# Recursive kernel distribution estimators defined by stochastic approximation method using Bernstein polynomials

Asma JMAEI[1,2] and Yousri SLAOUI[1]

[1] Université de Poitiers, Laboratoire de Mathématiques et applications,

UMR 7348, 11 Boulevard Marie et Pierre Curie, 86962 Futuroscope Chasseneuil, France

[2] Faculté des sciences de Bizerte, Tunisie

asma.jmaei@math.univ-poitiers.fr

yousri.slaoui@math.univ-poitiers.fr

**Résumé.** Nous proposons des estimateurs récursifs d'une fonction de distribution à l'aide de l'algorithme de Robbins-Monro et du polynôme de Bernstein. Nous étudions les propriétés de ces estimateurs et nous les comparons avec celle de l'estimateur de distribution de Vitale. Nous montrons qu'avec un choix optimal des paramètres, notre algorithme domine celui de Vitale en terms d'erreur quadratique moyenne intégrée. Ensuite, nous confirmons ces résultats théoriques par des simulations.
**Mots clés:** Estimation de la distribution; Algorithme d'approximation stochastique; Polynôme de Bernstein.

**Abstract.** We propose recursive estimators of a distribution function using Robbins-Monro algorithm and Bernstein polynomials. We study the properties of these estimators and compare them with that of the Vitale's distribution estimator. We show that, with optimal parameters, our proposed estimator dominates the Vitale's estimator in terms of their Mean Integrated Square Error performance. Then, we confirm these theoretical results by simulations.
**Key words:** Distribution estimation; Stochastic approximation algorithm; Bernstein polynomial.

## 1 Introduction

We consider $X_1, X_2, \ldots, X_n$ a sequence of i.i.d random variables having a common unknown distribution $F$ with associated density $f$ supported on $[0,1]$. Using Robbins-Monro's scheme (see Robbins and Monro (1951)), we construct a stochastic algorithm, which approximates the function $F$ at a given point $x$. We define the algorithm to search the zero of the function $h : y \mapsto F(x) - y$ as following : (i) we set $F_0(x) \in \mathbb{R}$ ; (ii) for all $n \geq 1$, we set

$$F_n(x) = F_{n-1}(x) + \gamma_n W_n,$$

where $W_n$ is an observation of the function $h$ at the point $F_{n-1}(x)$. To define $W_n$, we follow Vitale (1975) (see also Babu et al. (2002)) and introduce the Bernstein polynomial of order $m > 0$ (we assume that $m = m_n$ depends on $n$),

$$b_k(m,x) = C_m^k x^k (1-x)^{m-k} \quad \text{and set} \quad W_n = \sum_{k=0}^{m} \mathbb{I}\left\{X_n \leq \frac{k}{m}\right\} b_k(m,x) - F_{n-1}(x).$$

So the recursive estimator of the distribution $F$ at the point $x$ can be written as

$$F_n(x) = (1-\gamma_n)F_{n-1}(x) + \gamma_n \sum_{k=0}^{m} \mathbb{I}\left\{X_n \leq \frac{k}{m}\right\} b_k(m,x). \tag{1}$$

Throughout this paper, we suppose that $F_0(x) = 0$ and we let $\Pi_n = \prod_{j=1}^{n}(1-\gamma_j)$ and $Z_n(x) = \sum_{k=0}^{m} \mathbb{I}\left\{X_n \leq \frac{k}{m}\right\} b_k(m,x)$. Then, it follows from (1) that one can estimate $F$ recursively at the point $x$ by:

$$F_n(x) = \Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \gamma_k Z_k(x). \tag{2}$$

The aim of this paper is to study the properties of the recursive distribution estimator defined by the stochastic approximation algorithm (2), and its comparison with the distribution estimator introduced by Vitale (1975)

$$\tilde{F}_n(x) = \frac{1}{n} \sum_{k=0}^{m} \hat{F}_n\left(\frac{k}{m}\right) b_k(m,x), \tag{3}$$

where $\hat{F}_n$ is the empirical distribution function.
Some theoretical properties of the estimator $\tilde{F}_n$ have been investigated (see Leblanc (2012)).

# 2 Assumptions and Notations

*Definition* 1. Let $\gamma \in \mathbb{R}$ and $(v_n)_{n\geq 1}$ be a nonrandom positive sequence. We say that $(v_n) \in \mathcal{GS}(\gamma)$ if

$$\lim_{n\to+\infty} n\left[1 - \frac{v_{n-1}}{v_n}\right] = \gamma.$$

This condition was introduced by Galambos and Seneta (1973) to define regularly varying sequences (see also Bojanic and Seneta (1995)). Typical sequences in $\mathcal{GS}(\gamma)$ are, for $b \in \mathbb{R}$, $n^\gamma(\log n)^b$, $n^\gamma(\log\log n)^b$, and so on.

