N° d'ordre : 145/2025-D/MT

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université des Sciences et de la Technologie Houari Boumediène

Faculté de Mathématiques

THESE DE DOCTORAT EN SCIENCES

Présentée pour l'obtention du grade de **DOCTEUR**

En : MATHEMATIQUES

Spécialité : Probabilités et Statistique

Par : **Bachir Anis**

Sujet

Algorithmes Stochastiques pour l'estimation du Quantile et des paramètres de la loi de Poisson tronquée

Soutenue publiquement, le **14 Décembre 2024**, devant le jury composé de :

Mme	Sadki Ourida	Professeur	USTHB	Président
Mme	Djaballah Khedidja	Professeur	USTHB	Directrice de thèse
M	Hamaz Abdelghani	Professeur	UMMTO	Examinateur
M	Latreche Abelouahab	Professeur	ENSSEA	Examinateur

DEDICACES

À la mémoire de ma grande mère Zhira
J'aurais tant aimé que tu sois là aujourd'hui, mais le destin en a décidé autrement.
Puisse Dieu le Tout-Puissant t'accorder sa sainte miséricorde et t'accueillir dans son vaste paradis,

Anis BACHIR

REMERCIEMENTS

Ce travail a été effectué sous la direction de Madame DJABALLAH Khedidja, Professeur à l'U.S.T.H.B. Je ne saurais lui exprimer toute ma reconnaissance et ma gratitude pour sa générosité intellectuelle, ses encouragements et les précieux conseils qu'elle m'a prodigués et pour sa patience tout le long de cette thèse. Merci beaucoup Madame DJABALLAH.

Que Madame SADKI Ourida, Professeur à l'U.S.T.H.B., veuille bien trouver l'expression de ma respectueuse reconnaissance pour l'honneur qu'il me fait de présider le jury de cette thèse.

Je remercie vivement Monsieur HAMAZ Abdelghani, Professeur à l'université de Tizi-Ouzou U.M.M.T.O, pour l'honneur qu'il me fait de participer à ce jury de thèse.

Mes remerciements les plus sincères vont également à Monsieur LATRECHE Abelouahab, Professeur à E.N.S.S.E.A de Koléa, pour avoir accepté de rapporter juger ce travail.

Je ne citerai pas, mais ils se reconnaîtront, tous mes amis et collègues de l'USTHB et de l'ENSTA, qui m'ont soutenu, même de loin. Je leur en suis reconnaissant.

Enfin je voudrais remercier toute ma famille, en particulier mes parents, ma femme, mon frère, ma sœur et mes deux petites filles HANAA et SARAH, qu'aucune dédicace ne saurait exprimer mon respect et mes sentiments pour l'amour, l'attention et les sacrifices consentis.

Table des matières

Chapitre 1

Resume

1.1 Abstract

This thesis consists of two parts. The first concerns the estimation of quantile using stochastic algorithms and the second develops the parameter estimation of a a-truncated Poisson law. Both parties use stochastic algorithms.

1.1.1 Estimation of quantile using stochastic algorithms

The population mean of a variable X provides an important central measure, while the population median is an important alternative that is robust to po- tential outliers. The quantiles, a generalized concept of median, are capable of providing not only central features but also the tail properties of the response distribution. Our approach to estimating the quantile q is associated with a stochastic recursive algorithm similar to that of Robbins-Monro described in Robbins and Monro (1951) and Robbins and Siegmund (1971). Assume that one can find a function φ (called a contrast function) free of the parameter q, such that $\varphi(q) = 0$. Then, it is possible to estimate q by the Robbins-Monro algorithm

$$q_{n+1} = q_n + \gamma_n T_{n+1}$$

where (γ_n) is a positive sequence of real numbers decreasing to zero and (T_n) is a sequence of random variables such that $E\left(T_{n+1}|F_n\right) = \varphi(q_n)$, where F_n stands for the σ-algebra of the events occurring up to time n. Under standard

conditions on the function φ and on the sequence (γ_n), it is well-known (see Duflo (1997) that q_n tends to q almost surely. A lot of work has already been done around the rate of convergence of stochastic algorithms. Typically one normalises the iterates after centring them about the limit. Convergence in distribution to a random normal variable can then be proved for unconstrained algorithms (see [Duflo, 1997] for instance). The case of constrained algorithms is somehow related to the problem of multiple targets. The convergence rate of algorithms with multiple targets has been studied by Pelletier in [26].

Our aim is to provide an algorithm in order to approximate the a quantile q. To construct a stochastic algorithm, which approximates the function q, we define an algorithm of search of the zero of the function $\varphi : x \to \varphi(x) = \delta - F_n(x)$.

1.1.2 Estimation of Poisson a-truncated parameter

The basic Poisson model can be written as

$$P(Y_i = j) = \frac{e^{-\lambda} \times \lambda^j}{j!}$$

where there are $i = 1, 2, ..., n$ observations, Y_i is the i^{th} observation of the variable, $j = 0, 1, 2, ...$ are the possible values of Y_i (ie a set of natural numbers), and λ is the Poisson parameter. The zero-truncated Poisson distribution is a useful model for integer-valued random phenomena when the value of 0 is unobservable or simply out of domain. These distributions are found in many applications. Several examples have been given employing the truncated distribution in fitting rainfall data. Maximum likelihood estimation is a method of point estimation often used in statistics. Consider a sample of n independent observations $X_1, X_2, \ldots, X_n$, each with the probability density function $f(x)$.

The maximum likelihood estimator is the root of the equation

$$\frac{\lambda}{1 - e^{-\lambda}} = \overline{x}$$

where $\overline{x}$ is the empirical mean of the observations.

One part is denoted to the study of used discrete distribution with emphasis on their behavior under truncation. We works through the details of the a-truncated Poisson distribution, a special case of which is the zero-truncated Poisson distribution. The a-truncated Poisson distribution is the distribution of a Poisson random variable Y conditional on the event $Y > a$. Here, a is the cutoff value such that only values strictly larger than a are allowed, i.e. the probability mass function is

$$p_j = \frac{s_a^{-1} \lambda^j e^{-\lambda}}{j!}$$

$j = a+1, a+2, \ldots$ where $a \geq 0$ is an integer, $\lambda > 0$ is a parameter, and

$$s_a = 1 - \sum_{i=0}^{a} \frac{\lambda^i e^{-\lambda}}{i!}.$$

To illustrate the practical application of results obtained in above some simulations was conducted.

1.2 Résumé

Cette thèse se compose de deux parties. La première concerne l'estimation du quantile à l'aide des algorithms stochastiques et la seconde developpe l'estimation d'une loi de Poisson a-tronquée. Les deux parties utilisent les algorithmes stochastiques.

1.2.1 Estimation du quantiles à l'aide des algorithms stochastiques

Notre approche pour estimer le quantile q est associée à un algorithme récursif stochastique similaire à celui de Robbins-Monro décrit dans Robbins et Monro (1951) et Robbins et Siegmund (1971). Supposons que l'on puisse trouver une fonction φ (appelée fonction de contraste) libre du paramètre q, tel que $\varphi(q) = 0$. Ensuite, il est possible d'estimer q par l'algorithme de Robbins-Monro

$$q_{n+1} = q_n + \gamma_n T_{n+1}$$

où (γ_n) est une séquence positive de nombres réels diminuant à zéro et (T_n) est une séquence de variables aléatoires telles que $E[T_{n+1}|F_n] = \varphi(q_n)$, où F_n représente la σ-algèbre des événements se produisant jusqu'à n. Dans des conditions standard sur la fonction φ et sur la séquence (γ_n), il est bien connu (voir Duflo (1997) que q_n tend vers q presque sûrement. Beaucoup de travail a déjà été fait autour du taux de convergence des algorithmes stochastiques. Typiquement, on normalise les itérations après les avoir centrées sur la limite. La convergence dans la distribution vers une variable normale aléatoire peut alors être prouvée pour les algorithmes sans contrainte (voir [Duflo, 1997] par exemple). Le cas des algorithmes contraints est en quelque sorte lié au problème des cibles multiples. Le taux de convergence des algorithmes à cibles multiples a été étudié par Pelletier dans [26].

Notre objectif est de fournir un algorithme afin d'approcher le quantile q. Pour construire un algorithme stochastique, qui se rapproche de la fonction q, nous définissons un algorithme de recherche du zéro de la fonction $\varphi : x \to \varphi(x) = \delta - F_n(x)$.

1.2.2 Estimation du paramètre de la loi a-tronquée Poisson

Le modèle de Poisson de base peut être écrit

$$P(Y_i = j) = \frac{e^{-\lambda} \times \lambda^j}{j!}$$

où il y a $i = 1, 2, ..., n$ observations, Y_i est l'observation de ith de la variable, $j = 0, 1, 2, ...$ sont les valeurs possibles de Y_i (c'est-à-dire un ensemble de nombres naturels), et λ est le paramètre de Poisson. La distribution de Poisson tronquée à zéro est un modèle utile pour les phénomènes aléatoires à valeur entière lorsque la valeur de 0 est inobservable ou simplement hors domaine. Ces distributions se retrouvent dans de nombreuses applications. Plusieurs exemples ont été donnés en utilisant la distribution tronquée pour ajuster les données de pluie. L'estimation du maximum de vraisemblance est une méthode d'estimation ponctuelle souvent utilisée en statistique. Considérons un échantillon d'observations indépendantes n, $X_1, X_2, ..., X_n$, chacune avec la fonction de densité de probabilité $f(x)$.

L'estimateur du maximum de vraisemblance est la racine de l'équation

$$\frac{\lambda}{1 - e^{-\lambda}} = \overline{x}$$

où $\overline{x}$ est la moyenne empirique des observations.

Une partie est consacrée à l'étude de la distribution discrète utilisée en mettant l'accent sur leur comportement sous troncation. Nous travaillons sur les détails de la distribution a-tronquée de Poisson, dont un cas particulier est la distribution de Poisson zéro-tronquée. La distribution de Poisson a-tronquée est la distribution d'une variable aléatoire de Poisson Y conditionnelle à l'événement $Y > a$. Ici, a est la valeur seuil de sorte que seules les valeurs strictement supérieures à a sont autorisées, c.-à-d. que la fonction de masse de probabilité est

$$p_j = \frac{s_a^{-1} \lambda^j e^{-\lambda}}{j!}$$

, $j = a + 1, a + 2, \ldots$ où $a \geq 0$ est un entier, $\lambda > 0$ est un paramètre et

$$s_a = 1 - \sum_{i=0}^{a} \frac{\lambda^i e^{-\lambda}}{i!}.$$

Pour illustrer l'application pratique des résultats obtenus ci-dessus, des simulations ont été réalisées.

Chapitre 2

Introduction

2.1 Preliminaries

Our approach to estimating the quantile q is associated with a stochastic recursive algorithm similar to that of Robbins-Monro (1951) described in [31] and Robbins and Siegmund (1971). Assume that one can find a function φ (called a contrast function) free of the parameter q, such that $\varphi(q) = 0$. Then, it is possible to estimate q by the Robbins-Monro algorithm

$$q_{n+1} = q_n + \gamma_n T_{n+1} \qquad (1)$$

where (γ_n) is a positive sequence of real numbers decreasing to zero and (T_n) is a sequence of random variables such that $E(T_{n+1}|\mathcal{F}_n) = \varphi(q_n)$, where $\mathcal{F}_n$ stands for the σ-algebra of the events occurring up to time n. Under standard conditions on the function φ and on the sequence (γ_n), it is well-known see Duflo (1997) that q_n tends to q almost surely. A lot of work has already been done around the rate of convergence of stochastic algorithms. Typically one normalises the iterates after centring them about the limit. Convergence in distribution to a random normal variable can then be proved for unconstrained algorithms, see Duflo (1997) for instance. The case of constrained algorithms is somehow related to the problem of multiple targets. The convergence rate of algorithms with multiple targets has been studied by Pelletier in [26]. The first part concerns the estimation of the quantile q of the F. The second part concerns the solution of estimator of the problem addressed in this thesis is the non-parametric estimation of quantiles.

Given a random variable X with continuous distribution function F, we define the quantile as the solution to the a equation for some given value of q between 0 and 1. We assume in what follows that q is unique, i.e. that we are dealing with continuous or partly continuous distributions. Quantiles find application, for example, in testing statistical hypotheses and in characterizing the extreme values of the distribution of X when q is near or 1.

The estimation of distribution quantile is a classic topic. Estimation of the sample quantile via bootstrap was proposed by many authors. Suppose that a researcher wants to estimate the fifty percent quantile point, i.e. median. The aim is to investigate stochastic approximation and opportunities of applying stochastic approximation to statistical median point estimation for random variable with unknown distribution function.

The use of stochastic approximation algorithms in the frameworks of non parametric statistics was widely discussed. The most famous use of stochastic approximation algorithms in the frame work of nonparametric statistics is the work of Kiefer and Wolfowitz (1952), who build up an algorithm which allows the approximation of the maximizer of a regression function. Their well-known algorithm was widely discussed and extended in many directions. Stochastic approximation algorithms were also introduced by Révész (1973) to estimate a regression function and by Tsybakov (1990) to approximate the mode of a probability density. Djeddour *et al* (2008) proposed an recursive estimation of location and of size of the mode of a probability density. There is an extensive literature on so-called stochastic approximation methods we can cite Pelletier (1998), Duflo (1997) among others.

2.2 Stochastic algorithms and Quantile

2.2.1 Algorithm construction

The most famous use of stochastic approximation algorithms in the framework of nonparametric statistics is the work of Kiefer and Wolfowitz(1952), who build up an algorithm which allows the approximation of the maximizer of a regression function. Stochastic approximation algorithms were also introduced by Révész (1973) to estimate a regression function and by Tsybakov (1990) to approximate the mode of a probability density.

Let $(X_1, ..., X_n)$ are n random variables independents, taking its values in $\mathbb{R}$ and with the same density of probability f. The kernel estimator of density of probability f is defined by

$$f_n(x) = \frac{1}{nh_n} \sum_{i=1}^{n} k\left(\frac{x - X_i}{h_n}\right) \tag{2.2.1}$$

where the bandwidth (h_n) is a sequence of reals numbers strictly positives who goes to zero, and the kernel K is a function continuous on $\mathbb{R}$, with values in $\mathbb{R}$.

A distribution function is estimated by integrating a kernel estimator of the density say F_n. The kernel estimator of the quantile q of F is a random variable q_n satisfying

$$F_n(q_n) = \alpha \qquad (2)$$

. We apply the stochastic approximation method to construct a recursive kernel estimators of a quantile of a distribution function. The use of stochastic approximation algorithms in the frameworks of non parametric statistics was widely discussed. In particular, Mokkadem and Pelletier [26] provided a companion algorithm to Kiefer-Wolfowitz's algorithm in order to simultaneously approximate the location and the size of the mode of the regression function. Some preliminary results on stochastic approximation algorithm.

Kiefer and Wolfowitz (1952) and Blum (1954) (in the case d = 1), have introduced an algorithm, which allows to recursively approximate θ. We consider the stochastic approximation algorithm :

$$Z_{n+1} = Z_n + \gamma_n \left[h(Z_n) + r_{n+1}\right] + \sigma_n \varepsilon_{n+1} \tag{2.2.2}$$

where the random variable Z_0, $(r_n)_{n\geq 1}$ and $(\varepsilon_n)_{n\geq 1}$ are defined on $(\Omega, \mathcal{A}, P)$ equipped with a filtration $\mathcal{F} = (\mathcal{F}_n)$, and where the step-sizes (γ_n) et (σ_n) are two positive and non-random sequences that go to zero.

Such stochastic approximation algorithms used for the search of zeros of a function h have been widely studied under various assumptions ; see Nevels'on and Has'minskii (1976), Ljung, Pflug and Walk (1992), Duflo (1996), and the references therein.

2.2.2 Criteria of convergence

A straightforward application of Robbins-Monro's theorem gives the convergence of the sequence (Z_n) to a zero z^* of the function h (*see* Duflo [12]. The assumptions required on (γ_n), (σ_n), the function h, and the random perturbations (ε_n) and (r_n) are :

(A1) (γ_n) and (σ_n) verifies the conditions

$$(2.2.3) \gamma_n = O(\sigma_n), \quad \sigma_n^2 = o(\gamma_n), \quad \sum \gamma_n = \infty \text{ and } \sum \sigma_n^2 < \infty.$$

(A2) the vector Z_0 is $\mathcal{F}_0$ measurable.
(A3) there exists a differentiable function $V : \mathbb{R}^d \to \mathbb{R}^+$ with Lipschitz continuous differential dV, such that for all $x \in \mathbb{R}^d$ the estimation $\|h(x)\|^2 \leq cte(1 + V(x))$ holds.
(A4) the sequences (ε_n) and (r_n) are adapted to $\mathcal{F}$ and

$$E(\varepsilon_n|\mathcal{F}_{n-1}) = 0, \quad E(\|\varepsilon_n\|^2|\mathcal{F}_{n-1}) = O(1+V(Z_{n-1})), \text{ and } \sum \gamma_n \|r_n\|^2 < \infty a.s.$$

Théorème 2.2.1 (Robbins-Monro) *Under assumptions (A1)-(A4),and if* $< \nabla V(x), h(x) > \leq 0$ *for all* $x \in \mathbb{R}^d$, *then*

$$\textit{the sequence } V(Z_n) \textit{ converge and } \sum \gamma_n | < \nabla V(Z_n), h(Z_n) > | < \infty.$$

The application of this theorem to the function $V(z) = \|z - z^*\|^2$ gives the convergence of the sequence (Z_n) to a zero z^* of the function h.

