1)\}\dots \\
&\quad \exp\{-\Lambda(t_{n-1}+dt_{n-1},t_n)\}\Lambda(t_n,t_n+dt_n)\exp\{-\Lambda(t_n,t_n+dt_n)\} \\
&= \exp\{-\Lambda(0,t_n+dt_n)\}\prod_{i=1}^{n}[\Lambda(t_i,t_i+dt_i)] \\
&= \exp\{-\Lambda(0,t_n+dt_n)\}\prod_{i=1}^{n}[\lambda(t_i)dt_i+o(dt_i)]
\end{aligned}$$

Therefore the joint density of $(T_1, \dots, T_n)$ *is*

$$\begin{aligned}
f_{\underline{T}}(\underline{t}) &= \lim_{d\underline{t}\to 0}\frac{P(t_1 < T_1 \leq t_1+dt_1, \dots, t_n < T_n \leq t_n+dt_n)}{dt_1\dots dt_n} \\
&= \left[\prod_{i=1}^{n}\lambda(t_i)\right]\exp\{-\Lambda(t_n)\}.
\end{aligned}$$

In failure truncation scheme, the likelihood of a realization $(t_1, \dots, t_n)$ *is simply the density joint function of the first n jumps of the process, so*

$$L(\theta) = f_{\underline{T}}(\underline{t} \mid \theta) = \left[\prod_{i=1}^{n}\lambda(t_i)\right]\exp\{-\Lambda(t_n)\}.$$

□

Time truncation

Theorem J.0.2. *Suppose that the Poisson process* $N(t), t \geq 0$ *with intensity* $\lambda(t)$ *is observed in the time window* $(0,T)$ *and we obtain* $(t_1, \dots, t_n)$ *as n first jumps of the process. Denote* $\underline{T} = (T_1, \dots, T_n)$ *a random vector of n first events of the process and* $\underline{t} = (t_1, \dots, t_n)$ *a realization of that random vector. The conditional joint density of* $(T_1, \dots, T_n)$*, given* $N(T) = n$*, can be computed as the following.*

$$f^*_{\underline{T}}(\underline{t}) = n!\left[\prod_{i=1}^{n}\frac{\lambda(t_i)}{\Lambda(T)}\right]. \tag{J.3}$$

and the likelihood of a realization $(t_1, \dots, t_n)$ *is*

$$L(\theta) = f^*_{\underline{T}}(\underline{t} \mid \theta)P[N(T)=n] = \left[\prod_{i=1}^{n}\lambda(t_i)\right]\exp\{-\Lambda(T)\}. \tag{J.4}$$

Proof 16. *For $n \geq 1$ we have*

$$\begin{aligned}
&P(N(T)=n; t_1 < T_1 \leq t_1 + dt_1, \ldots, t_n < T_n \leq t_n + dt_n) \\
=&P[N(0,t_1)=0, N(t_1, t_1+dt_1)=1, \ldots, \\
&N(t_{n-1}+dt_{n-1}, t_n)=0, N(t_n, t_n+dt_n)=1] \\
=&P[N(0,t_1)=0]P[N(t_1,t_1+dt_1)=1]\ldots \\
&P[N(t_{n-1}+dt_{n-1},t_n)=0)P(N(t_n,t_n+dt_n)=1] \\
&=\exp\{-\Lambda(0,t_1)\}\,\Lambda(t_1,t_1+dt_1)\exp\{-\Lambda(t_1,t_1+dt_1)\}\ldots \\
&\exp\{-\Lambda(t_{n-1}+dt_{n-1},t_n)\}\,\Lambda(t_n,t_n+dt_n)\exp\{-\Lambda(t_n,t_n+dt_n)\} \\
&=\exp\{-\Lambda(0,t_n+dt_n)\}\prod_{i=1}^{n}[\Lambda(t_i,t_i+dt_i)] \\
&=\exp\{-\Lambda(0,T)\}\prod_{i=1}^{n}[\lambda(t_i)dt_i+o(dt_i)]
\end{aligned}$$