To obtain the behavior of the estimator defined in (2), we make to the following assumptions :

(A1) $F$ is continuous and admits two continuous and bounded derivatives.

(A2) $(\gamma_n) \in \mathcal{GS}(-\alpha)$,$\alpha \in ]\frac{1}{2}, 1]$.

(A3) $(m_n) \in \mathcal{GS}(a)$, $a \in ]0, 1[$.

(A4) $\lim_{n\to\infty}(n\gamma_n) \in ]\min\left(a,(2\alpha+a)/4\right),\infty]$.

Assumption $(A4)$ on the limit of $(n\gamma_n)$ as $n$ goes to infinity is usual in the framework of stochastic approximation algorithms.
Throughout this paper we will use the following notations:

$$\begin{aligned}
\Pi_n &= \prod_{j=1}^{n}(1-\gamma_j), \qquad Z_n(x)=\sum_{k=0}^{m}\mathbb{I}\left\{X_n\leq\frac{k}{m}\right\}b_k(m,x), \qquad \xi=\lim_{n\to\infty}(n\gamma_n)^{-1},\\
b(x) &= \frac{x(1-x)f'(x)}{2}, \qquad \sigma^2(x)=F(x)\left[1-F(x)\right], \qquad V(x)=f(x)[\frac{2x(1-x)}{\pi}]^{\frac{1}{2}}.\\
C_1 &= \int_0^1\sigma^2(x)dx, \qquad C_2=\int_0^1 V(x)dx, \qquad C_3=\int_0^1 b^2(x)dx.
\end{aligned}$$

# 3 Main Results

Our first result is the following proposition which gives the bias and the variance of $F_n$.
*Proposition* 1 (Bias and variance of $F_n$).
Let Assumptions $(A1)-(A4)$ hold.

1. If $0<a\leq\frac{2}{3}\alpha$, then

$$\mathbb{E}[F_n(x)]-F(x)=m_n^{-1}\frac{1}{1-a\xi}b(x)+o\left(m_n^{-1}\right). \tag{4}$$

   If $\frac{2}{3}\alpha<a<1$, then

$$\mathbb{E}[F_n(x)]-F(x)=o\left(\sqrt{\gamma_n m_n^{-1/2}}\right).$$

2. If $\frac{2}{3}\alpha\leq a<1$, then

$$Var[F_n(x)]=\gamma_n\frac{1}{2-\alpha\xi}\sigma^2(x)-\gamma_n m_n^{-1/2}\frac{2}{4-(2\alpha+a)\xi}V(x)+o\left(\gamma_n m_n^{-1/2}\right). \tag{5}$$

   If $\alpha/2\leq a<\frac{2}{3}\alpha$, then

$$Var[F_n(x)]=\gamma_n\frac{1}{2-\alpha\xi}\sigma^2(x)+o(\gamma_n).$$

   If $0<a<\alpha/2$, then

$$Var[F_n(x)]=o\left(m_n^{-2}\right).$$

3. If $\lim_{n\to\infty}(n\gamma_n)>\max\left(a,(2\alpha+a)/4\right)$, then (4) and (5) hold simultaneously.

The following corollary shows that, for $(\gamma_n)=(\gamma_0 n^{-1})$ with $\gamma_0\in]0,+\infty[$, the optimal value for the order $(m_n)$ depends on $\gamma_0$ and then the corresponding $MISE=\int_{\mathbb{R}}\mathbb{E}\left[F_n(x)-F(x)\right]^2dx$ depend also on the order $(m_n)$.

*Corollary* 1.
Let Assumptions (A1)-(A4) hold. To minimize the $MISE$ of $F_n$, the stepsize $(\gamma_n)$ must be chosen in $\mathcal{GS}(-1)$, $\lim_{n\to\infty} n\gamma_n = \gamma_0$, $(m_n)$ must equal to

$$\left(2^{2/3}\left(\gamma_0 - 2/3\right)^{-2/3}\left\{\frac{4C_3}{C_2}\right\}^{2/3} n^{2/3}\right), \tag{6}$$

and then

$$MISE\left(F_n\right) = n^{-1}\frac{\gamma_0^2}{2\gamma_0 - 1}C_1 - \frac{3}{4}\frac{1}{2^{4/3}}\frac{\gamma_0^2}{\left(\gamma_0 - 2/3\right)^{2/3}}\frac{C_2^{4/3}}{4^{1/3}C_3^{1/3}}n^{-4/3} + o\left(n^{-4/3}\right).$$

Let us now state the following theorem, which gives the weak convergence rate of the estimator $F_n$ defined in (2).

