Computational Statistics and Data Analysis 84 (2015) 68–84

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Iterated Bernstein operators for distribution function and density estimation: Balancing between the number of iterations and the polynomial degree

Claude Manté *

Aix-Marseille Université, Université du Sud Toulon-Var, CNRS/INSU, IRD, MIO, UM 110, Campus de Luminy, Case 901, F13288 Marseille Cedex 09, France

ARTICLE INFO

Article history:
Received 13 March 2013
Received in revised form 31 October 2014
Accepted 8 November 2014
Available online 15 November 2014

Keywords:
Density estimation
Bernstein operator
Roots of operators
Regular histogram
Shape restriction
Gnedenko test

ABSTRACT

Despite its slow convergence, the use of the Bernstein polynomial approximation is becoming more frequent in Statistics, especially for density estimation of compactly supported probability distributions. This is due to its numerous attractive properties, from both an approximation (uniform shape-preserving approximation, *etc.*) and a statistical (*bona fide* estimation, low boundary bias, *etc.*) point of view. An original method for estimating distribution functions and densities with Bernstein polynomials is proposed, which takes advantage of results about the eigenstructure of the Bernstein operator to refine a convergence acceleration method. Furthermore, an original data-driven method for choosing the degree of the polynomial is worked out. The method is successfully applied to two data-sets which are important benchmarks in the field of Density Estimation.

1. Introduction

Although S. Bernstein simultaneously introduced both the polynomials and the operator that bear his name in his famous constructive proof of the Stone–Weierstrass theorem (Bernstein, 1912), both of these objects naturally split up with time. While there is a large interest in the Bernstein operator in the literature on Approximation Theory, (see for instance Cooper and Waldron, 2000; Sevy, 1993, 1995; Sahai, 2004), researchers from other disciplines essentially focus on Bernstein polynomials. For instance, the attractive properties of this approximation prompted statisticians to apply it to Density Estimation (Vitale, 1975; Babu et al., 2002; Bouezmarni and Rolin, 2007; Leblanc, 2010, 2012a,b), Regression (Marco and Martinez, 2010; Curtis and Ghosh, 2011; Wang and Ghosh, 2012) or Bayesian Inference (Petrone, 1999). However, most of these authors paid little attention to the Bernstein operator itself.

Nevertheless, an operator is attached to a pair of vector spaces, and not to particular bases of these spaces. We highlight in Section 3 that Bernstein polynomials consist in a natural **output** basis for the eponym operator, while the natural **input** basis is a Lagrange polynomial basis (see also Cooper and Waldron (2000), Section 5). In addition, we must take into account the pair of bases associated with the eigendecomposition of the Bernstein operator, given by Cooper and Waldron (2000). Bearing in mind a generalization of the Sevy convergence acceleration method (Sevy, 1993, 1995), we further investigate in Section 3 the matrix representation of powers of the Bernstein operator with respect to these bases. This enables us to define first, in Section 4, fractional Bernstein operators and second, in Section 5, fractional Sevy approximation sequences.

* Tel.: +33 486 090 631; fax: +33 486 090 641.
E-mail address: claude.mante@mio.osupytheas.fr.

http://dx.doi.org/10.1016/j.csda.2014.11.003
0167-9473/

This constitutes the basis for refining results obtained by Manté (2012), where both the distribution function and the density approximation were obtained using Sevy's iteration scheme.

Now, roughly speaking, density estimation or approximation by Bernstein polynomials (Babu et al., 2002; Bouezmarni and Rolin, 2007; Leblanc, 2010, 2012a,b; Manté, 2012) consists in fitting a Bernstein polynomial of some order m on a distribution function, and in differentiating it. More precisely, these authors estimate the distribution function (d.f.) F associated with a random variable X from m values of the empirical distribution function (e.d.f.) F_N obtained from a N-sample of X:

$$\widetilde{F}_{N,m}(x) := \sum_{k=0}^{m} F_N\left(\frac{k}{m}\right) w_{m,k}(x),$$

where $w_{n,j}(x) := \binom{n}{j} x^j (1-x)^{n-j}$. The choice of an optimal number of bins m^* is always a critical step. In the density estimation setting, most authors recommend either choosing $m^* = \nu(N)$, where N is the sample size and ν is some function stemming from asymptotic results (Babu et al., 2002; Leblanc, 2010, 2012a), or else obtaining m^* from cross-validation methods (Bouezmarni and Rolin, 2007; Leblanc, 2010).