Recall that $N(T)$ is distributed by the law of Poisson with parameter $\Lambda(T)$ hence

$$P(N(T)=n)=\frac{[\Lambda(T)]^n}{n!}\exp\{-\Lambda(T)\}.$$

The conditional probability is

$$\begin{aligned}
&P(t_1 < T_1 \leq t_1+dt_1,\ldots,t_n < T_n \leq t_n+dt_n | N(T)=n) \\
=&\frac{n!}{[\Lambda(T)]^n}\prod_{i=1}^{n}[\lambda(t_i)dt_i+o(dt_i)]
\end{aligned}$$

Therefore the conditional joint density of $(T_1,\ldots,T_n)$, given $N(T)=n$, is

$$\begin{aligned}
f^*_{\underline{T}}(\underline{t}) =& \lim_{d\underline{t}\to 0}\frac{P(t_1 < T_1 \leq t_1+dt_1,\ldots,t_n < T_n \leq t_n+dt_n|N(T)=n)}{dt_1\ldots dt_n} \\
=&n!\left[\prod_{i=1}^{n}\frac{\lambda(t_i)}{\Lambda(T)}\right].
\end{aligned}$$

In time truncation scheme, the likelihood of a realization $(t_1,\ldots,t_n)$ is a product of the probability of having n jumps ($n \geq 1$) and the conditional joint density function of the first n jumps of the process, so

$$L(\theta)=f_{\underline{T}}(\underline{t}\,|\,\theta)=\left[\prod_{i=1}^{n}\lambda(t_i)\right]\exp\{-\Lambda(t_n)\}.$$

□

Bibliography

[1] AKAIKE, H. (1974). A new look ath the statistical model identification. *IEEE Transactions on Automatic control* **19**, 716–723.

[2] BAIN, L. J. & ENGELHARDT, M. (1980). Inference on the parameters and current system reliability for the time truncated Weibull process. *Technometrics* **22**, 421–426.

[3] BAR-LEV, S. K., LAVI, I. & REISER, B. (1992). Bayesian inference for the power law process. *Ann. Inst. Statist. Math.* **44**, 623–639.

[4] BOX, G. E. & TIAO, G. C. (1973). Bayesian inference in statistical analysis. *Addison-Wesley* .

[5] CALABRIA, R., GUIDA, M. & PULCINI, G. (1988). Some modified maximum likelihood estimators for the Weibull process. *Reliability Engineering and System Safety* **23**, 51–58.

[6] CALABRIA, R., GUIDA, M. & PULCINI, G. (1990). Bayesian estimation of prediction intervals for a power law process. *Comm. Statist. Theory Methods* **19**, 3023–3035.

[7] CHEN, F. & HALL, P. (2013). Inference for a nonstationary self-exciting point process with an application in ultra-high frequency financial data modeling. *Journal of Applied Probability* .

[8] CHORNOBOY, E., SCHRAMM, L. & KARR, A. (1988). Maximum likelihood identificatio of neuronal point process system. *Biol. Cybern.* **59**, 265–275.

[9] COWPERTWAIT, P., ISHAM, V. & ONOF, C. (2007). Point process models of rainfall: Development of pulses within cells. *Proceedings of The Royal Society A Mathematical Physical and Engineering Sciences* **463**, 2569–2587.

[10] COX, D. R. (1955). Some statistical methods connected with series of events. *J. Roy. Statist. Soc.* **17**, 129–164.

[11] COX, D. R. & ISHAM, V. (1980). *Point process*. London: Chapman & Hall.

[12] CRANE, R. & SORNETTE, D. (2018). Robust dynamic classes revealed by measuring the response function of a social system. *Archive* .

[13] CROW, L. H. (1974). Reliability analysis for complex repairable systems. *Reliability and Biometry: statistical analysis of life length, Edited by F. Proschan, SIAM Philadelpia* , 379–410.

[14] CROW, L. H. (1975). Tracking reliability growth. *Proceedings of the Twentieth Conference on the Design of Experiments, ARO Report* **75–2**, 741–754.

[15] CROW, L. H. (2004). An extended reliability growth model for managing and assessing corrective actions. *2004 PROCEEDINGS Annual RELIABILITY and MAINTAINABILITY Symposium* .

[16] DALEY, D. J. & VERE-JONES, D. (2003). *An Introduction to the Theory of Point Processes– Elementary Theory and Methods*, vol. I. Springer, New York, 2nd ed.