*Theorem* 1 (Weak pointwise convergence rate).
Let Assumption (A1)-(A4) hold.

1. If $\gamma_n^{-1/2}m_n^{-1} \to c$ for some constant $c \geq 0$, then

$$\gamma_n^{-1/2}(F_n(x) - F(x)) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\frac{c}{1-a\xi}b(x), \frac{1}{2-\alpha\xi}\sigma^2(x)\right).$$

2. If $\gamma_n^{-1/2}m_n^{-1} \to \infty$, then

$$m_n(F_n(x) - F(x)) \xrightarrow{\mathbb{P}} \frac{b(x)}{1-a\xi},$$

where $\xrightarrow{\mathcal{D}}$ denotes the convergence in distribution, $\mathcal{N}$ the Gaussian-distribution and $\xrightarrow{\mathbb{P}}$ the convergence in probability.

# 4 Simulations

The aim of this paragraph is to compare the performance of Vitale's estimator defined in (3) with that of the proposed estimator (2).
When applying $F_n$ one needs to choose two quantities:

- The stepsize $(\gamma_n) = (\gamma_0 n^{-1})$, where $\gamma_0 = 2/3 + c$, with $c \in ]0, 1/3]$.
- The order $(m_n)$ is chosen to be equal to (6).

In order to investigate the comparison between two estimators, we consider two samples sizes : $n = 50$ and $n = 100$, and the beta mixture distribution $0.5\mathcal{B}(2.5, 6) + 0.5\mathcal{B}(9, 1)$ (see Table 1). We compute the $MISE$.

From Tables 1, we conclude that:

- Using an appropriate choice of the stepsize $(\gamma_n)$, the $MISE$ of the distribution estimator defined by (2) can be smaller than that of the Vitale's estimator defined in (3).

| | Vitale | estimator 1 | estimator 2 | estimator 3 | estimator 4 |
|---|---|---|---|---|---|
| $n = 50$ | | | | | |
| $MISE$ | 0.002847 | 0.002527 | 0.003109 | 0.003154 | 0.002993 |
| $n = 100$ | | | | | |
| $MISE$ | 0.001509 | 0.001483 | 0.001684 | 0.001670 | 0.001567 |

Table 1: Quantitative comparison between Vitale's estimator (3) and four estimators; estimator 1 corresponds to the estimator (2) with $(\gamma_n) = ([2/3 + 0.02]n^{-1})$, estimator 2 corresponds to the estimator (2) with $(\gamma_n) = ([2/3+0.05]n^{-1})$, estimator 3 corresponds to the estimator (2) with $(\gamma_n) = ([2/3 + 0.1]n^{-1})$, estimator 4 corresponds to the estimator (2) with $(\gamma_n) = (n^{-1})$. Here we consider the beta mixture distribution $0.5\mathcal{B}(2.5, 6) + 0.5\mathcal{B}(9, 1)$, two sample sizes $n = 50$ and $n = 100$, and we compute the $MISE$.