2.2.3 Upper bound of the rate of convergence almost sure of general algorithm

Stochastic approximation algorithms (as (2.2.2)) used for the search of zeros of a function h, have been widely studied under various assumptions. The object of this section is not to give the most general existing result on (2.2.2), but only to precisely state the results we shall use in the sequel for the study of (2.2.2) the following conditions :

Let $s_n = \sum_{k=1}^{n} \gamma_k$ and $v_n = \frac{\gamma_n}{\sigma_n^2}$
(A1) There exists $z^* \in \mathbb{R}^d$ such that $Z_n^{(\mu)} \to z^*$ a.s.
(A2) h is differentiable at z^*, its Jacobian matrix H at z^* is symmetric, negative definite with maximal eigenvalue $-L < 0$, and there exists a neighborhood of z^* in which $h(z) = H(z - z^*) + O(\|z - z^*\|^2)$. (3)

(A3) i) $\lim_{n\to\infty}\sigma_n = 0$, $\lim_{n\to\infty}\gamma_n = 0$, $\lim_{n\to\infty} s_n = \infty$ and $\gamma_n/\gamma_{n+1} = 1 + O(\gamma_n)$.
ii) There exist $\xi \in [0, 2L[$ such that

$$\lim_{n\to\infty} \frac{1}{\gamma_n}\left[1 - \frac{v_{n-1}}{v_n}\right] = \xi. \qquad (4)$$

(A4) i) $E(\varepsilon_n + 1|\mathcal{F}_n) = 0$
ii) There exists a nonrandom,positive definite matrix C such that

$$\lim_{n\to\infty} E(\varepsilon_{n+1}\varepsilon_{n+1}^T|\mathcal{F}_n) = C \quad a.s. \qquad (5)$$

iii)there exists a sequence w_n such that $\lim_{n\to\infty} w_n^2\gamma_n \log s_n = 0$ a.s and such $\|\varepsilon_{n+1}\| \leq w_n$
(A5) i) The perturbation $r_{n+1} = R_{n+1}^{(1)} + R_{n+1}^{(2)}$ where $R_{n+1}^{(2)} = O(\|Z_n - z^*\|^2)$ a.s.
ii)There exist $\rho \in \mathbb{R}^d$ such that $\lim_{n\to\infty}\sqrt{u_n}R_{n+1}^1 = \rho$ a.s. and where u_n be a nonrandom sequence which verify the condition : there exist $\tilde{\xi} \in [0, 2L[$ such that

$$\lim_{n\to\infty} \frac{1}{\gamma_n}\left[1 - \frac{v_{n-1}}{v_n}\right] = \tilde{\xi}. \qquad (6)$$

We have the following theorem.

Théorème 2.2.2 *Under assumptions (A'1)-(A'5), we have*

$$\|Z_n - z^*\| = O\left(\max\left\{\frac{1}{\sqrt{u_n}}\ ;\sqrt{\frac{\log s_n}{v_n}}\right\}\right) \quad a.s. \qquad (7)$$

Let (L_n) and (Δ_n) be two sequences defined by :

$$L_{n+1} = e^{s_n H}\sum_{k=1}^{n} e^{-s_k H}\sigma_k\varepsilon_{k+1}, \qquad (8)$$

$$\Delta_{n+1} = (Z_{n+1} - z^*) - L_{n+1}. \qquad (9)$$

The theorem 2.2.2 is a direct consequence of combination of the two following lemmas :

Lemme 2.2.1 (upper bound a.s. of (L_n)) *Under assumptions (A'2)-(A'4), we have*

$$\|L_n\| = O\left(\sqrt{\frac{\log s_n}{v_n}}\right) \quad a.s. \tag{10}$$

Lemme 2.2.2 *Under assumptions (A'1)-(A'5), we have*

$$\|\Delta_n\| \quad = \quad O\left(\frac{1}{\sqrt{u_n}}\right) \; + \; o\left(\sqrt{\frac{\ln s_n}{v_n}}\right) \quad a.s. \tag{11}$$

Lemme 2.2.3 (Rate of convergence a.s. of (Δ_n)) *Under assumptions (A'1)-(A'5), we have*

$$\lim_{n\to\infty} \sqrt{u_n}\Delta_n = -\left[H + \frac{\tilde{\xi}}{2}I\right]^{-1} \rho \;\; a.s. \tag{12}$$

The asymptotic behavior of estimator Z_n of z^* is, governed by two terms :

• the term L_n, which we will call " variance term", which the rate of convergence almost sure to zero($O(\sqrt{\log s_n/v_n})$) depends of choice of (γ_n) et (σ_n) ;

• the term Δ_n, that we can interpret like "term of bias" ; whereas this term is random. Its rate of convergence to zero $(O(1/\sqrt{u_n}))$ is given by the rate of convergence of $R_n^{(1)}$ to zero

Demonstration of lemma 2.2.1

Let $-\lambda$ be an eigenvalue of $H(= H^T)$; using the assumptions (A2) and (A3) ii) we obtain,

(2.2.4) $$\lambda > L > \frac{\xi}{2} \, .$$

Let w be an eigenvector associated with eigenvalue $-\lambda$, and let M_n be the martingale defined as

$$\begin{aligned} M_{n+1} &= w^T \left(\sum_{k=1}^{n} e^{-s_k H} \sigma_k \varepsilon_{k+1} \right) \\ &= \sum_{k=1}^{n} e^{\lambda s_k} \sqrt{\frac{\gamma_k}{v_k}} w^T \ \varepsilon_{k+1}. \end{aligned}$$

applying theorem 6.4.24 in Duflo[12] we obtain :

$$< M >_n \quad = \quad \sum_{k=1}^{n} e^{2\lambda s_k} \frac{\gamma_k}{v_k} w^T C_k w \ ; \qquad (13)$$

The application of lemma 4 of Mokkadem and Pelletier ensures that

(2.2.5) $$\lim_{n\to\infty} v_n e^{-2\lambda s_n} < M >_n = w^T \quad a.s.$$

We deduce from (2.2.5) that

(2.2.6) $$\lim_{n\to\infty} < M >_n = \infty \quad a.s.$$

Noting that the assumption (A'3) ii) implies

$$\begin{aligned} \frac{v_{n-1}}{v_n} &= 1 - \xi\gamma_n + o(\gamma_n) \\ \frac{v_n}{v_{n-1}} &= 1 + \xi\gamma_n + o(\gamma_n) \qquad (14) \end{aligned}$$

Or

$$v_n = v_0 \prod_{k=1}^{n} \frac{v_k}{v_{k-1}}$$

we obtain

$$\begin{aligned} \log v_n &= \log v_0 + \sum_{k=1}^{n} \log \left(\frac{v_k}{v_{k-1}} \right) \\ &= \log v_0 + \sum_{k=1}^{n} \log \left(1 + \xi\gamma_k + o(\gamma_k) \right) \\ &= \xi s_n + o(s_n). \qquad (2.2.7) \end{aligned}$$

And then,

$$\begin{aligned}\log\left[v_n \exp(-2\lambda s_n)\right] &= \xi s_n + o(s_n) - 2\lambda s_n \\ &= (\xi - 2\lambda + o(1)) s_n.\end{aligned}$$

But $\lim_{n\to\infty} s_n = \infty$, (2.2.4) implies that

$$\lim_{n\to\infty} \log\left[v_n \exp(-2\lambda s_n)\right] = -\infty$$

i.e.

$$\lim_{n\to\infty} v_n \exp(-2\lambda s_n) = 0,$$

this show (2.2.6).

Let η be a function defined by $\eta(x) = \sqrt{2x \log\log x}$. It comes from (A'4) iii) and (2.2.5)

$$\begin{aligned}\frac{|M_{n+1} - M_n|}{< M >_n \left[\eta(< M >_n)\right]^{-1}} &= \frac{e^{\lambda s_n}\sqrt{\gamma_n v_n^{-1}}\left|w^T \varepsilon_{n+1}\right|\sqrt{2\log\log < M >_n}}{\sqrt{< M >_n}} \\ &\leq C_n \frac{e^{\lambda s_n}\sqrt{\gamma_n v_n^{-1}} w_n}{\sqrt{2e^{2\lambda s_n} v_n^{-1} w^T}} \sqrt{\log\log(e^{2\lambda s_n} v_n^{-1} w^T} \\ &\leq C'_n \sqrt{\gamma_n} w_n \sqrt{\log s_n}\end{aligned}$$

where (C_n) et(C'_n) are two nonrandom sequences, bounded a.s. The assumption (A'4) iii) ensures existence of a nonrandom sequence $(\tilde{C}_n)$ that goes to zero as n goes to infinity and

$$|M_{n+1} - M_n| \quad \leq \quad \tilde{C}_n < M >_n \left[\eta(< M >_n)\right]^{-1} \qquad (15)$$

The application of theorem 6.4.24 in Duflo (2) then ensures

$$\lim\sup_{n\to\infty} \frac{|M_n|}{\eta(< M >_n)} \leq 1 \quad a.s.$$

With (2.2.5), we have

$$\begin{aligned}|M_n| &= O\left(e^{\lambda s_n} v_n^{-1/2}\sqrt{\log\log(e^{2\lambda s_n} v_n^{-1})}\right) \quad a.s. \\ &= O\left(\frac{e^{\lambda s_n}\sqrt{\log s_n}}{\sqrt{v_n}}\right) \quad a.s. \qquad (16)\end{aligned}$$

Let $w^T L_{n+1} = e^{-\lambda s_n} M_{n+1}$, we have

$$\left|w^T L_n\right| = O\left(\frac{\sqrt{\log s_n}}{\sqrt{v_n}}\right) \quad a.s.,$$

This for all eigenvector w de H^T. It comes from assumption (A'2), the matrix H being diagonalizable, on $\mathbb{R}$; and it follows then

$$\|L_n\| = O\left(\sqrt{\frac{\log s_n}{v_n}}\right) \quad a.s. \qquad (17)$$

Demonstration of lemma 2.2.3

From assumptions (A'2) et (A'5), we can rewrite (3.2.2) as :

$$Z_{n+1} - z^* = (I + \gamma_n H)(Z_n - z^*) + \gamma_n r_{n+1} + \sigma_n \varepsilon_{n+1} \qquad (18)$$

Where r_{n+1} satisfying the assumption (A'5). One has

$$\begin{aligned} L_{n+1} &= \sigma_n \varepsilon_{n+1} + e^{\gamma_n H} L_n \\ &= \sigma_n \varepsilon_{n+1} + \left[I + \gamma_n H + O\left(\gamma_n^2\right)\right] L_n \end{aligned} \qquad (19)$$

we obtain

(2.2.8) $$\Delta_{n+1} = (I + \gamma_n H)\Delta_n + \gamma_n \left[O(\gamma_n) L_n + r_{n+1}\right].$$

Set $A \in]\xi/2, L[$; the matrix H is attractive, in view of proposition 3.I.2 in Duflo [12] there exist a matrix norm $|||.|||_A$ and a real $a \in]0, 1/A[$ such that, for all $\gamma \leq a$,

$$|||I + \gamma H|||_A \leq 1 - \gamma A \qquad (20)$$

.

Now, for all $x \in \mathbb{R}^d$, define the matrix $M(x) = [xx...x]$ all of whose columns are equal to x; the function $\|.\|_A$ defined on $\mathbb{R}^d$ by

$$\|x\|_A = |||M(x)|||_A \qquad (21)$$

is then a vector norm compatible with the matrix norm $|||.|||_A$; voir Horn and Johnson (1985). For n large enough, we thus obtain

$$\|\Delta_{n+1}\|_A \leq (1 - A\gamma_n)\|\Delta_n\|_A + \gamma_n \left[O(\gamma_n)\|L_n\|_A + \|r_{n+1}\|_A\right]. \qquad (22)$$

Since

$$\lim_{n\to\infty} \left[O(\gamma_n)\|L_n\|_A + \|r_{n+1}\|_A\right] = 0 \quad a.s.$$

The application of lemma 4.I.1 in Duflo [12] ensures that

$$\lim_{n\to\infty} \|\Delta_n\|_A = 0 \quad a.s.$$

In the other hand, in view of assumption (A'5), we have

$$\begin{aligned} r_{n+1} &= R^{(1)}_{n+1} + O\left(\|Z_n - z^*\|^2\right) \quad a.s. \\ &= R^{(1)}_{n+1} + O\left(\|L_n\|^2\right) + O\left(\|\Delta_n\|^2\right) \quad a.s. \end{aligned} \tag{23}$$

The equation (2.2.8) can then be rewritten in following form :

$$\Delta_{n+1} = [I + \gamma_n H]\,\Delta_n + \gamma_n\left[O(\gamma_n)\,L_n + O(\|L_n\|^2) + O(\|\Delta_n\|^2) + R^{(1)}_{n+1}\right].$$

Since $\lim_{n\to\infty}\Delta_n = 0$ a.s., we have

$$\Delta_{n+1} = [I + \gamma_n H + o(\gamma_n)]\,\Delta_n + \gamma_n\left[O(\gamma_n)\,L_n + O(\|L_n\|^2) + R^{(1)}_{n+1}\right].$$

We have then :

$$\begin{aligned} \sqrt{u_{n+1}}\Delta_{n+1} &= \left[\frac{u_{n+1}}{u_n}\right]^{1/2}(I + \gamma_n H + o(\gamma_n))\sqrt{u_n}\Delta_n \\ &+ \gamma_n\sqrt{u_{n+1}}\left[O(\gamma_n)\,L_n + O(\|L_n\|^2) + R^{(1)}_{n+1}\right]. \end{aligned}$$

We note that the condition of assumption (A'3) ensures that

$$\frac{u_n}{u_{n+1}} = 1 - \tilde{\xi}\gamma_n + o(\gamma_n)$$

i.e. that

$$\left[\frac{u_{n+1}}{u_n}\right]^{1/2} = 1 + \frac{\tilde{\xi}}{2}\gamma_n + o(\gamma_n). \tag{24}$$

We deduce that

$$\begin{aligned}\sqrt{u_{n+1}}\Delta_{n+1} &= \left(I+\gamma_n\left[H+\frac{\tilde{\xi}}{2}I\right]+o(\gamma_n)\right)\sqrt{u_n}\Delta_n \\ &+ \gamma_n\sqrt{u_{n+1}}\left[O\left(\gamma_n\right)L_n+O(\|L_n\|^2)+R_{n+1}^{(1)}\right].\end{aligned}$$

By setting $m=-\left[H+\frac{\tilde{\xi}}{2}I\right]^{-1}\rho$ and $\delta_n=\sqrt{u_n}\;\Delta_n-m$. we have :

$$\begin{aligned}\delta_{n+1} &= \left(I+\gamma_n\left[H+\frac{\tilde{\xi}}{2}I\right]+o(\gamma_n)\right)\delta_n+\left(\gamma_n\left[H+\frac{\tilde{\xi}}{2}I\right]+o(\gamma_n)\right)m \\ &\quad+\gamma_n\left[O\left(\gamma_n\right)\sqrt{u_{n+1}}L_n+\sqrt{u_{n+1}}O(\|L_n\|^2)+\sqrt{u_{n+1}}R_{n+1}^{(1)}\right] \\ &= \left(I+\gamma_n\left[H+\frac{\tilde{\xi}}{2}I\right]+o(\gamma_n)\right)\delta_n+\gamma_n\left[O\left(\gamma_n\right)\sqrt{u_{n+1}}L_n+o(1)+\sqrt{u_{n+1}}O(\|L_n\|^2)\right. \\ &\quad\left.+\sqrt{u_{n+1}}R_{n+1}^{(1)}+\left[H+\frac{\tilde{\xi}}{2}I\right]m\right] \qquad (25)\end{aligned}$$

Since $\tilde{\xi}<2L$, the matrix $H+\frac{\tilde{\xi}}{2}I$ is attractive. we can apply proposition 3.I.2 in Duflo [12] again. Let $A\in]\tilde{\xi}/2,L[$; there exist a matrix norm $|\!|\!|.|\!|\!|_A$ and a real $a\in]0,1/A[$ such that, for all $\gamma\leq a$,

$$\left|\!\left|\!\left|I+\gamma\left[H+\frac{\tilde{\xi}}{2}I\right]\right|\!\right|\!\right|_A\leq 1-\gamma A \qquad (26)$$

Moreover $\|.\|_A$ denotes the vector norm $|\!|\!|.|\!|\!|_A$ (for all $x\in\mathbb{R}^d$, $\|x\|_A=|\!|\!|M(x)|\!|\!|_A$) the matrix norm induced by the Euclidean vector norm. For n large enough, we have

$$\begin{aligned}\|\delta_{n+1}\|_A &\leq (1-A\gamma_n+o(\gamma_n))\|\delta_n\|_A+\gamma_n\left[O\left(\gamma_n\right)\sqrt{u_{n+1}}\|L_n\|_A+o(1)+\sqrt{u_{n+1}}O(\|L_n\|^2)\right. \\ &\quad\left.+\left\|\sqrt{u_{n+1}}R_{n+1}^{(1)}-\rho\right\|_A\right].\end{aligned}$$

Let $B\in]\tilde{\xi}/2,A[$; for n large enough, we get

$$\|\delta_{n+1}\|_A\leq\left(1-B\gamma_n\right)\|\delta_n\|_A+\gamma_n o(1)\quad a.s. \qquad (27)$$

The application of lemma 4.I.1 in Duflo [12] then ensures

$$\lim_{n\to\infty} \delta_n = 0 \quad a.s.$$

i.e.

$$\lim_{n\to\infty} \sqrt{u_n}\Delta_n = -\left[H + \frac{\tilde{\xi}}{2}I\right]^{-1} \rho \quad a.s. \tag{28}$$

2.2.4 Quantiles

The problem addressed in this thesis is the non-parametric estimation of quantile. In this part, we discuss quantile estimation in a fixed population. Supposed we are interested in estimating the q-quantile (or the q-th percentile) of an unknown but fixed distribution function F based on an i.i.d. random sample of size n.

Quantiles are important descriptors of univariate distributions in statistical practice. The extreme values which a random variable X may take on are usually best characterized by the quantiles of the random variable. Then quantiles are useful measures because they are less susceptible to long-tailed distributions and outliers. Empirically, if the data being analyzed are not actually distributed according to an assumed distribution, or if there are other potential sources for outliers that are far removed from the mean, then the quantiles may be more useful descriptive statistics than means and other moment-related statistics. For a variable $X \in \mathbb{R}$, the quantile function is defined from the inverse of its distribution function. When this distribution function is strictly increasing, its inverse is unambiguously defined. But a distribution function remains constant over any interval in which the random variable can not take values.

Given a random variable X with continuous distribution function F, representing some one-dimensional population, the α^{th} quantile (or quantile of order α) of the population is

$$F^{-1}(q) = \inf\{x : F(x) \leq q\}$$

with $0 < \alpha < 1$. Various quantile estimators have been proposed and investigated over the years, resulting in a large literature on this subject. See works in Azzalini [2] and Ralescu [29].