[17] DUANE, J. T. (1964). Learning curve approach to reliability monitoring. *IEEE Transactions on Aerospace* **2**, 563–566.

[18] ENGELHARDT, M. & BAIN, L. J. (1978). Prediction intervals for the Weibull process. *Technometrics* **20**, 167–169.

[19] FINKELSTEIN, J. M. (1976). Confidence bounds on the parametes of the Weibull process. *Technometrics* **18**, 115–117.

[20] GAUDOIN, O., YANG, B. & XIE, M. (2003). A simple goodness-of-fit test for the power-law process based on the duane plot. *IEEE Transactions on Reliability* **52**, 69–74.

[21] GOEL, A. L. & OKUMOTO, K. (1979). Time-dependent error-detection rate model for software reliability and other performance measures. *IEEE Transactions on Reliability* **28**, 206–211.

[22] GOUNO, E. & DAMAJ, R. (2016). Inference for a one-memory self-exciting point process. *Applied Mathematics in Engineering and Reliability, Proceeding of ICAMER, Ho Chi Minh City, Vietnam* , 3–8.

[23] GUIDA, M., CALABRIA, R. & PULCINI, G. (1989). Bayes inference for non-homogenuous Poisson process with power intensity law. *IEEE Transactions on Reliability* **38**, 603–609.

[24] HAWKES, A. (1971). Spectra of some self-exciting and mutually exciting point process. *Biometrika* **58**, 83–90.

[25] HAWKES, A. & OAKES, D. (1974). A cluster process representation of a sel-exciting process. *J. Appl. Prob* **11**, 493–503.

[26] HIGGINS, J. J. & TSOKOS, C. P. (1981). A quasi-bayes estimate of the failure intensity of a reliability-growth model. *IEEE Transactions on Reliability* **R-30**, 471–475.

[27] HUANG, Y. S. & BIER, V. M. (1998). A natural conjugate prior for the non homogeneous poisson process with a power law intensity function. *Communications in Statistics-Simulation and Computation* **27**, 525–551.

[28] HUANG, Y. S. & BIER, V. M. (1999). A natural conjugate prior for the non homogeneous poisson process with an exponential law intensity function. *Communications in Statistics-Theory and Methods* **25**, 1479–1509.

[29] KAZEMIPOUR, A., WU, M. & BABADI, B. (2016). Robust estimation of self-exciting point process models with application to neuronal modeling. *Archive* .

[30] KOOPMANA, S. J., LUCAS, A. & MONTEIRO, A. (2008). The multi-state latent factor intensity model for credit rating transitions. *Journal of Econometrics* **142**, 399–424.

[31] KUO, L. & YANG, T. (1996). Bayesian computation for nonhomogeneous poisson processes in software reliability. *Journal of the American Statistical Asociation* **434**, 763–773.

[32] KYPARISIS, J. & SINGPURWALLA, N. D. (1985). Bayesian inference for the Weibull process with applications to assessing software reliability growth and predicting software failures. *Computer Science and Statistics: Proceedings of the Sixteenth Symposium on the Interface (ed. L. Billard), Elsevier Science Publishers B. V. (North Holland), Amsterdam* , 57–64.

[33] LEE, L. & LEE, S. (1978). Some results on inference for the Weibull process. *Technometrics* **20**, 41–45.

[34] LEWIS, P. A. W. & SHEDLER, G. S. (1979). Simulation of nonhomogeneous poisson processes by thinning. *Naval Research Logistics Quarterly* **26**, 403–413.

[35] LORENZEN, F. (2012). Analysis of order clustering using high frequency data: A point process approach. *Thesis. Tilburg School of economics and Management Finance Department* .

[36] MÄKELÄINEN, T., SCHIMDT, K. & STYAN, G. (1981). On the existence and uniqueness of the maximum likelihood estimate of vector-valued parameter in fixed-size samples. *The Annals of Statistics* **9**, 758–767.

[37] MASUDA, N., TAKAGUCHI, T., SATO, N. & YANO, K. (2012). Self-exciting point process modeling of conversation event sequences. *Temporal Network. Understanding Complex Systems. In Holmes P., Saramäki J. (eds) .Springer.* , 245–264.