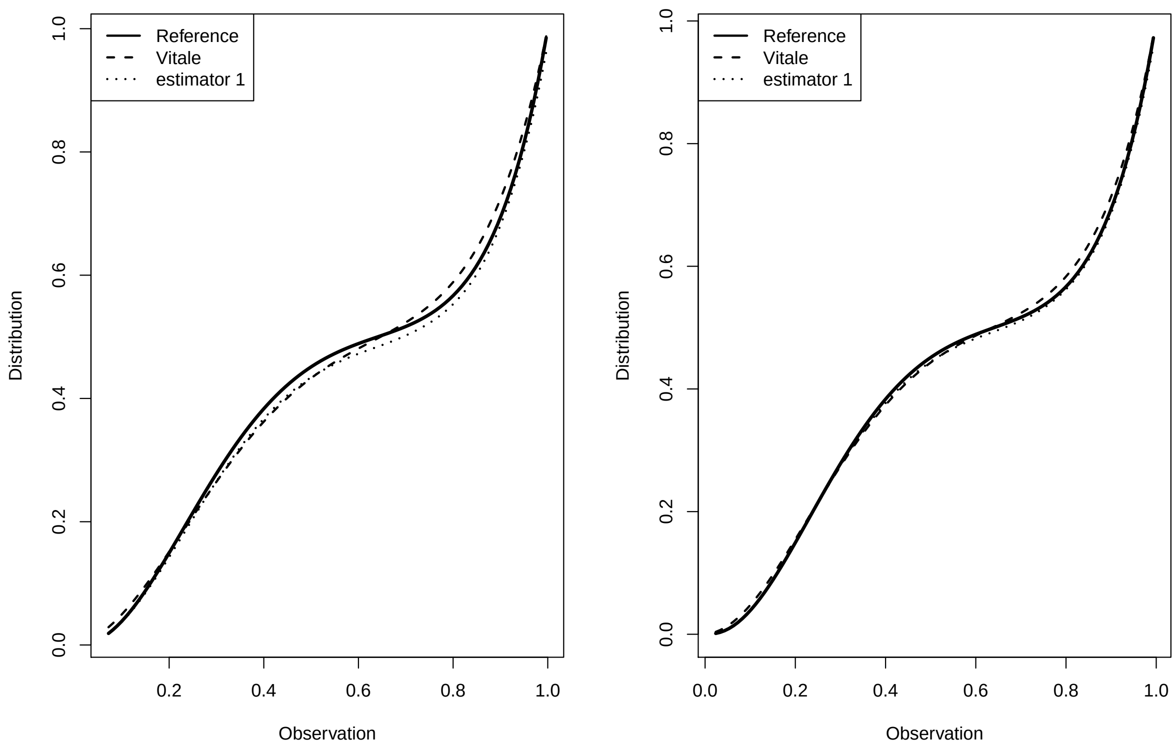


Figure 1: Qualitative comparison between the estimator $\tilde{F}_n$ defined in (3) and the proposed distribution estimator (2) with stepsize $(\gamma_n) = ([2/3 + 0.02]n^{-1})$, for 500 samples respectively of size 50 (left panel) and of size 100 (right panel) for the beta mixture distribution $0.5\mathcal{B}(2.5, 6) + 0.5\mathcal{B}(9, 1)$.

- The $MISE$ decreases as the sample size increases.

From Figures 1, we conclude that:

- Our proposal (2) is closer to the true distribution function than that Vitale's estimator (3).

- When the sample size increases, we get closer estimation of the true distribution.

# 5 Conlusion

In this paper, we propose an estimator of the distribution function. We compare our proposed estimator to the Vitale's distribution estimator throught simulations, by plotting figures and computing the $MISE$. For all the cases, the $MISE$ of the our proposed estimator (2) with an appropriate choice of the stepsize $(\gamma_n)$ and using the corresponding order $(m_n)$ is smaller than Vitale's estimator introduced in (3). Moreover, the figures show that our estimator $F_n$ is closer to the true distribution function than Vitale's estimator $\tilde{F}_n$. In addition, the proposed estimators have a major advantage is that their update, from a sample of size $n$ to one of size $n+1$, requires less computations than the Vitale's estimator.
In conclusion, using the proposed estimators $F_n$ we obtain better results than those given by the Vitale's distribution estimator. Hence, we plan to work on automatic choice of the order $(m_n)$ through plug-in method and then we can compare the proposed work to the one given in Slaoui (2014b).

# References

Babu, G. J. Canty, A. J. and Chaubey, Y. P. (2002), 'Application of Bernstein polynomials for smooth estimation of a distribution and density function', *Journal of Statistical Planning and Inference*, **105**, 377–392.

Bojanic, R. and Seneta, E. (1973), 'A unified theory of regularly varying sequences', *Mathematische Zeitschrift*, **134**, 91–106.

Galambos, J. and Seneta, E. (1973), 'Regularly varying sequences', *Proceedings of the American Mathematical Society*, **41**, 110–116.

Leblanc, A. (2012), 'On estimationg distribution function using Bernstein polynomials', *Annals of the Institute of Statistical Mathematics*, **64**, 919–943.

Robbins, H and Monro, S. (1951), 'A stochastic approximation method', *Annals of Mathematical Statistics*, **22**, 400–407.

Slaoui, Y. (2014b), 'The stochastic approximation method for the estimation of a distribution function'. *Mathematical Methods of Statistics*, **23**, 306–325.

Vitale, R. A. (1975), 'A Bernstein polynomial approach to density function estimation, *Statistical Inference and Related Topics'*, **2**, 87–99.