The most important class of functions to be used in sequential quantile estimation schemes are stochastic approximation estimators. There is an extensive literature on so-called stochastic approximation methods; these methods are intended to find the root $x = q$ of

$$F(x) = q.$$

Most work on stochastic approximation has been concerned with specifying conditions under which the sequence of estimators converges probabilistically to the correct value. One sequential design strategy known as stochastic approximation is to choose $x_1, x_2, \ldots$ such that x_n go to 0 in probability. Robbins and Monro(1951) proposed the procedure

$$x_{n+1} = x_n - \gamma_n(y_n - q)$$

where y_n is the response observed at x_n and γ_n is a pre-specified sequence of positive constants. Robbins and Monro showed that x_n go to θ in probability. Sequential quantile estimation, which is called incremental quantile estimation in the computer science literature, is not a new topic. Robbins and Monro (1951) introduced the idea of stochastic approximation for quantile estimation, for example. Tierney (1983) used it for monitoring computer simulations. Stochastic approximation is best suited for continuous data because it requires an estimate of the density near the quantile. The data in our application, such as packet sizes, are often discrete and can often have preferred values and spikes, so any continuity assumption is suspect. The quality of quantile estimates depends on the quality of the estimate of the cumulative distribution function CDF.

According to Thiam [1], quantile estimation is an active field in statistics and has been widely explored in various areas, including finance, economics, biology and hydrology. There are many circumstances under which it is essential to know the quantiles of a given distribution. An overview of the use of special quantiles such as quartiles, deciles and percentiles in statistical applications appears in the monograph of Fisz (1963)and Gourieroux et al. (2000) and the references therein. In financial econometrics, technological innovations and trade globalization have provided the financial community the opportunity to develop sophisticated statistical techniques for risk management. Several measures of risk have been introduced, the most popular being the so-called Value at Risk. This measure has become the standard and most useful risk measure for some financial institutions. Although, the kernel estimator can be preferable to the sample estimator, it suffers from computational problems.

In fact, there is no explicit expression for the inverse of the kernel distribution function estimator. Thus, quantile estimation is obtained by numerical approximation methods such as the Newton–Raphson method. However, the convergence of this algorithm requires a large number of iterations and can thus be time-consuming for large sample sizes. This situation is more apparent in data streams, i.e., in which sample data are obtained by an observational mechanism that allows for an increase in sample size over time. In recent years, data streams have become an increasingly important area of research. Common data streams include Twitter activity, the Facebook news stream, Internet packet data, stock market activity, credit card transactions and Internet and phone usage.

Consider a probability space (Ω, F, P) and a random variable X admitting cumulative distribution function (CFD)

$$F(x) = P(X \leq x)$$

and density function

$$f(x) = F^{'}(x)$$

for all $x \in \mathbb{R}$. Let q_p be its pth quantile, that is,

$$q_p = \inf\{x : F(x) \geq p\}$$

According to Azzalini, a distribution function is estimated by integrating a kernel estimator of the density. Quantiles are estimated by inverting the estimate of the distribution function. Second-order properties of both these estimators are studied. In Chen approaches Recursive Quantile Estimation with Application to Value at Risk. The simulation study in Tierney (1983) showed that the standardized mean squared errors of the stochastic algorithms estimator and the sample quantile were quite close even for small sample sizes. However, the stochastic algorithms estimator involves an indicator function in the incremental estimation of the quantile. This discrete component will increase abruptness, the step makes no difference between the observations that are close to the quantile estimate and the observations that are extremely larger or smaller than the quantile estimate. It increases the quantile estimates by the same amount regardless of whether the observation exceeds the last quantile estimate by a small margin or a large margin ; besides, consider the problem of computing an estimate of a percentile or quantile of an unknown population based on a random sample of n observations. By viewing this problem as a problem in stochastic approximation, we obtain an estimator that requires only a small amount of direct access storage space that does not increase with the sample size. We show that a modified version of the simple stochastic approximation estimator has the same large-sample behavior as the sample quantile, which has the smallest asymptotic variance among all reasonable estimators. The modified procedure also yields an estimate of the asymptotic variance of the estimator. Some simulation results are presented to show that the proposed estimator performs well in samples of moderate size. Holst approximation theorems can be used to obtain the asymptotic properties when the observation are independent. For dependent sequences martingale theory cannot be applied straight forwardly as the tool for asymptotic analysis. In his paper, he consider both the case when the observations are i.i.d. and when they form a stationary and strongly regular process. The main result is sufficient condition for almost sure convergence in the strongly regular case.

2.3 Truncated Poisson distribution

2.3.1 The Poisson distribution

In this section we introduce a probability which can be used when the outcome of an experiment is a random variable taking on positive integer values and where the only information available is a measurement of its average value. The Poisson distribution is a discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time if these events occur with a known constant mean rate and independently of the time since the last event. It can also be used for the number of events in other types of intervals than time, and in dimension greater than 1.

The Poisson distribution is a probability model which can be used to find the probability of a single event occurring a given number of times in an interval of (usually) time. The occurrence of these events must be determined by chance alone which implies that information about the occurrence of any one event cannot be used to predict the occurrence of any other event. It is worth noting that only the occurrence of an event can be counted ; the non-occurrence of an event cannot be counted. This contrasts with Bernoulli trials where we know the number of trials, the number of events occurring and therefore the number of events not occurring. The Poisson distribution has widespread applications in different areas. Poisson distributions have been used to describe many events. For example, they can be used to explain or predict : number of text messages per hour, number of male sheeps per hectare, number of machine breakdowns per year, number of website visitors per day, number of cases of influenza per year... The Poisson distribution is also commonly used to model financial count data where the tally is small and is often zero. As one example in finance, it can be used to model the number of trades that a typical investor will make in a given day, which can be 0 (often), 1, or 2.

Under a Poisson distribution with the expectation of λ events in a given interval, the probability of k events in the same interval is :

$$(Y_i = j) = \frac{\exp(-\lambda) \times \lambda^j}{j!}$$

For instance, consider a call center which receives, randomly, an average of $\lambda = 3$ calls per minute at all times of day. If the calls are independent, receiving one does not change the probability of when the next one will arrive. Under these assumptions, the number k of calls received during any minute has a Poisson probability distribution.

2.3.2 The truncated Poisson distribution

Zero-truncated Poisson distribution

The zero-truncated Poisson distribution is a certain discrete probability distribution whose support is the set of positive integers. This distribution is also known as the conditional Poisson distribution or the positive Poisson distribution. It is the conditional probability distribution of a Poisson-distributed random variable, given that the value of the random variable is not zero. Thus it is impossible for a zero-truncated Poisson random variable to be zero. Consider for example the random variable of the number of items in a shopper's basket at a supermarket checkout line. Presumably a shopper does not stand in line with nothing to buy (i.e., the minimum purchase is 1 item), so this phenomenon may follow a zero-truncated Poisson distribution.The zero-truncated Poisson distribution is a certain discrete probability distribution whose support is the set of positive integers. This distribution is also known as the conditional Poisson distribution or the positive Poisson distribution. It is the conditional probability distribution of a Poisson-distributed random variable, given that the value of the random variable is not zero. Thus it is impossible for a zero-truncated Poisson random variable to be zero. Consider for example the random variable of the number of items in a shopper's basket at a supermarket checkout line. Presumably a shopper does not stand in line with nothing to buy (i.e., the minimum purchase is 1 item), so this phenomenon may follow a zero-truncated Poisson distribution.

Let X denote the random variable representing the number of insurance claims. If λ is the average rate of claims, the zero-truncated Poisson probability mass function takes the form :

$$P(X = k|X > 0)) = \frac{\lambda^k e^{-\lambda}}{k!\,(1 - e^{-\lambda})} \qquad (29)$$

for $k = 1, 2, 3, ...$

This formula encapsulates the probability of observing k claims given that at least one claim has transpired. The denominator ensures the exclusion of the improbable zero-claim scenario. By utilizing the zero-truncated Poisson distribution, the manufacturing company can analyze and predict the frequency of defects in their products while focusing on instances where defects exist. This distribution helps in understanding and improving the quality control process, especially when it's crucial to account for at least one defect.

a-truncated Poisson distribution

This section works of the a-truncated Poisson distribution, a special case of which is the zero-truncated Poisson distribution. The a-truncated Poisson distribution is the distribution of a Poisson random variable Y conditional on the event $Y > a$. It has one parameter, which we may take to be $\mu = E(Y)$. Since μ is not the mean of the distribution of Y conditioned on the even $Y > a$ we do not call μ the mean, rather we call it the original parameter. If f_μ is the probability mass function of Y, then the probability mass function g_μ of the a-truncated Poisson distribution is defined by

$$g_\mu(x) = \frac{f_\mu(x)}{1 - \sum_{j=0}^{a} f_\mu(x) f_\mu(j)}$$

pour $x = a + 1, a + 2, ...$

The truncated Poisson distribution can be used to model count data where only values above a certain threshold can be observed. If we study observations whose lower bound is a, the frequency of observations less than or equal to a is not zero while the density of the a priori probability law chosen to carry out a statistical adjustment on these data is zero or undefined at this point.If we study observations whose lower bound is k, the frequency of observations less than or equal to k is not zero while the density of the a priori probability law chosen to carry out a statistical adjustment on these data is zero or undefined at this point. We therefore consider that all observations below a strictly positive threshold k reflect precipitation below k ; among precipitation below k, we do not know the number of days when we should have observed a non-zero precipitation height. We are interested in the probabilities of occurrence of precipitation above k, and in particular the probability of absence of rain.

We introduce the a-truncated degenerate Poisson random variable with parameter $\lambda > 0$, whose probability mass function is given by p properties of this random variable, $(i = a+1, a+2, a+3, ..)$, and investigate various properties of this random variable.They applies to cornavirus pandemic. The cornavirus pandemic has been spreading unpredictably around the world, terrorizing many people. Although several vaccines for the coronavirus have been developed and many people are getting vaccinated, they have many unexpected side effects as well. We would like to predict stability of the coronavirus vaccines after a-days the vaccines were shot. For this purpose, we study the a-truncated degenerate Poisson random variables, which has the 'degenerate factor' λ reflecting abnormal situations. Indeed, we elabore a theorem is useful in predicting the probability of the coronavirus vaccines becoming stable after the r-days of getting vaccinated.

If the entire population is of interest, then conventional statistical inference (such as estimation) on the truncated sample produces a systematic bias known as (of course) "truncation bias." This would arise, for example, if an ordinary Poisson model intended to characterize the full population is fit to the sample from a truncated population. However, it is also possible to reconstruct the estimators specifically to account for the truncation or censoring in the data

Chapitre 3

Quantile recursive

3.1 Introduction

Quantile plays a fundamental role in various statistical applications. For example, a company wishes to have information on the rate of electro- magnetic waves absorbed by a fetus when its mother uses a cell phone. Various studies have been carried out, and it is now known that this rate depends on several parameters, each parameter has values in $[0, 1]$. Thus, one goal will be to find a good estimate of the quantile of the absorption rate distribution, with as few measurements of this rate as possible. Hydrologists were among the first to take an interest in these two problems. Having a sample of the heights of a watercourse, they asked themselves the following two questions : 1) what is the height of water that is reached or exceeded for a given low probability ? 2) for a fixed great water level, what is the probability of observing a higher water level ?

Questions 1) and 2) therefore relate respectively to the estimation of an extreme quantile and of a small probability. Quantiles are more useful descriptive sta- tistics than means because they are less susceptible to long-tailed distributions and outliers. They often arise as the natural parameters to estimate when the distribution is skewed.

Let $0 < q < 1$ be a probability. The q^{th} quantile is the smallest number $x \in \mathbb{R}$ so that $P(X > x) \leq 1 - q$, given a random variable X with a continuous distribution function $F(.)$. The statistical inference will be based on independently and identically distributed observations $X_1, ..., X_n$ for a fixed n.

Nonparametric is used to say that observations $X_1, ..., X_n$ come from an unknown distribution $F \in \mathcal{F}$ with $\mathcal{F}$ being the class of all continuous and strictly increasing distribution functions and, for a given $q \in (0,1)$, we are interested in estimating the q^{th} quantile θ_q of the distribution F.

For a distribution function F, the q^{th} quantile θ_q of F is defined as $\theta_q = F^{-1}(q)$ with

(3.1.1) $$F^{-1}(q) = \inf\{x : F(x) \geq q\}.$$

For $F \in \mathcal{F}$ and $q \in (0,1)$ it always exists and is uniquely determined. Rather than direct inversion of the distribution function, the quantile can be determined through a traditional nonparametric estimator of the distribution function, noted $F_n(x)$. If $F_n(x)$ is continuous and strictly increasing empirical distribution function, so is its inverse

(3.1.2) $$F_n^{-1}(q) = \inf\{x : F_n(x) \geq q\}$$

as an estimator $\widehat{\theta}_q$ of θ_q, and can be considered as a continuous and strictly increasing function of $q \in (0,1)$. Different definitions of F_n lead of course to different estimators θ_q.

One can say that the variety of definitions of F_n (left- or right-continuous step functions, smoothed versions, F_n as a kernel estimator of F, etc) is what produces the variety of estimators to be found in abundant literature in mathematical statistics.

Let $X_1, ..., X_n$ be independent identically distributed random variables, and let f and F denote respectively the probability density and the distribution function of X_1. The kernel estimator of density of probability f is defined by

(3.1.3) $$f_n(x) = \frac{1}{nh_n} \sum_{i=1}^{n} k\left(\frac{x - X_i}{h_n}\right)$$

where the bandwidth (h_n) is a sequence of real numbers strictly positives who goes to zero, and the kernel k is a function continuous on $\mathbb{R}$, with values in $\mathbb{R}$.

A distribution function is estimated by integrating a kernel estimator of the density. We introduce a bandwidth (h_n) (that is, a sequence of positive real numbers that goes to zero), and a kernel k (that is, a function satisfying $\int\limits_{-\infty}^{\infty} k(x)dx = 1$).

The kernel estimator of the quantile q of F is a random variable q_n satisfying $F_n(q_n) = \delta$. There is no explicit expression for the inverse of the kernel distribution function estimator. Thus,quantile estimation is obtained by numerical approximation methods.

Our aim is to provide an algorithm in order to approximate the a quantile q. To construct a stochastic algorithm, which approximates the function q, we define an algorithm of search of the zero of the function $\varphi : x \to \varphi(x) = \delta - F_n(x)$. We thus proceed in the following way :

1. we set $q_0 \in [0, 1]$
2. for all $n \geq 1$, we set

$$q_n = q_{n-1} + \gamma_n W_n$$

where $(W_n)_n$ is a sequence of functions, $W_n : \mathbb{R} \to \mathbb{R}$ is an observation of the function φ at the point x_{n-1}, and the stepsize (γ_n) is a sequence of positive real numbers that goes to zero.

The use of stochastic approximation algorithms in the frameworks of non parametric statistics was widely discussed. In particular, Mokkadem and Pelletier [26] provided a companion algorithm to Kiefer-Wolfowitz's algorithm in order to simultaneously approximate the location and the size of the mode of the regression function.

Since the distribution function F is rarely known exactly, quantiles and other features of the distribution function F must be estimated from the data.

Quantiles finds application, for example, in testing statistical hypotheses and in characterizing the extreme values of the distribution of X when δ is near 0 or 1 [31]. The extreme values which a random variable X may take on are usually best characterized by its quantiles. Various quantile estimators have been proposed and investigated over the years, resulting in a large literature on this subject. See also works in Azzalini [2] and Ralescu [29]. If nothing is known about $F(.)$, one must resort to non-parametric for estimating q.

The most important class of functions to be used in sequential quantile estimation schemes is stochastic approximation estimators. Stochastic approximation algorithms were also introduced by Révész ([30]) to estimate a regression function and by Tsybakov [36] to approximate the mode of a probability density. There is an extensive literature on so-called stochastic approximation methods ([26], [12], [31], [23]).

3.2 Some preliminary results

Let $X_1, ..., X_n$ be a random sample drawn from unknown continuous distribution function $F(x)$ with density function $f(x)$. We assume that the underlying distribution function $F(x)$ satisfies

(H1) $F(x)$ is twice continuously differentiable with $f(x)$ a bounded differentiable on a neighborhood of quantile.

A natural apporach to estimating a quantile is firstly estimating the distribution and then calculating the target quantile.

Let f_n be the usual kernel estimate of $f(x)$ with appropriate kernel function $k(t)$ and smoothing parameter h_n what is a sequence of positive constants which tends to zero with increasing n; for example $h_n = hn^{-1/3},\ h > 0$.

The empirical density function estimator at the point x is then given by (3.1.3). Using the relationship between the density and the distribution function, that is $F(x) = \int_{-\infty}^{x} f(t)dt$; it is easy to construct a kernel estimator for the distribution function as :

$$
\begin{aligned}
F_n(x) &= \int_{-\infty}^{x} f_n(t)dt \\
&= \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right) \qquad (3.2.1)
\end{aligned}
$$

with $K(x) = \int_{-\infty}^{x} k(t)dt$. It is in a sense already quite smooth, although it is known that further smoothing can be an advantage. $k(t)$ is a probability density such that it is bounded, symmetric around zero ($k(t) = k(-t)$) and has finite support. Moreover $k(t)$ satisfies

(H2) k is a Lipschitz and continuous function such that $\int k(u)du = 1$, $\int uk(u)du = 0$ and $0 < \int u^2 k(u)du < \infty$.

An example of kernel is the triangular weight function

$$
\begin{aligned}
k(x) &= 1 - |x| \qquad |x| \leq 1 \\
&= 0 \ \ \text{otherwise.}
\end{aligned}
$$

Using this kernel, Lejeune and Sarda [22] shows that

$$
\begin{aligned}
EK^2\left(\frac{x-X}{h_n}\right) &= F(x) + h_n f(x)\int_{-1}^{1} K^2(z)dz - h_n f(x)\int_{-1}^{1} K(z)dz + O(h_n^2) \\
&= F(x) + h_n f(x)\int_{-1}^{1}\left(K^2(z) - 1\right)dz + O(h_n^2) \qquad (30)
\end{aligned}
$$

Subsequently we shall use this kernel. It was shown that kernel estimator $F_n(x)$ is better than usual empirical distribution function in the sense of mean integrated squared error. Several properties of $F_n(x)$ have been investigated; among others, Nadaraya [24] proved the uniform convergence of $F_n(x)$ to $F(x)$ with probability one.