In Section 6 we propose another method, starting with Babu et al. (2002) upper value $m_0 := N/\ln(N)$. It consists in selecting $m^* \leq m_0$ in order that the same optimal m^* should be obtained with a high probability from different N-samples (stability), and that the "coarsened" distribution functions associated with these m^* bins should be close to the classical empirical distribution function F_N (fidelity). The method is tested on real data in Section 7.

2. Notation

We will work in the Banach space $C[0,1]$ of continuous functions on $[0,1]$, equipped with the Chebyshev norm $\|f\| := \max_{x\in[0,1]} |f(x)|$. $\mathfrak{P}_n$ denotes the subspace of $C[0,1]$ consisting of polynomials of degree $k \leq n$, and $\overline{\mathfrak{P}_n}$ denotes the complement of $\mathfrak{P}_1$ in $\mathfrak{P}_n$ *i.e.* the vector space of polynomials of degree $1 < k \leq n$.

Consider an operator $U : C[0,1] \to C[0,1]$; for $n \geq 2$ (fixed), its restriction to $\mathfrak{P}_n$ (*i.e.* the operator $U|_{\mathfrak{P}_n} : \mathfrak{P}_n \to C[0,1]$ such that $\forall f \in \mathfrak{P}_n$, $U|_{\mathfrak{P}_n}(f) = U(f)$) will be denoted $\overset{\circ}{U}$, and its restriction to $\overline{\mathfrak{P}_n}$ will be denoted $\overline{U}$. For the sake of simplicity, the restrictions of the identity operator to these subspaces will be denoted 1, instead of $\overset{\circ}{1}$ or $\overline{1}$.

In the finite dimensional setting, we will use the matrix p-norm (or ℓ^p-norm) $\|U\|_p := \sup_{v\neq 0} \frac{\|U(v)\|_p}{\|v\|_p}$ where $\|v\|_p$ is the usual vector ℓ^p-norm. Notice that $\|U\|_1$ and $\|U\|_\infty$ are the greatest sum of the absolute values of the matrix elements along columns and rows, respectively, while $\|U\|_2$ is the spectral norm (Farouki, 1991). In this setting, $Mat(U; L_n, W_n)$ will denote the matrix representation of the operator U with respect to the bases L_n and W_n.

Finally, the expression $Y \overset{\mathfrak{L}}{=} X$ denotes that both of the random variables X and Y obey the same probability law. The integer value of some real number x will be denoted $\lfloor x \rfloor$.

3. Expression of powers of the Bernstein operator into different bases

The Bernstein operator $B_n : C[0,1] \to C[0,1]$ is defined (Cooper and Waldron, 2000; Manté, 2012; Sevy, 1995) by:

$$B_n[f](x) := \sum_{j=0}^{n} w_{n,j}(x) f\left(\frac{j}{n}\right),$$

with $w_{n,j}(x) := \binom{n}{j} x^j (1-x)^{n-j}$. Of course, its image $\mathcal{R}(B_n)$ is included in $\mathfrak{P}_n$. In this section, we will focus on the matrix representation of powers of B_n with respect to three bases of $\mathfrak{P}_n$: Lagrange and Bernstein bases, and the eigenfunctions of B_n.

3.1. Expression of powers of B_n relative to Lagrange and Bernstein bases

First, let us consider the Lagrange interpolation operator $\mathcal{L}_n : C[0,1] \to C[0,1]$, defined by

$$\mathcal{L}_n[f](x) := \sum_{j=0}^{n} \ell_{n,j}(x) f\left(\frac{j}{n}\right),$$

where $\ell_{n,j}(x) := \prod_{\substack{k=0 \\ k\neq j}}^{n} \frac{nx-k}{j-k}$ is the jth Lagrange polynomial in the equally spaced case. Clearly, $\mathcal{R}(\mathcal{L}_n) = \mathfrak{P}_n$ and, since $\mathcal{L}_n$ is idempotent and the Lebesgue constant $\|\mathcal{L}_n\| = \max_{\|f\|\neq 0} \frac{\|\mathcal{L}_n[f]\|}{\|f\|} \sim \frac{2^n}{e\, n \log(n)}$ (see Mills and Smith, 1992) is bounded for any finite n, $\mathcal{L}_n$ is the projection onto $\mathfrak{P}_n$. Consequently, any $f \in C[0,1]$ is the direct sum of two components: $\mathcal{L}_n[f]$ and the "Lagrange residual" $(f - \mathcal{L}_n[f])$.