[38] MAUD, B. (1981). *Point processes and queues*. Berlin: Springer-Verlag.

[39] MINO, H. (2001). Parameter estimation of the intensity process of self-exciting point processes using the EM algorithm. *IEEE Transactions on Instrumentation and Measurement* **50**, 658–664.

[40] MOLHER, G. O., SHORT, M. B., BRANTINGHAM, P. J., SCHOENBERG, F. P. & TITA, G. E. (2011). Self-exciting point process modeling of crime. *Journal of the American Statistical Association* , 100–108.

[41] MUSA, J. (1979). Software reliability date. *deposited in IEEE Computer Society Repository, New York* .

[42] OGATA, Y. (1981). On lewis' simulation method for point processes. *IEEE Transactions on Information Theory* **27**, 23–31.

[43] OGATA, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Ann. Inst. Statist. Math.* **83**, 9–27.

[44] OGATA, Y. & AKAIKE, H. (1982). On linear intensity models for mixed doudly stochastic poisson and self-exciting point processes. *J. R. Statist. Soc. B* **44**, 102–107.

[45] OLIVEIRA, M. D., COLOSIMO, E. A. & GILARDONI, G. L. (2012). Bayesian inference for power law processes with applications in repairable systems. *Journal of Statistical Planning and Inference* **142**, 1151–1160.

[46] OZAKI, T. (1979). Maximum likelihood estimation of hawkes self-exiting point processes. *Journal of the American Statistical Association.* **31**, 145–155.

[47] PAPANGELOUS (1972). Integrability of expected increments of points processes and a related random change of scale. *Trans. Amer. Math. Soc.* **165**, 483–506.

[48] PERUGGIA, M. & SANTNER, T. (1996). Bayesian analysis of time evolution of earthquakes. *Journal of the American Statistical Association* **91**, 1209–1218.

[49] RATHBUN, S. L. (1996). Asymptotic properties of the maximum likelihood estimator for spatio-temporal point processes. *Journal of Statistical Planning and Inference* **51**, 55–74.

[50] RIGDON, S. E. (2002). Properties of the duane plot for repairable systems. *Qual. Reliab. Engng. Int* **18**, 1–4.

[51] RIGDON, S. E. & BASU, A. P. (1989). The power law process: a model for the reliability of repairable systems. *Journal of Quality Technology* **21**, 251–260.

[52] RUGGERI, F. & SOYER, R. (2008). Advances in bayesian software reliability modelling. *Advances in Mathematics Modelling for Reliability, Bedford et al. (Eds)* , 145–155.

[53] SCHWARZ, G. (1978). Estimating the dimension of a model. *The Annals of Statistics* **6**, 461–464.

[54] SEN, A. (2002). Bayesian estimation and prediction of the intensity of the power law process. *Journal of Statistical Computation and Simulation* **72**, 613–631.

[55] SEN, A. & KHATTREE, R. (1998). On estimating the current intensity of failure for the power law process. *Journal of Statistical Planning and Inference* **74**, 252–272.

[56] SNYDER, L. & MILLER, I. (1991). *Random Point processes in Time and Space*. Springer.

[57] SOLAND, R. M. (1968). Bayesian analysis of Weibull process with unknown scale and its application to acceptance sampling. *IEEE Transactions on Reliability* **R-18**, 84–90.

[58] SOLAND, R. M. (1969). Bayesian analysis of Weibull process with unknown scale and shape parameters. *IEEE Transactions on Reliability* **R-18**, 181–184.

[59] TEICH, M. C. & SALEH, B. E. A. (2000). Branching processes in quantum electronics. *IEEE Journal of Selected Topics in Quantum Electronics* **6**, 1450–1457.

[60] UTSU, T. (1961). A statistical study of the occurrence of aftershocks. *Geophysical Magazine* **30**, 521–605.

[61] VERE-JONES, D. (1995). Forecasting earthquakes and earthquake risk. *International journal of forecasting* **11**, 503–538.

[62] ZADEH, A. H. & SHARDA, R. (2015). *Reshaping Society through Analytics*. Springer International Publishing Switzerland.