As pointed out by many authors, the choice of the kernel k is not very crucial but the choice of the smoothing parameter is a serious problem that has been addressed in the literature extensively. If the bandwidth is small, we will obtain an under-smoothed estimator, with high variability. On the contrary, if the value of h_n is big, the resulting estimator will be very smooth and farther from the function that we are trying to estimate.

According to [31], the most important class of functions to be used in sequential quantile estimation schemes are stochastic approximation estimators. There is an extensive literature on so-called stochastic approximation methods ; these methods are intended to find the root $x = \theta$ of the regression function $E[Y(x)] = M(x) = a$ where the only information available consists of independent observations on the random variable $Y(x)$. We note that this is a more general problem than the quantile estimation problem considered here. Most work on stochastic approximation has been concerned with specifying conditions under which the sequence of estimators converges probabilistically to the correct value.

The first, and prototypical, algorithms of this kind are the Robbins-Monro and Kiefer-Wolfowitz algorithms [12]. We present the basic form of the stochastic approximation algorithm as it applies to quantile estimation. We consider the stochastic algorithm defined by

$$Z_{n+1} = Z_n + \gamma_n \left(h(Z_n) + r_{n+1} + c_n^{-1} \varepsilon_{n+1} \right) \tag{3.2.2}$$

where the random variable Z_0, $(r_n)_{n \geq 1}$ and $(\varepsilon_n)_{n \geq 1}$ are defined on a space of probability $(\Omega, \mathcal{A}, P)$ and adapted to the filtration $\mathcal{G} = (\mathcal{G}_n)$, and where (γ_n) et (c_n) are two deterministic positive sequences which go to zero when $n \to \infty$. The sequences (ε_n) and (r_n) are adapted to $\mathcal{G}$.

An attractive feature of recursive stochastic algorithms [37], is that they allow fast updating at each instant when new data arrive and therefore can be used to support decisions "on-line", i.e., during the operation of the system. Moreover, they can be asymptotically as efficient as their "off-line" counterparts that require much greater computational complexity and whose memory requirements grow to ∞ with the sample size. However, the loss of information in using only the previous estimate q_n and the current observation x_n to form the current estimate Z_n of Z often results in unsatisfactory finite-sample performance except for linear problems in which off-line estimates have simple recursive forms.

For example [37], in applying the Robbins-Monro scheme to estimate the p^{th} quantile q_p of a distribution function F, the nonlinearity of F and the binary nature of the response suggest that there is substantial loss of information to use an stochastic approximation scheme instead of regression modeling with all the data collected so far. The convergence of the sequence (Z_n) to a zero z^* of the function h can be proved with use of Robbins-Monro theorem. Thus we need the following Robbins-Monro's Theorem (see for instance [12]).

Théorème 3.2.1 *(Robbins-Monro) We assume that assumptions* $\gamma_n = O(c_n),\ \ c_n^2 = o\left(\gamma_n\right), \sum_n \gamma_n = \infty \ \ \sum \gamma_n^2 c_n^{-2} < \infty$ *holds and the vector* Z_0 *is* $\mathcal{F}_0$ *measurable.*

There exist a continuously differentiable function $V : \mathbb{R} \to \mathbb{R}^+$, *such that :* ∇V *is Lipschitz-continuous and such that, for all* $x \in \mathbb{R}$, $\|h(x)\|^2 \le cte(1 + V(x))$ *and*

- $[\nabla V(x)]^T h(x) \le 0\,;\ E(\varepsilon_n|\mathcal{G}_{n-1}) = 0 \ \ a.s,$
- $E(\|\varepsilon_n\|^2|\mathcal{G}_{n-1}) = O(1 + V(Z_{n-1})) \ \ a.s,$
- $\sum \frac{c_n^2}{\gamma_n}\|r_{n+1}\|^2 < \infty$ *holds, and*

if $\langle\nabla V(x), h(x)\rangle \le 0$ *for* $\forall x \in \mathbb{R}^d$, *then the sequence* $V(Z_n)$ *converges and* $\sum |\gamma_n| \ \langle\nabla V(Z_n), h(Z_n)\rangle < \infty$

Remarque 3.2.1 *The application of this theorem to the function* $V(z) = \|z - z^*\|^2$ *give the convergence of the sequence* (Z_n) *to a zero* z^* *of the function* h.

A quantile estimation is not a new topic. Robbins and Monro [31] introduced the idea of stochastic approximation for quantile estimation, Tierney [35] used it for monitoring computer simulations.

In order to construct an algorithm, which approximates the quantile of distribution function (in other words, which approximates q, we define an algorithm searching the zero of the function $h : z \mapsto \delta - F(z)$). Following Robbins-Monro's scheme, we set $q_0 \in \mathbb{R}$, and, for $n \ge 1$,

$$q_n = q_{n-1} + \gamma_n \left(\delta - K\left(\frac{q_{n-1} - X_n}{h_n}\right)\right) \tag{3.2.3}$$

The usual steps (see [10]) : $\gamma_n = \frac{\gamma_0}{n^\alpha}$ and $\sigma_n = \frac{\sigma_0}{n^s}$ with $\gamma_0 > 0$, $\sigma_0 > 0$ and $\frac{1}{2} < s \le \alpha < 1$; or, more generally, by the steps of type (α, β) defined in Pelletier [26].

Also, we can use $\gamma_n = \frac{\gamma_0}{n}$ and $\sigma_n = \frac{\sigma_0}{n^s}$ with $\frac{1}{2} < s < 1$, $\sigma_0 > 0$ provided that γ_0 is chosen so that $\gamma_0 > \frac{2s-1}{2L}$. However, we can not use the steps $\gamma_n = \sigma_n = \frac{\gamma_0 \log n}{n}$ where $\gamma_0 > 0$ since, in this case, $\lim_{n\to\infty} \left[1 - \frac{v_{n-1}}{v_n}\right] = \infty$.

We adopted the approach developed in [36]. In order to avoid the estimation of the density f on all $\mathbb{R}^d$ to estimate the mode θ, in a pioneering work, Tsybakov introduced a recursive estimator of θ. This recursive estimator is defined by

$$\theta_n = \theta_{n-1} + \frac{\gamma_n}{h_n^{d+1}} \nabla K \left(\frac{\theta_{n-1} - X_n}{h_n}\right) \tag{31}$$

where θ_0 is an arbitrary vector of $\mathbb{R}^d$ and where the step (γ_n) is a sequence of strictly positive real numbers that tends to zero and such that $\sum \gamma_n = \infty$.

We prove that q_n is strongly consistent, and we establish the weak convergence rate of q_n defined by the algorithms (3.2.3).

3.2.1 Writing (3.2.3) in the form of stochastic approximation algorithm

The approach adopted is to reduce the study of recursive estimators q_n to that of stochastic algorithms. Let $\mathcal{G} = (\mathcal{G}_n)$ be the natural filtration associated to the sequence (X_n).

$$\begin{aligned} q_n &= q_{n-1} + \gamma_n \left\{\delta - E\left[K\left(\frac{q_{n-1} - X_n}{h_n}\right) |\mathcal{G}_{n-1}\right]\right\} \\ &\quad -\gamma_n \left\{K\left(\frac{q_{n-1} - X_n}{h_n}\right) - E\left[K\left(\frac{q_{n-1} - X_n}{h_n}\right) |\mathcal{G}_{n-1}\right]\right\}. \end{aligned} \tag{32}$$

Let's calculate

$$E\left[K\left(\frac{q_{n-1} - X_n}{h_n}\right) |\mathcal{G}_{n-1}\right]$$

We have

$$E\left[K\left(\frac{q_{n-1} - X_n}{h_n}\right) |\mathcal{G}_{n-1}\right] = \int_{-\infty}^{\infty} K\left(\frac{q_{n-1} - x}{h_n}\right) f(x)dx \tag{33}$$

$$
\begin{aligned}
&= \int_{-\infty}^{\infty}\int_{-\infty}^{q_{n-1}} k\left(\frac{t-x}{h_n}\right) f(x)dtdx \\
&= \int_{-\infty}^{q_{n-1}}\int_{-\infty}^{\infty} k(y)f(t-h_ny)dydt \\
&= \int_{-\infty}^{\infty} k(y)\int_{-\infty}^{q_{n-1}} f(t-h_ny)dtdy \\
&= \int_{-\infty}^{\infty} k(y)F(q_{n-1}-h_ny)dy.
\end{aligned}
$$

Using a Taylor expansion of F, then we deduce :

$$
\begin{aligned}
E\left[K\left(\frac{q_{n-1}-X_n}{h_n}\right)|\mathcal{G}_{n-1}\right] &= F(q_{n-1}) \\
&+ \frac{h_n^2}{2}F^{''}(\zeta_n)\left[\int_R x^2k(x)dx\right] + o(h_n^2).
\end{aligned} \tag{3.2.4}
$$

Consequently, (3.2.3) can be written as :

$$
q_n = q_{n-1} + \gamma_n\left[h(q_{n-1}) + r_n + \varepsilon_n\right]. \tag{3.2.5}
$$

With

$$
\begin{aligned}
r_n &= F(q_{n-1}) - E\left[K\left(\frac{q_{n-1}-X_n}{h_n}\right)|\mathcal{G}_{n-1}\right] \\
c_n &= 1 \\
\varepsilon_n &= K(q_{n-1}, X_n) - E\left[K\left(\frac{q_{n-1}-X_n}{h_n}\right)|\mathcal{G}_{n-1}\right].
\end{aligned}
$$

Remarque 3.2.2 *With this notations, we have*

$$
E(\varepsilon_n|\mathcal{G}_{n-1}) = 0
$$

Lemme 3.2.1 *With this notations, we have*

$$
E(\varepsilon_n^2|\mathcal{G}_{n-1}) = F(q_{n-1})[1-F(q_{n-1})] + h_nf(q_{n-1})\left(\int K^2(z)dz - 1\right) + O(h_n^2).
$$

Proof 3.2.1 *see [22]. We have*

$$EK^2\left(\frac{x-X}{h_n}\right) = h_n\int_1^x f(x-h_nu)du + h_n\int_{-1}^1 K^2(u)f(x-h_nu)du$$
$$= F(x) + h_nf(x)\int_{-1}^1 (K^2(u)-K(u))du + O(h_n^2)$$

$\Longrightarrow$

$$Var(\varepsilon) = \frac{1}{n}\left[F(x)(1-F(x)) + h_nf(x)\left(\int_{-1}^1 K^2(u)du - 1\right) + O(h_n^2)\right] \quad (34)$$

Remarque 3.2.3 *The algorithm defined by (3.2.5) is then an algorithm of search of zero of the function $h : z \to h(z) = \delta - F(z)$. In practice, F in (3.2.4) is unknown can only be estimated from the data.*

3.3 Main results

Assumptions and notations

$\mathcal{G} = (\mathcal{G}_n)$ denotes the σ-fields spanned by $(X_1, \ldots, X_n)$, and q^* the zero (which we assume unique) of $h(z)$. Let $s_n = \sum_{i=1}^{n} \gamma_n$ and $v_n = {\gamma_n}^{-1}$.

(H3) $\sum \gamma_n = \infty$, $\lim_{n\to\infty} \gamma_n = 0$, $\sum \gamma_n^2 < \infty$.

(H4) $(\delta - h(q_n))\,(q_n - q^*) \neq 0$ for all $q_n \neq q^*$

(H5) h is differentiable at q^*, its derivative H at q^* is symmetric, and there exists a neighborhood of q^* in which $h(z) = H(z-q^*) + O(|z-q^*|^2)$.

(H6) i) $\lim_{n\to\infty} \gamma_n \log s_n = 0$.

ii) There exist $\xi \in [0, 2L[$ such that $\lim_{n\to\infty} \frac{1}{\gamma_n}\left[1 - \frac{v_{n-1}}{v_n}\right] = \xi$, $L = -H$.

(H7) There exists a sequence w_n such that $\lim_{n\to\infty} w_n^2\gamma_n \log s_n = 0$ a.s and such $\|\varepsilon_{n+1}\| \leq w_n$.

The condition "H is symmetric" which appears in (H5) is not generally required for the study of the rate of convergence of stochastic algorithms ; we add here this condition (which is satisfied by the algorithm (3.2.5)) to simplify the demonstrations.

The assumption $\sum \gamma_n^2 < \infty$ ensures that $\sum \gamma_n \varepsilon_n$ converges in L^2 and $a.s.$ for many stochastic models of random noise. Under certain regularity conditions, this in turn implies that $q_n - q^*$ converges in L^2 and a.s., and the assumption $\sum \gamma_n = \infty$ then ensures that the limit of $q_n - q^*$ is 0 ([26], [12] and [10]). Since F is twice differentiable and $\|F^{''}\|_\infty < \infty$, a Taylor expansion gives :

$$\lim_{n\to\infty} \frac{1}{h_n^2} r_n = r$$

with

$$r = -\frac{1}{2} F^{''}(q) \left[\int_R x^2 k(x) dx \right]. \tag{35}$$

We first establish the consistency of q_n.

3.3.1 Strong consistency

To prove consistence of q_n, we apply theorem (3.2.1).

Proposition 3.3.1 *Let q_n be the sequence obtained from algorithm (3.2.5). Under assumptions (H_1)-(H_4),*

$$\lim_{n\to\infty} q_n = q^* \quad a.s.$$

Proof 3.3.1 *The application of Robbins-Monro theorem (3.2.1) with*

$$V \;:\; z \mapsto |z - q^*|^2$$

to algorithm defined by (3.2.5) ensure that the sequence ($|q_n - q^*|$) converges almost surely and that, see (3.2.1) :

$$\sum \gamma_n |\, (q_n - q^*)\, (\delta - K(q_n))| < \infty \quad a.s$$

Since $\sum_n \gamma_n = \infty$ and $< \delta - K(q_n), q_n - q^* > \neq 0$ for all $q_n \neq q^*$, this implies that $(q_n - q^*)$ converges almost surely to zero.

3.3.2 Upper bound of the strong convergence of q_n

Proposition 3.3.2 *Under the assumptions* (H_1)-(H_7),

$$|q_n - q^*| = O\left(\sqrt{\gamma_n \log s_n}\right) \qquad a.s.$$

Remarque 3.3.1 *Note that the rate of convergence of the general algorithm has been studied under the assumption :* $\exists\, b > 2$ *such that*

$$\sup_n E[|\varepsilon_{n+1}|^b \, |\mathcal{G}_n] < \infty \tag{36}$$

Let (L_n) and (Δ_n) be two sequences defined by :

$$L_{n+1} = e^{s_n H} \sum_{k=1}^{n} e^{-s_k H} c_k \varepsilon_{k+1}, \tag{37}$$

$$\Delta_{n+1} = (q_{n+1} - q^*) - L_{n+1}. \tag{38}$$

To establish proposition 3.3.2, we first prove that the sequence (L_n) satisfies an CLT.a.s. Then, we suppose that Δ_n the difference between (L_n) and $(q_n - q^*)$ is "small enough" so that, the sequence $(q_n - q^*)$ fulfills the same CLT as (L_n). We suppose then that the asymptotic behavior of q_n is dominated by the behavior of (L_n), the term (Δ_n) is negligible. The proposition 3.3.2 is a direct consequence of the lemma below.

Lemme 3.3.1 *Under assumptions (H1)-(H7), we have*

$$|L_n| = O\left(\sqrt{\gamma_n \log s_n}\ \right) \quad a.s.$$

Since K is bounded on $\mathbb{R}$, we have $\|\varepsilon_n\| \le 2\|K\|_\infty$ and the assumption H_6 is hold true with $w_n = 2\|K\|_\infty$, using H_6i).

3.3.3 Weak convergence rate of q_n

Théorème 3.3.1 *Suppose the assumptions (H1)-(H7) holds, then we have :*

$$\sqrt{\gamma_n^{-1}}(q_n - q^*) \to N(0, \delta(1-\delta))$$

In the theorem, the asymptotic behavior of q_n is dominated by the behavior of (L_n), the term (Δ_n) is negligible. The weak convergence rate of $F_n(q)$ to zero is governed by the weak convergence rate of the variance term $F_n(q) - E(F_n(q))$ on one hand and by the deterministic convergence rate of the bias term on the other hand.

Since the variance term converges at the rate $\sqrt{\gamma_n \log s_n}$, the condition $\lim_{n\to\infty} \gamma_n \log s_n = 0$ is necessary to make the bias term negligible in front of the variance term, and thus to establish a central limit theorem.

The theorem 3.3.1 is a straightforward consequence of the application of following Lemma.

Lemme 3.3.2 *Under the assumptions (H1)-(H7),*

$$\sqrt{(\gamma_n)^{-1}} L_n \to N(0, \delta(1-\delta)).$$

Remarque 3.3.2 *Note that the choice of the step $(\gamma_n) \equiv (\gamma_0 n^{-1})$, which makes it possible to obtain the optimal convergence rate, induces - as is very often the case in the context of stochastic algorithms, a very inconvenient condition on the parameter γ_0, this condition involving the unknown parameter.*

We can introduce the principle of averaging stochastic algorithms to build recursive estimators that converge at the optimal rate, but for which no condition on the parameter γ_0 is required.

3.4 On the averaged version of the stochastic approximation algorithm

First, consider the stochastic approximation algorithm

$$Z_{n+1} = Z_n + \gamma_n[h(Z_n) + r_{n+1}] + \frac{\gamma_n}{b_n^\beta}\varepsilon_{n+1} \tag{3.4.1}$$

The principle of averaging stochastic algorithms was introduced by Polyak (1990) and Ruppert (1985) for the Robbins-Monro algorithm. The principle of averaging stochastic algorithms makes it possible to construct an algorithm that converges at the same rate as the algorithm (3.4.1) when for the latter one chooses the optimal step $(\gamma_n) = (\gamma_0 n^{-\alpha})$; the advantage of the averaged algorithm is that the additional condition relating to the parameter γ_0 and involving the unknown parameter L is no longer required.

The principle is to proceed in two steps, consists in :
(i) running the approximation algorithm (3.4.1) by using a slower stepsize ;

(ii) computing an average of the approximations obtained in (i). To find out what is the correct mean of the Z_k to be computed, a positive deterministic sequence (ϕ_n) is introduced such that $\sum \phi_n \gamma_n = \infty$ and let

$$\overline{Z}_n = \frac{1}{\sum\limits_{k=1}^{n} \phi_k \gamma_k} \sum_{k=1}^{n} \phi_k \gamma_k Z_k \tag{39}$$

(The choice $(\phi_n) = (\gamma_n^{-1})$ corresponding to the case where $\overline{Z}_n$ is the arithmetic mean of Z_k). The choice of the averaging of Z_k (ie the choice of the sequence (ϕ_n)) does not influence the order of the speed, this one being always the optimal speed $\sqrt{n^{1-2\beta\tau}}$(the window (h_n) being fixed) ; it only affects the value of the asymptotic variance of $(\overline{Z}_n - z^*)$.

For this to be the optimal variance, we must choose $\phi = \alpha - 2\beta\tau$ since only this choice ensures that the convergence of $(\overline{Z}_n)$.

Note that the value of ϕ_0 does not intervene ; this choice of the sequence (ϕ_n) thus amounts to taking $(\phi_n) = (\gamma_n^{-1} b_n^{2\beta})$.

In order to get the asymptotic efficiency of the averaged version of the stochastic approximation algorithm (3.4.1), the average of the Z_k must be weighted by the c_k^2. We set

$$\overline{Z}_n = \frac{1}{\sum_{k=1}^{n} b_k^{2\beta}} \sum_{k=1}^{n} b_k^{2\beta} Z_k$$

It was establish (see [10]) in particular that, in order to get the asymptotic efficiency of the averaged version of the stochastic approximation algorithm (3.2.2), the average of the Z_k must be weighted by the c_k^2. We set

$$\overline{Z}_n = \frac{1}{\sum_{k=1}^{n} c_k^2} \sum_{k=1}^{n} c_k^2 Z_k \tag{40}$$

The aim of this part is to state the different properties, which will enable us to establish the asymptotic behaviour of the algorithms (3.2.2). To this end, we consider the assumptions (A1)-(A4)

(A1) $\lim_{n\to 1} Z_n = z^*$ $a.s.$

(A2) (i) h is differentiable at q^*, its derivative H at q^* is symmetric, and there exists a neighborhood of z^* in which $h(z) = H(z - z^*) + O(|z - z^*|^2)$.

(ii) H is diagonalizable and its largest eigenvalue $-L$ is negative.

(A3) Either $(c_n) \in GS(-\tau)$ with $\tau \in (0, 1/2)$ or $(c_n) = 1$, in which case we set $\tau = 0$.

(A4) (i) $E(\varepsilon_{n+1}|\mathcal{F}_n) = 0$.

(ii) There exists a nonrandom, positive de nite matrix Γ such that

$$\lim_{n\to\infty} E(\varepsilon_{n+1}\varepsilon_{n+1}^t|\mathcal{F}_n) = \Gamma \quad a.s.$$

The asymptotic behaviour of $(\overline{Z}_n)$ is given by those of the sequences(M_n) and (R_n) defined by

$$M_{n+1} = \frac{-1}{\sum_{k=1}^{n} c_k^2} H^{-1} \sum_{k=1}^{n} c_k^2 c_{k+1}^{-1} \varepsilon_{k+1} \tag{41}$$

$$R_{n+1} = (\overline{Z}_n - z^*) - M_{n+1} \tag{42}$$

We have the following lemma.

Lemme 3.4.1 *Let (A2)-(A4) hold. Moreover, assume that there exists $m > 2$ such that*

$$\lim_{n\to 1} n^{1-m/2} E[\|\varepsilon_{n+1}\|^m |F_n] = 0 a.s.$$

Then, we have

$$\sqrt{nc_n^2} M_{n+1} \to N(0,(1-2\tau)H^{-1}\Gamma[H^{-1}]^t)$$

The Lemmas (3.4.1) gives the weak convergence rate of $(\overline{Z}_n)$ under assumptions (A1)-(A4) : If $\lim_{n\to 1} \gamma_n c_n^2 \gamma_n^2 = 0$, then

$$\sqrt{nc_n^2}(\overline{Z}_n - z^*) \to N(0,(1-2\tau)H^{-1}\Gamma[H^{-1}]^t)$$

To apply the averaging principle to the approximating algorithms (2.2.2), we proceed as follows. First, we run the algorithms (2.2.2) with a slower stepsize satisfying $\lim_{n\to\infty} n\gamma_n = \infty$. Then, we define the average $\overline{q}_n$ of the q_k by setting

$$\overline{q}_n = \frac{1}{n}\sum_{k=1}^{n} q_k \qquad (3.4.2)$$

We establish the weak convergence rate of $(\overline{q}_n)$.

We adopt an approach that involves introducing the principle of averaging stochastic algorithms to the problem of quantile estimation. The following theorem, giving the weak convergence rate of the averaged algorithms (3.2.5), shows that $(\overline{q}_n)$ can be asymptotically efficient, and this without any tedious condition on the stepsize (γ_n).

Théorème 3.4.1 *Let (q_n) be defined by the stochastic approximation algorithms (3.2.5), let $\overline{q}_n$ be the averaged algorithms defined by (3.4.2), and assume that (H1)-(H7) hold, then*

$$\sqrt{n}(\overline{q}_n - q) \to N(0,\delta(1-\delta))$$

3.5 Proofs

3.5.1 Proof of lemma 3.3.1

Let M_n be the martingale defined as

$$\begin{aligned} M_{n+1} &= \sum_{k=1}^{n} e^{-s_k H} c_k \varepsilon_{k+1} \\ &= \sum_{k=1}^{n} e^{L s_k} \sqrt{\frac{\gamma_k}{v_k}} \varepsilon_{k+1}. \end{aligned}$$

We obtain the predictable quadratic variation of (M_n) (see Duflo [12]) :

$$< M >_n = \sum_{k=1}^{n} e^{2Ls_k} \frac{\gamma_k}{v_k} E\left[|\varepsilon_k|^2 | \mathcal{G}_{k-1} \right] \qquad (43)$$

and the application of Lemma 4 in [23] ensures that

$$\lim_{n\to\infty} v_n e^{-2Ls_n} < M >_n = \Gamma \quad a.s. \tag{3.5.1}$$

where $\Gamma = \lim_{n\to\infty} E\left[|\varepsilon_n|^2 | \mathcal{G}_{n-1} \right]$ (it is assumed that it exists). From (3.5.1) and in view of assumptions (H3), we deduce that

$$\lim_{n\to\infty} < M >_n = \infty \quad a.s. \tag{3.5.2}$$

Noting that the assumption (H6) implies

$$\begin{aligned} \frac{v_{n-1}}{v_n} &= 1 - \xi\gamma_n + o(\gamma_n) \\ \frac{v_n}{v_{n-1}} &= 1 + \xi\gamma_n + o(\gamma_n) \end{aligned}$$

Or $v_n = v_0 \prod_{k=1}^{n} \frac{v_k}{v_{k-1}}$, we obtain

$$\begin{aligned} \log v_n &= \log v_0 + \sum_{k=1}^{n} \log \frac{v_k}{v_{k-1}} \\ &= \log v_0 + \sum_{k=1}^{n} \log\left(1 + \xi\gamma_k + o(\gamma_k)\right) \\ &= \xi s_n + o(s_n). \end{aligned} \tag{3.5.3}$$

And

$$\begin{aligned}\log\left[v_n \exp(-2Ls_n)\right] &= \xi s_n + o(s_n) - 2Ls_n \\ &= (\xi - 2L + o(1))s_n.\end{aligned}$$

Since $L > \frac{\xi}{2}$, and $\lim_{n\to\infty} s_n = \infty$, this implies that

$$\lim_{n\to\infty} \log\left[v_n \exp(-2Ls_n)\right] = -\infty$$

i.e.

$$\lim_{n\to\infty} v_n \exp(-2Ls_n) = 0,$$

this prove lemma 3.3.1.

Let η be a function defined by $\eta(x) = \sqrt{2x \log\log x}$; it comes from (H7) and (3.5.1)

$$\begin{aligned}\frac{|M_{n+1} - M_n|}{< M >_n \left[\eta(< M >_n)\right]^{-1}} &= \frac{e^{Hs_n}\sqrt{\gamma_n v_n^{-1}}\,\|\varepsilon_{n+1}\|\,\sqrt{2\log\log < M >_n}}{\sqrt{< M >_n}} \\ &\leq C_n \frac{e^{Hs_n}\sqrt{\gamma_n v_n^{-1}} w_n}{\sqrt{2e^{2Hs_n} v_n^{-1}\Gamma}} \sqrt{\log\log(e^{2Hs_n} v_n^{-1}\Gamma)} \\ &\leq C_n^{'} \sqrt{\gamma_n} w_n \sqrt{\log s_n}\end{aligned}$$

where (C_n) et$(C_n^{'})$ are two nonrandom sequences, bounded a.s. The assumption (H7) ensures existence of a nonrandom sequence $(\widetilde{C}_n)$ that goes to zero as n goes to ∞ and

$$|M_{n+1} - M_n| \leq \widetilde{C}_n < M >_n \left[\eta(< M >_n)\right]^{-1}$$

The application of theorem 6.4.24 in Duflo [12] then ensures

$$\limsup_{n\to\infty} \frac{|M_n|}{\eta(< M >_n)} \leq 1 \quad a.s. \tag{44}$$

With (3.5.1), we have

$$\begin{aligned} |M_n| &= O\left(e^{Ls_n} v_n^{-1/2} \sqrt{\log\log(e^{2Ls_n} v_n^{-1})}\right) \quad a.s. \\ &= O\left(\frac{e^{Ls_n}\sqrt{\log s_n}}{\sqrt{v_n}}\right) \quad a.s. \end{aligned}$$

$L_{n+1} = e^{-Ls_n} M_{n+1}$, we have then

$$|L_n| = O\left(\sqrt{\frac{\log s_n}{v_n}}\right) \quad a.s. \tag{45}$$

3.5.2 Proof of lemma 3.3.2

Let

$$M_j^{(n)} = \sqrt{v_n} e^{s_n H} \sum_{k=1}^{j} e^{-s_k H} \gamma_k \varepsilon_k.$$

For n fixed, $M^{(n)} = (M_j^{(n)})_{j \geq 1}$ is a martingale we applied the CLT for the martingales (*see* Duflo [12]). We obtain

$$M_n^{(n)} = \sqrt{v_n}\; L_n. \tag{46}$$

We have

$$\begin{aligned} < M >_n^{(n)} &= v_n e^{2s_n H} \sum_{k=1}^{n} E\left[e^{-2s_k H \gamma_k^2} |\varepsilon_k|^2 | \mathcal{G}_{k-1}\right] \\ &= v_n e^{2s_n H} \sum_{k=1}^{n} \frac{\gamma_k}{v_k} e^{-2s_k H} E\left[|\varepsilon_k|^2 | \mathcal{G}_{k-1}\right]. \end{aligned}$$

Then we obtain :

$$< M >_n^{(n)} = v_n e^{2s_n H} \left\{ \sum_{k=1}^{n} \frac{\gamma_k}{v_k} e^{-2s_k H} E\left[|\varepsilon_k|^2 | \mathcal{G}_{k-1}\right] \right\} \tag{47}$$

We have

$$\lim_{k\to\infty} E\left[|\varepsilon_k|^2|\mathcal{G}_{k-1}\right] = \delta(1-\delta)$$

We obtain, with application of lemma 4 of Mokkadem and Pelletier [23],

$$\lim_{n\to\infty} < M >_n^{(n)} = \delta(1-\delta)$$

We check that $(M_j^{(n)})_{j\geq 1}$ satisfy the Lindeberg condition.

• Let $b > 2$; Since K is bounded on $\mathbb{R}$, we have

$$\begin{aligned}
\sum_{k=1}^{n} E\left[\left\|\sqrt{v_n}e^{s_nH}e^{-s_kH}c_k\varepsilon_k\right\|^b|\mathcal{G}_{k-1}\right] &= O\left(\gamma_n^{-b/2}e^{-bLs_n}\sum_{k=1}^{n}e^{bLs_k}\gamma_k^b E\left[\|\varepsilon_k\|^b|\mathcal{G}_{k-1}\right]\right)\\
&= O\left(\gamma_n^{-b/2}e^{-bLs_n}\sum_{k=1}^{n}e^{bLs_k}\gamma_k^b h_k\right)\\
&= O\left(e^{-bLs_n}\sum_{k=1}^{n}e^{bLs_k}h_k^{b/2}\gamma_k\right)\\
&= o(1) \qquad (3.5.4)
\end{aligned}$$

the lemma 3.3.2 yields.

Some graphic representations The outputs on R which simulates an exponential sample and the median by changing the n and U : an estimate by the kernel method, taking the Gaussian kernel, and a comparison by the MSE

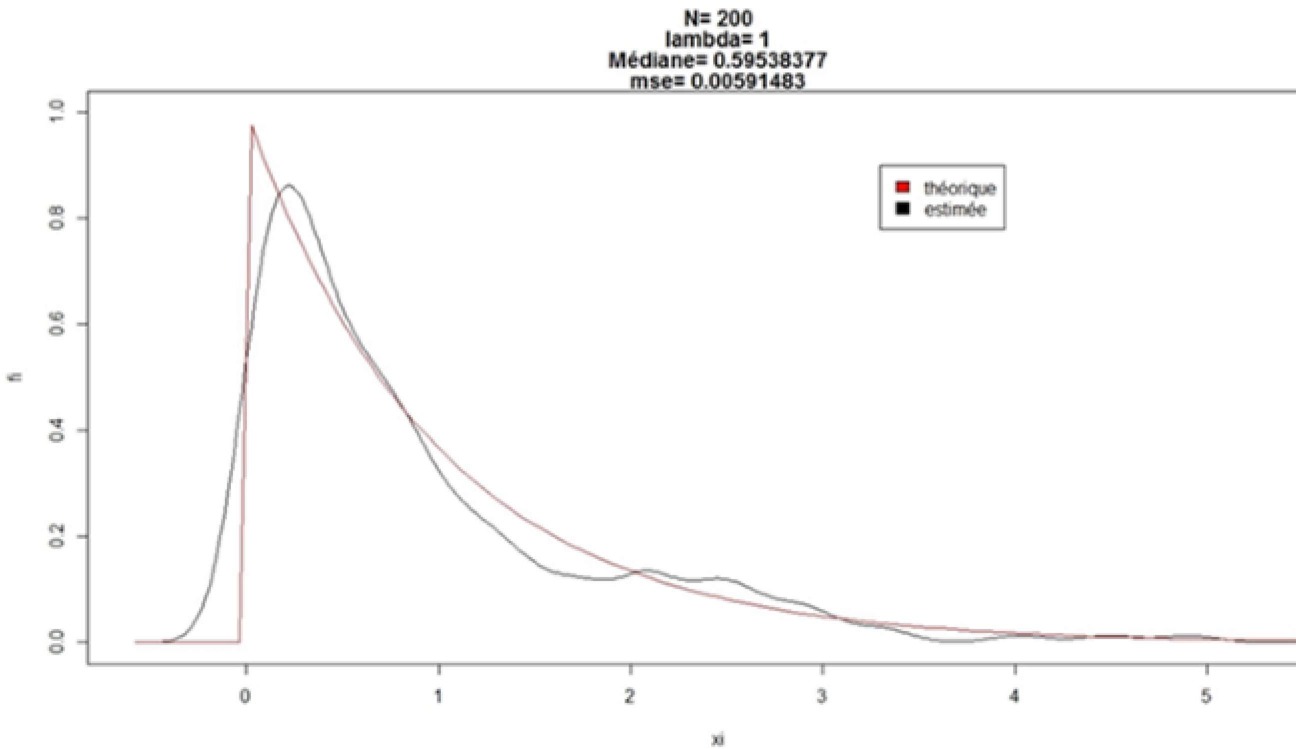

Figure 3.1 – nag, fig1

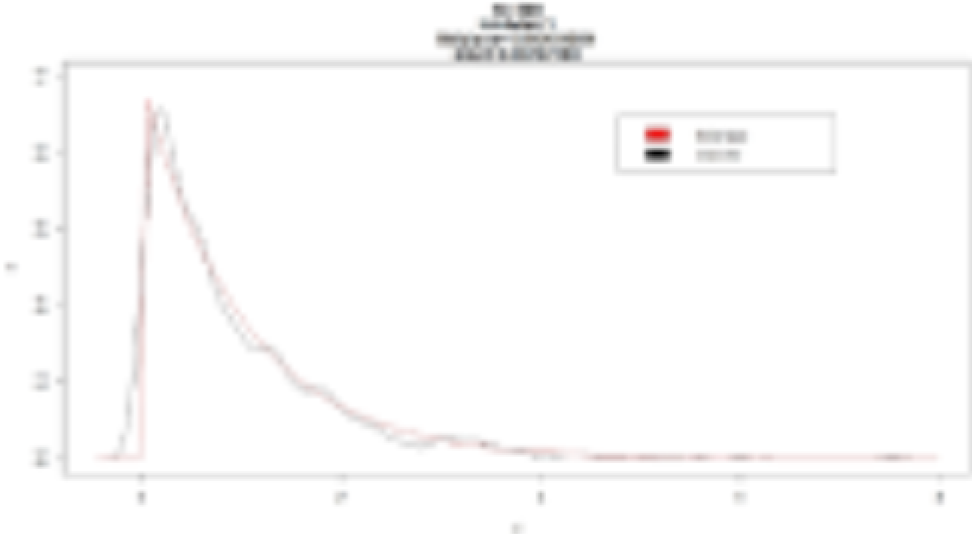

FIGURE 3.2 – nag, fig2

Chapitre 4

Iterative Solution

4.1 Introduction

Poisson's law is a probability distribution that applies to rare events. Among the areas of application, we can mention : quality controls (if we assume that errors are rare), the probabilities of credit default... The Poisson distribution is a discrete distribution that measures the probability of a given number of events happening in a specified time period. In finance, the Poisson distribution could be used to model the arrival of new buy or sell orders entered into the market or the expected arrival of orders at specified trading venues or dark pools. In these cases, the Poisson distribution is used to provide expectations surrounding confidence bounds around the expected order arrival rates. Poisson distributions are very useful for smart order routers and algorithmic trading. The Poisson distribution arises from either of two models. In one model, quantities, for example, bacteria are assumed to be distributed at random in some medium with a uniform density of λ per unit area. The number of bacteria colonies found in a sample area of size A follows the Poisson distribution with a parameter μ equal to the product of λ and A.

In terms of the model over time, we assume that the probability of one event in a short interval of length t_1 is proportional to t_1, that is, Pexactly one event is approximately λt_1. Another assumption is that t_1 is so short that the probability of more than one event during this interval is almost zero. We also assume that what happens in one time interval is independent of the happenings in another interval. Finally, we assume that λ is constant over time. Given these assumptions, the number of occurrences of the event in a time interval of length t follows the Poisson distribution with parameter μ, where μ is the product of λ and t.

Many situations are related to the study of the realization of an event in a given interval of time (appearances of failures of a computer network in a year, arrival of patients to the emergency of a hospital in a night,...). The phenomena thus studied are phenomena of waiting. To describe the achievements in time of a given event, one can look for the number of achievements of the event in a given time interval that is distributed according to a Poisson law. The Poisson distribution is the discrete probability distribution of the number of events occurring in a given time period, given the average number of times the event occurs over that time period.

The Poisson distribution, like the binomial, is a counted number of times something happens. The difference is that there is no specified number n of possible tries. Here is one way that it can arise. If an event happens independently and randomly over time and the mean rate of occurrence is constant over time, then the number of occurrences in a fixed amount of time will follow the Poisson distribution.The Poisson is a discrete distribution (because you can list the possibilities as 0, 1, 2, 3, ...) and depends only on the mean number of occurrences expected.

Here are some random variables that might follow a Poisson distribution :

1.The number of orders your firm receives tomorrow.

2.The number of people who apply for a job tomorrow to your human resources division.

3.The number of defects in a finished product.

4.The number of calls your firm receives next week for help concerning an “easy-to-assemble” toy.

5.A binomial number X when n is large and p is small.

The basic Poisson model can be written as :

$$P(Y_i = j) = \frac{\exp(-\lambda) \times \lambda^j}{j!} \tag{4.1.1}$$

where there are $i = 1, 2, ..., n$ observations, Y_i is the i^{th} observation of the variable, $j = 0, 1, 2, ...$ are the possible values of Y_i (ie a set of natural numbers), and λ is the Poisson parameter.

The Poisson distribution is constructed with a single parameter λ, which is both the expectation and the variance of the law. The mean and variance are both equal to λ;

$$E(Y) = Var(Y) = \lambda.$$

In fact, this property characterizes the Poisson distribution. A related property is that the cumulant generating function of a Poisson random variable is given by

$$K(t) = \log\left((M_Y(t)\right) = \lambda(\exp(t) - 1)$$

where $M_Y(t)$ is the moment generating function of Y. This can be proved by simple calculations, but for this presentation, it is instructive to consider the Poisson distribution as a member of the natural exponential family of distributions.

Let $f(x;\theta)$ denote the density function of the natural exponential family with parameter θ, i.e.,

$$f(x;\theta) = h(x)\exp(\theta x - b(\theta))$$

with $x \in A$ $\theta = \log\lambda$, $b(\theta) = \exp(\theta)$ and $h(x) = 1/x!$. h, b are known functions and A is a subset of $\mathbb{R}$. Estimation and inference are based on the maximum likelihood theory,this topic has been described in several texts.

The zero-truncated Poisson distribution is a useful model for integer-valued random phenomena when the value of 0 is unobservable or simply out of domain. These distributions are found in many applications. Several examples have been given employing the truncated distribution in fitting rainfall data.

Some zero-truncated models have been proposed already. In Cohen [6], the ML estimation of a zero-truncated Poisson model is used as estimation of the Poisson parameter of more general zero-modified Poisson distributions. They simply ignore the zero observations in the sample. The latter ones are only used to estimate the additional model parameter in a second stage. Cohen [5] calls it conditional Poisson distribution. The truncated distribution was first examined by David and Johnson [8]. In particular, they derived maximum likelihood estimation (MLE) from and its asymptotic variance, and discussed the efficiency of moment estimation. In David and Johnson [8] the ML estimation of zero-truncated Poisson distributions is also considered. We used the probability theory definitions , drawing on the approach adopted by Johnson and Kotz [16].

Niyomdecha and Srisuradetchai [25] proposes a new continuous distribution called the complementary gamma zero-truncated Poisson distribution, which combines the distribution of the maximum of a series of independently identical gamma-distributed random variables with zero-truncated Poisson random variables. The unknown parameters are estimated using the maximum likelihood method, whose asymptotic properties are examined. Plackett [27] provided a similar estimate of the parameter of zero-truncated Poisson distribution, to show that it is highly efficient, and to estimate its sampling variance.

4.2 Parameter Estimation of the Poisson distribution truncated at 0

The zero-truncated Poisson (ZTP) distribution is a certain discrete probability distribution whose support is the set of positive integers. This distribution is also known as the conditional Poisson distribution or the positive Poisson distribution. It is the conditional probability distribution of a Poisson-distributed random variable, given that the value of the random variable is not zero. Thus it is impossible for a ZTP random variable to be zero. Consider for example the random variable of the number of items in a shopper's basket at a supermarket checkout line. Presumably a shopper does not stand in line with nothing to buy (i.e., the minimum purchase is 1 item), so this phenomenon may follow a ZTP distribution.

4.2.1 The Poisson distribution truncated at 0

We recall here the estimation of parameter λ of the Poisson distribution truncated at 0. In some situations, counts that are zero are not recorded in the data, and so fitting a Poisson distribution is not straightforward because of those missing zeros. We will fit out them with a distribution which is identical to a Poisson law on positive integers, but which has no probability at zero.

In the marketing literature, modeling of customer response behavior to price and promotion effects is widely accepted. The consumer quantity decision to a given purchase impact can be modeled as a stochastic variable, after a null truncated fish distribution. Typically, the Poisson parameter in these models is estimated based on client-specific variables and marketing variables.

This way, we can estimate the Poisson parameter while accounting for missing zeros. A law that is truncated is a conditional law that is derived from another probability law. More clearly, for a random variable X, with a Poisson distribution, the zero-truncated distribution is merely the conditional distribution :

$$\begin{aligned} P(R=r|R>0)) &= \frac{P(R=r)}{P(R>0)} \\ &= \frac{\lambda^r}{r!}\frac{1}{e^\lambda - 1} \end{aligned} \quad (48)$$

The density function of a zero-truncated Poisson variable is given by (Johnson and Kotz [16])). The difference with the standard Poisson distribution lies in the correction factor $(e^\lambda - 1)^{-1}$, which reflects the fact that a value of 0 cannot occur.

Proposition 4.2.1 *A random variable R, with a zero-truncated Poisson distribution possesses the first moments :*

$$E(R) = \frac{\lambda}{1-e^{-\lambda}}$$

and

$$Var(R) = \frac{\lambda(1-(1+\lambda)e^{-\lambda})}{(1-e^{-\lambda})^2}$$

Proof 4.2.1 *of Proposition 4.2.1*

1. *The expectation of R*

$$\begin{aligned} E(R) &= \sum_{r=1}^{\infty} \frac{r\lambda^r}{r!}\frac{1}{e^\lambda - 1} \\ &= \frac{1}{e^\lambda - 1}\lambda\sum_{r=1}^{\infty}\frac{\lambda^{r-1}}{(r-1)!} \\ &= \frac{\lambda}{1-e^{-\lambda}} \end{aligned}$$

2. *The variance of R :*

$$E(R^2) = \sum_{r=1}^{\infty} \frac{r^2\lambda^r}{r!}\frac{1}{e^\lambda - 1}$$

$$
\begin{aligned}
&= \frac{\lambda}{e^{\lambda}-1}\left[\sum_{r=1}^{\infty}\frac{r\lambda^{r-1}}{(r-1)!}\right] \\
&= \frac{\lambda}{e^{\lambda}-1}\sum_{r=1}^{\infty}\frac{d}{d\lambda}\left[\frac{\lambda^{r}}{(r-1)!}\right] \\
&= \frac{\lambda}{e^{\lambda}-1}\frac{d}{d\lambda}\left[\lambda\sum_{r=1}^{\infty}\frac{\lambda^{r-1}}{(r-1)!}\right] \\
&= \frac{\lambda}{e^{\lambda}-1}\frac{d}{d\lambda}\left[\lambda e^{\lambda}\right] \\
&= \frac{\lambda(\lambda+1)e^{\lambda}}{e^{\lambda}-1} \qquad (49)
\end{aligned}
$$

We can deduce

$$
\begin{aligned}
Var(R) &= \left[\frac{\lambda(\lambda+1)}{1-e^{-\lambda}}\right]-\left[\frac{\lambda}{1-e^{-\lambda}}\right]^{2} \\
&= \frac{\lambda(1-(1+\lambda)e^{-\lambda})}{(1-e^{-\lambda})^{2}} \qquad (50)
\end{aligned}
$$

Remark : the Poisson cumulative distribution function does not have a simple form, though it can be easily calculated.

In order for the Poisson distribution to be accurate, all events are independent of each other, the rate of events through time is constant, and events cannot occur simultaneously. Moreover, the mean and the variance will be equal to one another.The Poisson distribution can be a helpful tool to evaluate and predict financial and trading operations.

4.2.2 Maximum likelihood estimate

Maximum likelihood estimation is a method of point estimation often used in statistics. Consider a sample of n independent observations $X_1, X_2, \ldots, X_n$, each with the probability density function $f(x)$.

Proposition 4.2.2 *The maximum likelihood estimator is the root of the equation*

(4.2.1)
$$\frac{\lambda}{1-e^{-\lambda}} = \overline{x}$$

where $\overline{x}$ is the empirical mean of the observations.

Proposition 4.2.2 Note that truncation here means that 0-data is removed. We have

$$L(x,\lambda) = \sum_{i=1}^{n} \log(f(x_i,\lambda))$$

with

$$f(x_i,\lambda) = \frac{\lambda^{x_i}}{x_i!} \times \frac{1}{e^{\lambda}-1}$$

Then we have

$$L(x,\lambda) = \log(\lambda)\sum_{i=1}^{n} x_i - \sum_{i=1}^{n}\log(x_i!) - n\log(e^{\lambda}-1) \tag{51}$$

from where

$$\frac{dL(x,\lambda)}{d\lambda} = \frac{1}{\lambda}\sum_{i=1}^{n} x_i - n\frac{e^{\lambda}}{(e^{\lambda}-1)}$$

Then we deduce (4.2.1).

According to [8], the maximum likelihood estimate in the truncated case is difficult to find explicitly, but we can get an expression of this estimate. The r^th term of the truncated distribution is

$$p_r \;= \tfrac{\lambda^r}{1-e^{-\lambda}} \times \tfrac{e^{-\lambda}}{r!} \qquad\qquad r = 1, 2, 3, ...$$

Whence the joint probability function is

$$p = \frac{1}{(1-e^{-\lambda})^n}\lambda^{n\overline{x}}e^{-\lambda n}.$$

where $\overline{x}$ is the truncated sample mean and n is the number in the truncated sample. Equating the differential of $L = \log(p)$ to zero we have

(4.2.2) $$\frac{\partial L}{\partial \lambda} = -\frac{ne^{-\lambda}}{1-e^{-\lambda}} + \frac{n\overline{x}}{\lambda} - n$$

$\Rightarrow$

(4.2.3) $$\overline{x} = \widehat{\lambda}(1-e^{-\widehat{\lambda}})^{-1}$$

from which it does not seem possible to obtain an explicit expression for $\widehat{\lambda}$.

A technique to find a value of λ, which approximately solves this equation, is to use an iterative method. One way to find the roots of a function is the Newton-Raphson method or a modified version that is Robin-Monro's algorithm [31].

It is noted that the parameter estimate is lower than the sample average. This is quite normal, because the maximum likelihood counts the missing zeros, which are not present in the data.

Corollaire 4.2.1 *We deduce*

$$E\hat{\lambda} = \lambda$$

and

$$Var(\hat{\lambda}) = \frac{\lambda}{n} \times \frac{(1-e^{-\lambda})^2}{1-(\lambda+1)e^{-\lambda}}$$

Plackett [27] presents a very straightforward estimator which is unbiased and is extremely simple to calculate. The estimate is not fully efficient, but the loss in the efficiency is offset by the gain in simplicity.

Proposition 4.2.3 *(without demonstration) The maximum likelihood estimator can be expressed by*

$$\begin{aligned} \hat{\lambda} &= \overline{x} + \sum_{r=1}^{\infty} \frac{(-1)^r \overline{x}^r}{r!} \frac{d^{r-1}}{d\overline{x}^{r-1}} e^{-r\overline{x}} \\ &= \overline{x} - \sum_{r=1}^{\infty} \frac{r^{r-1}}{r!} (\overline{x} e^{-\overline{x}})^r \end{aligned} \tag{52}$$

The proof can be found in Kendall [19].

The sum of identical independent zero-truncated Poisson distributed random variables We consider the case of a random variable X which can be expressed as the sum of m independent identical zero-truncated Poisson distributed random variables X_i ; $i = 1, ..., m$:

$$X = \sum_{i=1}^{m} X_i$$

In order to get insight into the distribution of such a random variable, we first construct the probability function :

$$P\left(\sum_{i=1}^{m} X_i = n\right)$$

More specifically, we have the following theorem :

Théorème 4.2.1 *Let Xi, $i = 1, ..., m$ be m independent identical zero-truncated Poisson distributed random variables and let X be the random variable defined as $\sum_{i=1}^{m} X_i$. The probability function for X is given by*

$$\begin{aligned} P(X = n) &= \frac{\lambda^n}{n!(e^\lambda - 1)^m} \sum_{j=0}^{k} (-1)^k (m-k)^n \binom{m}{k} \quad m \leq n \\ &= 0 \quad m \geq n \end{aligned}$$

Since the m variables X_i are independent, the mean and variance as well as the moment generating function of $X = Pmi = 1Xi$ can easily be derived. The mean is given by :

$$E(X) = \sum_{i=1}^{m} E(X_i) = \frac{m\lambda e^\lambda}{e^\lambda - 1} \tag{53}$$

and the variance by :

$$Var(X) = \sum_{i=1}^{m} Var(X_i) = \frac{m\lambda e^\lambda}{e^\lambda - 1} \times \left(1 - \frac{\lambda}{e^\lambda - 1}\right) \tag{54}$$

while the moment generating function is given by :

$$M_X(t) = \prod_{i=1}^{m} M_{X_i}(t) = \frac{e^{\lambda e^t} - 1}{e^\lambda - 1} \tag{55}$$

We have derived expressions for the density function and cumulative distribution function of the sum of m independent.

4.3 Parameter estimation of the Poisson distribution truncated at a

The a-truncated Poisson distribution is the distribution of a Poisson random variable Y conditional on the event $Y > k$. It has one parameter, which we may take to be $\lambda = E(Y)$. Since λ is not the mean (or anything else simple)of the distribution of Y conditioned on the even $Y > k$, we do not call λ the mean,rather we call it the original parameter.

4.3.1 The Poisson distribution truncated at a

This part is denoted to the study of used discrete distribution with emphasis on their behavior under truncation. We works through the details of the a-truncated Poisson distribution, a special case of which is the zero-truncated Poisson distribution. The a-truncated Poisson distribution is the distribution of a Poisson random variable Y conditional on the event $Y > a$. Here, a is the cutoff value such that only values strictly larger than a are allowed, i.e. the probability mass function is ([38])

$$p_j = \frac{s_a^{-1}\lambda^j e^{-\lambda}}{j!} \tag{4.3.1}$$

$j = a+1, a+2, \ldots$ where $a \geq 0$ is an integer, $\lambda > 0$ is a parameter, and

$$s_a = 1 - \sum_{i=0}^{a} \frac{\lambda^i e^{-\lambda}}{i!}.$$

Clearly 4.3.1 is the PMF of a one parameter exponential family having canonical statistic Y and canonical parameter $\theta = \log(\lambda)$. Of course, the original parameter is $\lambda = \exp(\theta)$.

1. **The exceptation**
   Recall that the mean of $X \sim \text{Poisson}(\lambda)$ can be found as follows :

$$E(X) = \sum_{x=0}^{\infty} x p_x$$

The case of the a-truncated distribution make : first sum start at $x = k+1$, since $p_x = 0$ for $x \leq k$, and the probability mass function has an extra factor of s_a^{-1} which we may as well factor outside. We again factor a λ and set $y = x-1$ to obtain a Poisson probability mass function :

$$\begin{aligned} E(X) &= \sum_{x=a+1}^{\infty} x p_x = s_a^{-1} \sum_{x=a+1}^{\infty} x \times \left(\frac{\lambda^x e^{-\lambda}}{x!}\right) \\ &= \lambda s_a^{-1} \sum_{x=a+1}^{\infty} \frac{\lambda^{x-1} e^{-\lambda}}{(x-1)!} = \lambda s_a^{-1} \sum_{z=a}^{\infty} \frac{\lambda^z e^{-\lambda}}{(z)!} \end{aligned}$$

This sum does not come to unity, but

$$\sum_{z=a}^{\infty} \frac{\lambda^z e^{-\lambda}}{z!} = 1 - \sum_{z=0}^{a-1} \frac{\lambda^z e^{-\lambda}}{(z)!} = s_{a-1}$$

(where, in the special case that $a = 0$, we understand $q_{-1} = 1$).
Therefore we obtain :

$$E(X) = \frac{\lambda s_{a-1}}{s_a}$$

2. **The variance**
We can do something similar to work out the variance, for which it is easier to use factorial moments to help with the factorial-cancelling trick [38].
The second factorial moment of a Poisson distribution is :

$$E((X)_2) = E(X(X-1)) = \sum_{x=0}^{\infty} x(x-1)p_x = \sum_{x=0}^{\infty} x(x-1) \times \frac{\lambda^x e^{-\lambda}}{(x)!}$$

Again, we attempt to repeat this analysis for the a-truncated distribution.

$$\begin{aligned} E((X)_2) &= E(X(X-1)) = \sum_{x=0}^{\infty} x(x-1)p_x \\ &= s_a^{-1} \sum_{x=a+1}^{\infty} x(x-1) \left(\frac{\lambda^x e^{-\lambda}}{(x)!} \right) \qquad (4.3.2) \end{aligned}$$

Assuming $a \geq 1$, we can cancel the factorial as before to produce :

$$\begin{aligned} E((X)_2) &= \lambda^2 s_a^{-1} \sum_{x=a+1}^{\infty} \frac{\lambda^{x-2} e^{-\lambda}}{(x-2)!} \\ &= \lambda^2 s_a^{-1} \sum_{z=a-1}^{\infty} \frac{\lambda^z e^{-\lambda}}{z!} = \frac{\lambda^2 s_{a-2}}{s_a} \end{aligned}$$

Applying $Var(X) = E((X)_2) + E(X) - E(X)^2$, we obtain :

$$Var(X) = \frac{\lambda^2 (s_{a-2} s_a - s_{a-1}^2)}{s_a} + \frac{\lambda s_{a-1}}{s_a}$$

We need to reason separately about the case $a = 0$ in (4.3.2), since the first term of the sum, where $x = 1$, will then be $\frac{\lambda e^{-\lambda}}{1!}$ and we can't cancel the factorial to $(-1)!$. However, this term is zero so we can begin the summation at $x = 2$ instead. This is exactly the sum we performed for the un-truncated Poisson distribution, so we obtain $E(X) = s_a^{-1}\lambda^2$. If we take $s_{-2} = 1$ then we can use $\frac{\lambda^2 s_{a-2}}{s_a}$ for $a = 0$ as well as $a \geq 1$; there is no need to treat $a = 0$ as a special case in our final formula.

The cumulant function for the family is then

$$\psi(\theta) = \log\left(e^{e^\theta} - \sum_{j=0}^{k} e^{j\theta}/j!\right)$$

and has derivatives

$$\begin{aligned}
\tau(\theta) &= \psi^{'}(\theta) \\
&= \frac{e^\theta \times e^\theta - \sum_{j=1}^{k} \frac{e^{j\theta}}{(j-1)!}}{e^{e^\theta} - \sum_{j=0}^{k} \frac{e^{j\theta}}{j!}} \\
&= \frac{\mu - e^{-\mu} \sum_{j=1}^{k} \frac{\mu^j}{(j-1)!}}{1 - e^{-\mu} \sum_{j=0}^{k} \frac{\mu^j}{j!}}
\end{aligned}$$

and

$$\begin{aligned}
\psi^{''}(\theta) &= \frac{(e^\theta + e^{2\theta}) \times e^{e^\theta} - \sum_{j=1}^{k} \frac{j e^{j\theta}}{(j-1)!}}{e^{e^\theta} - \sum_{j=0}^{k} \frac{e^{j\theta}}{j!}} - \tau(\theta)^2 \\
&= \frac{(\mu + \mu^2) - e^{-\mu} \sum_{j=1}^{k} \frac{j\mu^j}{(j-1)!}}{1 - e^{-\mu} \sum_{j=0}^{k} \frac{\mu^j}{j!}} - \tau(\theta)^2
\end{aligned}$$

By exponential family theory we know

$$\psi^{'}(\theta) = E_\theta(X)$$

and

$$\psi^{''}(\theta) = Var_{\theta}(X)$$

,

where X is the canonical statistic. Thus from our definition of $\tau(\theta)$ in 4.3.3 it follows that

$$\tau^{'}(\theta) = \psi^{''}(\theta) > 0 \quad \forall\theta.$$

Hence the map τ is one-to-one and defines an invertible change of parameter. Since $\tau(\theta) = E_{\theta}(X)$, it is called the mean value parameter. It is the mean of the distribution under discussion : a-truncated Poisson.

Using $\mu = \exp(\theta)$,

$$\begin{aligned} \psi(\theta) &= \mu + \log\left(1 - e^{-\mu}\sum_{j=0}^{k}\frac{\mu^j}{j!}\right) \qquad (56) \\ &= \mu + \log\left(P_{\mu}(Y > k)\right) \end{aligned}$$

where $Y \sim Poi(\mu)$. This looks fairly stable whether μ is large or small. We will leave the calculation of the log-Poisson probability to R.

4.3.2 A natural estimate of the parameter

If the values $\leq a$ are removed by truncation, we consider the estimator of the parameter λ of a Poisson distribution as follows : if we have a sample $X_1, ..., X_n$ then the natural estimator is $T = \frac{1}{n}\sum_{i=1}^{n} X_i$. We have

$$\begin{aligned} T &= \frac{1}{n}\sum_{i=1}^{n}\sum_{r=a+2}^{\infty}\left(r \times I_{(X_i=r)}\right) \\ &= \frac{1}{n}\sum_{r=a+2}^{\infty}\left(\sum_{i=1}^{n} r \times I_{(X_i=r)}\right) \\ &= \frac{1}{n}\sum_{r=a+2}^{\infty} r\left(\sum_{i=1}^{n} I_{(X_i=r)}\right) \\ &= \frac{1}{n}\sum_{r=a+2}^{\infty} n_r r \end{aligned}$$

n_r is the number of times r is observed.

Proposition 4.3.1 *T is an unbiased estimator of λ variance*

$$Var(T) = \frac{\lambda}{n}\left(1 + \frac{\lambda^{a+1}}{a!(e^{\lambda} - \sum_{r=0}^{a} \frac{\lambda^r}{r!})}\right)$$

We conclude that T is an unbiased estimate with

$$Var(T) = \frac{1}{n}Var(X_i) = \frac{\lambda}{n}\left[1 + \frac{\lambda^{a+1}}{a!} \times \frac{1}{e^{\lambda} - S_a}\right]$$

We obtain an unbiased estimate of the variance of T, given by :

Corollaire 4.3.1

$$\widehat{Var(T)} = \frac{1}{n}\left(T + \frac{(a+1)(a+2)n_{a+2}}{n}\right) \quad (57)$$

4.3.3 The maximum likelihood estimate

We give here the estimation of parameter λ of the Poisson distribution truncated at a.

Proposition 4.3.2 *The maximum likelihood estimator is the root of the equation*

$$\lambda\left(1 + \frac{\lambda^a e^{-\lambda}}{a! S_a}\right) = \overline{x} \quad (4.3.3)$$

$\overline{x}$ is the empirical mean of the observations

We note that the solution of equation (4.3.3) can not obtained analytically. Algorithms must be used to find a λ estimate.

From this latter we may make a comparison of the efficiency of the maximum likelihood method.

4.4 Numerical results

To illustrate the practical application of results obtained in above, an example considered in Feller [13] which is the connections to a wrong telephone number (slightly modified) :

Table 1

k	n_k	k	n_k	k	n_k	k	n_k	k	n_k
2	1	5	14	8	31	11	20	14	7
3	5	6	22	9	40	12	18	15	6
4	11	7	43	10	35	13	12	≥ 16	2

This table shows the statistics of phone connections to a wrong number. A total of $n = 267$ numbers was observed, n_k indicates how many numbers had k wrong connections.

Here $\overline{x} = 8.7416$ $\quad a = 2,$ recall that $T = \frac{1}{n} \sum_{r=a+2}^{\infty} n_r r$

Evaluating of the estimate, variance and the confidence interval of estimate obtained from the sample :

T	8.6779
$\widehat{Var(T)}$	3.4353×10^{-2}
IDC	$(8.3072, 9.0486)$

Remarque 4.4.1 *Note that the parameter estimate is smaller than the sample mean. That's just as it should be, because the estimate accounts for the missing values not present in the data.*

4.5 Iterative solution

For a random variable X, the Poisson distribution truncated at zero is the following distribution conditional :

$$P(X = r|X > 0)) = \frac{\lambda^r}{r!} \times \frac{1}{e^{\lambda} - 1}$$

Let n independent observations $X_1, X_2, ..., X_n$, each have probability density function $f(x)$, the maximum likelihood estimate is the root of the equation :

$$\frac{\lambda}{1 - e^{-\lambda}} = \overline{x} \qquad (4.5.1)$$

where $\overline{x}$ is the empirical mean of the observations. Admittedly, maximum likelihood estimates are troublesome to calculate without proper tables since it is necessary to solve a somewhat complicated non-linear equation.

To simulate a a-truncated Poisson distribution, the simplest method is to simulate ordinary Poisson random variates(using the rpois function in R) and reject all of the simulations less than or equal to a. This works well unless $\mu = \exp(\theta)$,the mean of the untruncated Poisson distribution is nearly zero, in which case the acceptance rate is also nearly zero. In that case,another simple rejection sampling scheme, simulates $Y \sim Poi(\mu)$ and uses $X = Y + m$ as the rejection sampling proposal, where m is a nonnegative integer(the case $m = 0$ is the case already discussed).

4.5.1 Exact Solution

The proposition 4.2.3 gives the exact solution of equation (4.5.1). This proposition is in Kendall's book [19]. This allows us to compare the approximate solution obtained using iterative methods to this solution.

The proposition 4.2.3 gives the expression of maximum likelihood estimate :

$$\hat{\lambda} = \overline{x} + \sum_{r=1}^{\infty} \frac{(-1)^r \overline{x}^r}{r!} \frac{d^{r-1}}{d\overline{x}^{r-1}} e^{-r\overline{x}} = \overline{x} - \sum_{r=1}^{\infty} \frac{r^{r-1}}{r!} (\overline{x} e^{-\overline{x}})^r$$

Table 2

n / λ	4	10	50
50	3.9420	10.3397	49.7551
100	4.0692	9.7094	50.8000
200	4.1379	9.8145	50.300

We can observe that estimates are very close to the actual values.

4.5.2 Using fixed-point theorem

An equation $f(x) = 0$ may be expressed as a fixed-point equation $\varphi(x) = x$. Fixed Point Theory provides essential tools for solving problems arising in various branches of mathematical analysis, such as nonlinear optimization problems, and problems of finding the existence of solution of equations. An element $x \in \mathbb{R}$ is a fixed point of φ if

$$\varphi(x) = x. \tag{4.5.2}$$

Many problems of applied mathematics may be formulated as a fixed point equation or reformulated in this way. The essential step to prove the existence of a solution of the equation (4.5.2) via the iteration.

$$x_{n+1} = \varphi(x_n)$$

$n \in \mathbb{N}$ and $x_0 := x$ is to prove that the orbit $(\varphi^n(x))$, $n \in \mathbb{R}$ is a Cauchy sequence (for every x).The assumption which guarantees this is the assumption that φ is Lipschitz continuous with Lipschitz constant α satisfying $\alpha < 1$.

Fixed-point theorem is very useful for finding out if an equation has a solution. Whether or not this method yields a solution (i.e.,whether or not a fixed-point can be found) depends on the exact nature of the function from which a solution is sought. Fixed point theorem asserts that for a function φ there is at least one point x such that $\varphi(x) = x$; in other words, such that the function φ maps x to itself. Such a point is called a fixed point of the function.

Transform the problem $f(x) = 0$ where $f : \mathbb{R} \to \mathbb{R}$ by an equivalent problem

$$x - \varphi(x) = 0$$

that is to say $x = \varphi(x)$ where $\varphi(.)$ is such that $\varphi(c) = c$. This transformation is always possible but there is not unique function φ. An element $x \in \mathbb{R}$ is a fixed point of φ if $\varphi(x) = x$.

Fixed point theorems are examples of existence theorems, in the sense that they assert the existence of objects, such as solutions to equations, but not necessarily methods for finding such solutions. However, some of these theorems are coupled with algorithms that produce solutions.

Théorème 4.5.1 *Let $\varphi(x)$ be a contraction, Then the following statements hold.*

a)There exists a uniquely determined fixed point z of $\varphi(x)$.

b)The successive iteration $x_{n+1} = \varphi(x_n)$ defines a sequence $(x_n)_n \in \mathbb{N}$ which converges to z.

In our case, we have

$$\varphi(\lambda) = \overline{x}(1 - e^{-\lambda}). \qquad (58)$$

The results obtained using simulated data are :

Table 3

X_0		3			16			50	
n / λ	4	10	50	4	10	50	4	10	50
50	3.94	10.34	49.75	3.94	10.34	49.75	3.94	10.34	49.75
N° iteration	6	4	3	5	3	2	5	3	2
100	4.07	9.71	50.8	4,07	9,71	50.8	4.07	9.71	50.8
N° iteration	5	4	3	6	3	2	6	3	2
200	4.14	9.81	50.3	4,14	9.82	50.3	4.14	9.81	50.3
N° iteration	5	4	3	5	3	2	5	3	2

4.5.3 Robbins-Monro Algorithm

Let us recall Robbins-Monro's scheme [31] to construct approximation algorithms searching the zero θ^* of an unknown function $h : \mathbb{R}^p \to \mathbb{R}^p$, which is observable at any level. First, $Z_0 \in \mathbb{R}^p$ is arbitrarily chosen, and then the sequence (Z_n) is recursively defined by setting

$$Z_n = Z_{n-1} + \gamma_n W_n$$

where W_n is an observation of the function h at the point Z_{n-1}, and where the stepsize (γ_n) is a sequence of positive real numbers going to zero. Let $X_1, ..., X_n$ be independent, identically distributed $\mathbb{R}^p$-valued random vectors, let f denote a function, and assume that f has a unique maximizer θ. Djeddour and all [9] introduced the method of stochastic approximation to construct a recursive estimator of the size of the mode.

Let $f : \mathbb{R} \to \mathbb{R}$ continuous, bounded, strictly increasing and vanishes at a unique point z_0. Let $(\gamma_n)_{n\geq 1}$ be a sequence of positive real numbers such that

$$\sum \gamma_n^2 < \infty \qquad \sum \gamma_n = \infty$$

Let $(U_n)_n$ a sequence of random variables. i.i.d. uniformly distributed on $[-1, 1]$. We define a sequence $(Z_n)_{n\geq 0}$ by

$$Z_0 \in \mathbb{R} \qquad Z_n = Z_{n-1} - \gamma_n(f(Z_{n-1}) + U_n).$$

Then

$$Z_n \to Z_0$$

as $n \to \infty$ almost surely. Note that for any $\alpha > 0$,

$$\gamma_n = \frac{1}{n^\alpha}$$

satisfies the above condition if and only if $0.5 < \alpha < 1$. The convergence of iterative methods for determining the roots of a nonlinear equation in general depends on the choice of the initial data X_0. The Robbins-Monro algorithm [31] is a basic stochastic approximation scheme involve finding the roots of a function under noisy observations.

Statistical studies were developed using stochastic algorithms in [11] and in [9]. We obtain an iteration scheme

$$\lambda(t+1) = \overline{y}(1 - exp(-\lambda(t))) \qquad (59)$$

to compute the ML estimate starting from an appropriate initial value, for which the sample mean $\lambda(0) = \overline{y}$ can be used.

Simulation results

$\alpha = 0.54$

X_0		3			16			50	
n / λ	4	10	50	4	10	50	4	10	50
50	3.94	10.34	49.75	3.94	10.34	49.75	3.94	10.34	49.75
N° iteration	27	29	23	20	4	2	20	4	2
100	4.07	9.71	50.8	4,07	9,71	50.8	4.07	9.71	50.8
N° iteration	19	22	23	37	6	2	37	6	2
200	4.14	9.81	50.3	4,14	9.82	50.3	4.14	9.81	50.3
N° iteration	22	29	23	22	4	2	22	4	2

$\alpha = 0.6$

X_0		3			16			50	
n / λ	4	10	50	4	10	50	4	10	50
50	3.94	10.34	49.75	3.94	10.34	49.75	3.94	10.34	49.75
N° iteration	35	38	29	26	4	2	26	4	2
100	4.07	9.71	50.8	4,07	9,71	50.8	4.07	9.71	50.8
N° iteration	25	28	29	50	6	2	50	6	2
200	4.14	9.81	50.3	4,14	9.82	50.3	4.14	9.81	50.3
N° iteration	29	40	29	28	5	2	27	5	2

$\alpha = 0.8$

X_0		3			16			50	
n / λ	4	10	50	4	10	50	4	10	50
50	3.94	10.34	49.75	3.94	10.34	49.75	3.94	10.34	49.75
N° iteration	140	164	105	83	4	2	83	4	2
100	4.07	9.71	50.8	4,07	9,71	50.8	4.07	9.71	50.8
N° iteration	19	22	23	37	6	2	37	6	2
200	4.14	9.81	50.3	4,14	9.82	50.3	4.14	9.81	50.3
N° iteration	100	173	107	94	6	2	94	6	2

$\alpha = 0.9$

X_0		3			16			50	
n / λ	4	10	50	4	10	50	4	10	50
50	3.94	10.34	49.75	3.94	10.34	49.75	3.94	10.34	49.75
N° iteration	497	634	333	234	5	2	234	5	2
100	4.07	9.71	50.8	4,07	9,71	50.8	4.07	9.71	50.8
N° iteration	214	306	340	1202	11	2	1202	12	2
200	4.14	9.81	50.3	4,14	9.82	50.3	4.14	9.81	50.3
N° iteration	308	686	340	282	6	2	282	6	2

Conclusion 4.5.1 *The fixed point theorem method gives better results than that of Robbins Monroe. In the sense that the convergence towards the exact solution is faster. Additionally, the Robbins-Monro algorithm method depends on the initial value X_0 and the value of the step (more precisely the value of α) unlike the fixed point which does not depend on the initial value.*

4.6 Proofs

4.6.1 Proof of Proposition 4.2.2

Note that truncation here means that 0-data is removed. We have

$$L(x, \lambda) = \sum_{i=1}^{n} \log(f(x_i, \lambda))$$

with

$$f(x_i, \lambda) = \frac{\lambda^{x_i}}{x_i!} \times \frac{1}{e^{\lambda} - 1}$$

Then we have

$$L(x, \lambda) = \log(\lambda) \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \log(x_i!) - n \log(e^{\lambda} - 1)$$

from where

$$\frac{dL(x, \lambda)}{d\lambda} = \frac{1}{\lambda} \sum_{i=1}^{n} x_i - n \frac{e^{\lambda}}{(e^{\lambda} - 1)}$$

Then we deduce (4.2.1).

4.6.2 Proof of Corollary 4.2.1

Indeed we use a limited expansion of the function $g(\lambda) = \frac{\lambda}{1-e^{-\lambda}}$ in $\hat{\lambda}$ in the neighborhood of λ which is given by

$$g(\hat{\lambda}) = g(\lambda) + g^{'}(\lambda)(\hat{\lambda} - \lambda) + O(1).$$

We deduce

$$\begin{aligned} E(g(\hat{\lambda})) &= g(\lambda) + g^{'}(\lambda)(E(\hat{\lambda}) - \lambda) \\ &= g(\lambda) - \lambda g^{'}(\lambda) + g^{'}(\lambda)E(\hat{\lambda}) \end{aligned}$$

If we take the expectation of the two members of the equation (4.5.1), we get $E\left(\frac{\hat{\lambda}}{1-e^{\hat{\lambda}}}\right) = E\overline{x}$, but

$$E\overline{x} = \frac{\lambda}{1-e^{-\lambda}} = g(\lambda)$$

we conclude that $E(\hat{\lambda}) = \lambda$

Let's move on to the variance : if we take the variance of the two sides of the equation (4.5.1), we get

$$Var(\overline{x}) = Var\left(\frac{\hat{\lambda}}{1-\hat{\lambda}}\right) \tag{4.6.1}$$

But $Var(\overline{x}) = \frac{\lambda(1-(\lambda+1)e^{-\lambda})}{n(1-e^{-\lambda})^2}$ and setting $g(\lambda) = \frac{\lambda}{1-e^{-\lambda}}$; we deduce

$$\begin{aligned} Var\left(\frac{\hat{\lambda}}{1-\hat{\lambda}}\right) &= (g^{'}(\lambda))^2 Var(\hat{\lambda}) \\ &= \left(\frac{1-(\lambda+1)e^{-\lambda}}{(1-e^{-\lambda})^2}\right)^2 var(\hat{\lambda}) \end{aligned}$$

By replacing in (4.6.1), we have

$$Var(\hat{\lambda}) = \frac{\lambda}{n} \times \frac{(1-e^{-\lambda})^2}{1-(\lambda+1)e^{-\lambda}},$$

hence the result of the corollary.

4.6.3 Proof of Proposition 4.3.1

The truncated random variable X_i has distribution

$$P(X_i = r|X_i > a)) = \frac{P(X_i = r)}{P(X_i > a)} = \frac{\lambda^r}{r!} \times \frac{1}{e^\lambda - S_a}$$

where we set $S_a = \sum_{r=0}^{a} \frac{\lambda^r}{r!}$. Starting from $Var(T) = \frac{1}{n} Var(X_i)$, we calculate $Var(X_i)$.

— The expectation of X_i

$$\begin{aligned} E(X_i) &= \sum_{r=a+2}^{\infty} r \frac{\lambda^r}{r!} \times \frac{1}{e^\lambda - S_a} = \frac{\lambda}{e^\lambda - S_a} \left(\sum_{r=a+2}^{\infty} \frac{\lambda^{r-1}}{(r-1)!} \right) \\ &= \frac{\lambda}{e^\lambda - S_a} \left(e^\lambda - S_a \right) = \lambda \end{aligned}$$

— The variance of X_i

$$\begin{aligned} E(X_i^2) &= \sum_{r=a+2}^{\infty} r^2 \frac{\lambda^r}{r!} \times \frac{1}{e^\lambda - S_a} = \frac{1}{e^\lambda - S_a} \sum_{r=a+2}^{\infty} r^2 \frac{\lambda^r}{r!} \\ &= \frac{1}{e^\lambda - S_a} \left(\lambda^2 \sum_{r=a+2}^{\infty} \frac{\lambda^{r-2}}{(r-2)!} + \lambda \sum_{r=a+2}^{\infty} \frac{\lambda^{(r-1)}}{(r-1)!} \right) \\ &= \frac{1}{e^\lambda - S_a} \left[\lambda^2 (e^\lambda - S_{a-1}) + \lambda (e^\lambda - S_a) \right] = \left[\lambda^2 \frac{e^\lambda - S_{a-1}}{e^\lambda - S_a} + \lambda \right] \end{aligned}$$

From where

$$Var(X_i) = \lambda^2 \times \frac{e^\lambda - S_{a-1}}{e^\lambda - S_a} + \lambda - \lambda^2 = \lambda^2 \times \frac{\lambda^a}{a!} \times \frac{1}{e^\lambda - S_a} + \lambda$$

We deduce that T is an unbiased estimate with

$$Var(T) = \frac{1}{n} Var(X_i) = \frac{\lambda}{n} \left[1 + \frac{\lambda^{a+1}}{a!} \times \frac{1}{e^\lambda - S_a} \right]$$

4.6.4 Proof of Corollary 4.3.1

Indeed $n_{(a+2)}$ is the number of observations equal to $(a+2)$; the true associate is $N_{(a+2)}$. It follows a binomial distribution with parameter $\frac{\lambda^{(a+2)}}{(a+2)!}e^{-\lambda}$.

We can therefore approximate $e^{-\lambda}\lambda^{(a+2)}$ by $\frac{(a+2)!n_{(a+2)}}{n}$.

4.6.5 Proof of Proposition 4.3.2

Let $x_1, x_2, ..., x_n$ be a random sample with

$$p(X = x_i) = \frac{\lambda^{x_i}e^{-\lambda}}{x_i!\left(1 - \sum\limits_{k=0}^{a} \frac{\lambda^{x_k}e^{-\lambda}}{x_k!}\right)} = \frac{\lambda^{x_i}e^{-\lambda}}{x_i!S_a} \tag{4.6.2}$$

The likelihood function is

$$\begin{aligned} L(x,\lambda) &= \prod_{i=1}^{n}\left[\frac{\lambda^{x_i}e^{-\lambda}}{x_i!S_a}\right] = \frac{e^{-n\lambda}}{S_a^n}\prod_{i=1}^{n}\left[\frac{\lambda^{x_i}}{x_i!}\right] \\ &= \frac{e^{-n\lambda}}{S_a^n} \times \frac{\lambda^{\sum x_i}}{\prod\limits_{i=1}^{n} x_i!} \end{aligned}$$

The log-likelihood function is

$$\log L(x,\lambda) = -n\lambda - n\log S_a + \left(\sum x_i\right) \times \log\lambda - \sum \log\left(x_i!\right)$$

Then

$$\widehat{\lambda} = \arg\max_\lambda\left[\log L(x,\lambda)\right]$$

i.e.

$$\widehat{\lambda} = \arg\max_{\lambda}\left[-n\lambda - n\log S_a + \left(\sum x_i\right) \times \log\lambda - \sum \log\left(x_i!\right)\right]$$

We have

$$\frac{\partial}{\partial\lambda}\log\left(1-\sum_{k=0}^{a}\frac{\lambda^{x_k}e^{-\lambda}}{x_k!}\right)=\frac{\lambda^a e^{-\lambda}}{a!\left(1-\sum\limits_{k=0}^{a}\lambda^{x_k}\frac{e^{-\lambda}}{x_k!}\right)} \tag{4.6.3}$$

The derivative of $L(x,\lambda)$ is deduced :

$$\frac{\partial}{\partial\lambda}\left(L(x,\lambda)\right)=-n-n\times\left[\frac{\lambda^a e^{-\lambda}}{a!\left(1-\sum\limits_{k=0}^{a}\lambda^{x_k}\frac{e^{-\lambda}}{x_k!}\right)}\right]+\frac{\sum x_i}{\lambda}$$

To find the maximum of $L(x,\lambda)$, one must have :

$$\frac{\partial}{\partial\lambda}\left(L(x,\lambda)\right)=0 \tag{4.6.4}$$

$\Rightarrow$

$$-n-\frac{n\lambda^a e^{-\lambda}}{a!\left(1-\sum\limits_{k=0}^{a}\lambda^{x_k}\frac{e^{-\lambda}}{x_k!}\right)}+\frac{\sum x_i}{\lambda}=0$$

Thus, λ has to check the following equation :

$$\frac{n\overline{x}}{\lambda}-\frac{n\lambda^a e^{-\lambda}}{a!S_a}-n=0$$

Which gives

$$\lambda\left(1+\frac{\lambda^a e^{-\lambda}}{a!S_a}\right)=\overline{x} \tag{4.6.5}$$

Example : simulate zero-truncated Poisson distribution

```
x <- rtpois(1e5, 14, 16)
xx <- seq(-1, 50)
plot(prop.table(table(x)))
```

```
lines(xx, dtpois(xx, 14, 16), col = "red")
hist(ptpois(x, 14, 16))
xx <- seq(0, 50, by = 0.01)
plot(ecdf(x))
lines(xx, ptpois(xx, 14, 16), col = "red", lwd = 2)
uu <- seq(0, 1, by = 0.001)
lines(qtpois(uu, 14, 16), uu, col = "blue", lty = 2)
# Zero-truncated Poisson
x <- rtpois(1e5, 5, 0)
xx <- seq(-1, 50)
plot(prop.table(table(x)))
lines(xx, dtpois(xx, 5, 0), col = "red")
hist(ptpois(x, 5, 0))
xx <- seq(0, 50, by = 0.01)
plot(ecdf(x))
lines(xx, ptpois(xx, 5, 0), col = "red", lwd = 2)
lines(qtpois(uu, 5, 0), uu, col = "blue", lty = 2)
```

Chapitre 5

Conclusion

This thesis consists of two parts namely :

1. a part which studies the estimation of quantile.
2. The second concerns the study of iterative solution of truncated Poisson distribution.

Estimation of quantile

For a distribution function F, the q^{th} quantile θ_q of F is defined as $\theta_q = F^{-1}(q)$ with $F^{-1}(q) = \inf\{x : F(x) \geq q\}$. Rather than direct inversion of the distribution function, the quantile can be determined through a nonparametric estimator of the distribution function, noted $F_n(x)$.

Our aim was to provide an algorithm in order to approximate the quantile q. To construct a stochastic algorithm, which approximates the function q, we define an algorithm of search of the zero of the function $\varphi : x \rightarrow \varphi(x) = \delta - F_n(x)$, where δ satisfying $F_n(q_n) = \delta$ and $F_n(x)$ is estimated by integrating a kernel estimator of the density. In order to construct an algorithm, which approximates the quantile of distribution function (in other words, which approximates q, we define an algorithm searching the zero of the function $h : z \mapsto \delta - F(z)$). Following Robbins-Monro's scheme, we set $q_0 \in \mathbb{R}$, and, for $n \geq 1$. We adopted the approach developed in [36]. In order to avoid the estimation of the density f on all $\mathbb{R}^d$ to estimate the mode θ, in a pioneering work, Tsybakov introduced a recursive estimator of θ.

The main contributions of this part are

(i) the generalization of the previous work of Djeddour *et all* [9] to a quantile of a distribution,

(ii) the illustration of the theoretical results developed in the above-cited references through a simulation study,

(iii) the improvement of the algorithm with regard to kernel estimation.

Iterative solution of truncated Poisson distribution

This part works through the details of the a-truncated Poisson distribution, a special case of which is the zero-truncated Poisson distribution. The a-truncated Poisson distribution is the distribution of a Poisson random variable Y conditional on the event $Y > a$. It has one parameter, which we may take to be $\lambda = E(Y)$. Since λ is not the mean(or anything else simple) of the distribution of Y conditioned on the even $Y > a$, we do not call λ the mean, rather we call it the original parameter. The performance of this algorithm seems to be fine for small a.

The convergence of iterative methods for determining the roots of a nonlinear equation in general depends on the choice of the initial data X_0. We have used several methods for the numerical results; namely the fixed point theorem, the Robin-Monro algorithm, the exact solution (not demonstrated, found in Feller's book) of the equation obtained by the estimation of the maximum likelihood. The best results are obtained by the fixed point method. The latter method does not depend on the initial point X_0.

Perspectives

Use stochastic algorithms to statistical problems to give perspectives that are different from classical methods.

• We wish to continue to explore stochastic algorithms and be able to use them to other statistical applications, estimation etc...

• Testing is also part of our field. We are talking about developing some of them. Given a sample of independent measurements, we can note the signs of the deviations from the median specified in H_0 and note the magnitude of each deviation. If H_0 is true, deviations from a given quantity are equally likely, for a symmetric distribution, to be positive as negative

• Another example, in hydrology, we sometimes want to get the value exceeded by the height of a river with a given probability of confidence. This information is generally the basis for decision-making and regulatory action. We can stretch some of our work and explore more specifically, certain quantiles.

Bibliographie

[1] A. Amiri and Baba Thiam (2014 A smoothing stochastic algorithm for quantile estimation. Statistics and Probability Letters 93 116-125.

[2] A. Azzalini(1981). A note on the estimation of a distribution function and quantiles by a kernel method. Biometrika 68 326-328.

[3] Blum, J.R., 1954. Multidimensional stochastic approximation methods. Ann. Math. Statist. 25, 737–744.Blum, J.R., 1954. Multidimensional stochastic approximation methods. Ann. Math. Statist. 25, 737–744.

[4] Chen, H.F., Guo, L., Gao, A.J., (1988). Convergence and robustness of the Robbins-Monro algorithm truncated at randomly varying bounds. Stochastic Process. Appl., 217-231

[5] Cohen, A.C. (1960). Estimating the parameter in a conditional Poisson distribution, Biometrics 16, 203-211.

[6] Cohen, A.C. (1960). An extension of a truncated Poisson distribution. Biometrics, 16, 446-450.

[7] Cohen, A.C. (1960). Estimation in a truncated Poisson distribution when zeroes and some ones are missing. Journal of the American Statistical Association 55, 342- 348.

[8] David, F.N and Johnson, N.L.(1952). The truncated Poisson. Biometrics 8, 275-285.

[9] Djeddour, K., Mokkadem A., and Pelletier M. (2008). On the recursive estimation of the location and of the size of the mode of a probability density. Serdica Math. J. 34 651-688.

[10] K. Djeddour (2003). Estimation récursive du mode et de la valeur modaled'une densité, Test d'ajustement de loi, Ph.D. Thesis. University of Versailles, France (2003).

[11] Djeddour-Djaballah, K. and Bachir, A. (2020). Recursive quantile estimation using a stochastic algorithm. Philippine Statistical journal Volume 69, Number 1 ; 89-109.

[12] Duflo,M (1997). Random Iterative Models. Collection Applications of mathematics, Springer (1997).

[13] Feller,E (1957). An introduction to probability and its applications. Wiley. New York

[14] Fisz, M., (1963). Probability Theory and Mathematical Statistics. Wiley, New-York. Fisz, M., 1963. Probability Theory and Mathematical Statistics. Wiley, New-York.

[15] Gourieroux, C., Laurent, J.P., Scaillet, O., (2000). Sensitivity analysis of values at risk. J. Empir. Finance 7, 225–245.Gourieroux, C., Laurent, J.P., Scaillet, O., 2000. Sensitivity analysis of values at risk. J. Empir. Finance 7, 225–245.

[16] N.L. Johnson and S. Kotz (1969). Distributions in statistics, Discrete distributions. John Wiley and Sons, Ltd., Chichester, Sussex, England, p.104–109.

[17] Holst, U. (1987). Recursive estimation of quantiles using recursive kernel density estimators. Sequential Anal. 6 (3), 217–237.

[18] Horn, R.A., and C.R. Johnson. (1985). Matrix Analysis. Cambridge : Cambridge University Press

[19] Kendall M, Stuart A (1954). Advanced theory of statistics, 2nd edition. London : Charles Griffen and Company Limited.

[20] J. Kiefer and J. Wolfowitz (1952). Stochastic approximation of the maximum of a regression functions. Ann.Math.Statist. 23 462-466.

[21] T. Kim, D. Kim, S. Lee, S. Park, L. Jang (2021). arXiv :2104.13522 [math.PR] (or arXiv :2104.13522v1 [math.PR] for this version) https ://doi.org/10.48550/arXiv.2104.13522

[22] M. Lejeune and Sarda, P. (1992). Smoothing estimators of distribution and density function. Computational Statistics and Data Analysis 14 457-471.

[23] A. Mokkadem and M. Pelletier (2007). The Annals of Statistics Vol. 35, No. 4, pp. 1749-1772.

[24] E.A. Nadaraya (1964). Some new estimates for distribution function. Theory Probab. Appl. 9 497-500.

[25] Niyomdecha, A. ; Srisuradetchai, P. (2023). Complementary Gamma Zero-Truncated Poisson Distribution and Its Application. Mathematics 2023, 11, 2584.

[26] M. Pelletier (1998). On the almost sure asymptotic behaviour of stochastic algorithms. Stochastic Process. Appl. 78 217-244.

[27] Plackett, R.L. (1953). The truncated Poisson distributions, Biometrics 9, 485-488.

[28] B. T. Polyak (1990). A new method of stochastic approximation type. Avtomatika i Telemekhanika 51(7) 98-107.

[29] Ralescu, S. (2012). Asymptotic theorems for kernel U-quantiles. Electronic Journal of Statistics, Vol.6, 664-671.

[30] P. Revesz, Robbins-Monro procedure in a Hilbert space and its application in the theory of learning processes. I. Studia Sci. Math. Hung. 8 (1973) 391-398.

[31] H. Robbins and S. Monro(1951). A stochastic approximation method. The Annals of Mathematical Statistics 22, 3 400-407.

[32] D. Robinson (1975). Non-parametric quantile estimation through stochastic approximation. Thesis and Dissertation Collection.

[33] D. Ruppert (1985). A Newton-Rafson version of the multivariate Robbins-Monro procedure. Ann. Statist. 13 236-245.

[34] Tate, R.F. and Goen, R.L. (1958). Minimum variance unbiased estimation for the truncated Poisson distribution, Annals of Mathematical Statistics 29, 755- 765.

[35] L. Tierney (1983). A space-efficient recursive procedure for estimating a quantile of an unknown distribution. SIAM Journal of Scientific and Statistical Computing, 4, 4.

[36] A.B. Tsybakov (1990). Recurrent estimation of the mode of a multidimensional distribution. Problems Inform. Transmission, 26 31-37.

[37] Tze Leung Lai (2002). Stochastic Approximation. Technical Report No. 2002-31. Department of Statistics Stanford University Stanford, California 94305-4065.

[38] https ://stats.stackexchange.com/questions/179426/how-to-derive-the-mean-and-variance-of-a-k-truncated-poisson