ACCESSIBLE METHODOLOGIES FOR ESTIMATING DENSITY FUNCTIONS

Serge B. Provost
Department of Statistical and Actuarial Sciences,
The University of Western Ontario, Canada

It is often the case that the moments of a distribution can be readily determined, while its exact density function is mathematically intractable. We show that the density function of a continuous distribution defined on a closed interval can be easily approximated from its exact moments by solving a linear system involving a Hilbert matrix. When sample moments are being used, the same linear system will yield density estimates. A simple formula that is based on an explicit representation of the elements of the inverse of a Hilbert matrix is being proposed as a means of directly determining density estimates or approximants without having to resort to kernels or orthogonal polynomials. As illustrations, density estimates will be determined for the 'Buffalo snowfall' data set and the density of the distance between two random points in a cube will be approximated. Finally, an alternate methodology is proposed for obtaining smooth density estimates from averaged shifted histograms.

INTRODUCTION

A variety of nonparametric techniques are available for estimating density functions; for a comprehensive account, the reader is referred to Izenman (1991). Some estimators, such as the histogram or the frequency polygon (formed by linear interpolation of adjacent mid-bin values in a histogram), fail to provide a smooth functional representation of the underlying density and are significantly affected by bin edge location. Others, such as those that are based on kernels, splines or orthogonal series, rely on concepts which may be too advanced for certain audiences.

Another problem of interest consists of approximating a density function from the exact moments of a given distribution. In this case, one could for instance make use of Edgeworth expansions, Pearson or Johnson curves, linear combinations of orthogonal functions, or the inverse Mellin transform technique. However, these methodologies can prove difficult to implement and their applicability is often subject to restrictive conditions. Moreover, their derivations require a certain level of mathematical sophistication.

We propose a simple unified technique which applies to both density approximation and density estimation problems when the underlying distribution is continuous and confined to a closed interval. Its derivation relies on basic concepts such as polynomials and their properties, the moments of a distribution and the solution of a system of linear equations. The resulting approximants or estimates are in fact represented by a single polynomial whose coefficients are determined by solving a linear system of equations. An explicit representation of the coefficients is also provided, thus allowing for direct evaluation. The polynomial form of the density makes it easy to report as well as amenable to complex calculations. Moreover, for all intents and purposes, the percentiles evaluated from such approximants can be viewed as exact since their accuracy can always be improved by making use of higher degree polynomials. A density estimation technique which is based on the moments of averaged shifted histograms is also discussed. As will be shown, the proposed methodologies consistently produce reasonable estimates.

DENSITY FUNCTION APPROXIMATION BY MEANS OF A SINGLE POLYNOMIAL

Let X be a continuous random variable whose support is the interval $[a, b]$. Its exact density function, $f(x)$, is approximated by an rth degree polynomial, denoted by $p(x)$, which integrates to one over the support of the distribution and whose first r moments coincide with those of X. Thus, for $h = 0, 1, 2, \ldots, r$, we let

$$\begin{aligned}
\textstyle\int_a^b x^h f(x)\mathrm{d}x &= \textstyle\int_a^b x^h p(x)\mathrm{d}x \\
&= \textstyle\int_a^b x^h (\sum_{j=0}^r \alpha_j x^j)\,\mathrm{d}x \\
&= \textstyle\sum_{j=0}^r \alpha_j (b^{j+h+1} - a^{j+h+1})/(j+h+1) \\
&= \boldsymbol{c}'_{h+1}\,\boldsymbol{\alpha}
\end{aligned} \tag{1}$$

where $\boldsymbol{\alpha}$ is an unknown vector whose components α_j are the coefficients of x^j, $j = 0, \ldots, r,$ in the polynomial $p(x)$ and $\boldsymbol{c}'_{h+1}$ is the $(h+1)$th row of C, an $(r+1)\times(r+1)$ Hilbert matrix (see, for instance, Burden and Faires (1997, p. 489)) whose $(h+1, j+1)$th element is $(b^{j+h+1} - a^{j+h+1})/(j+h+1)$, $h = 0, \ldots, r;\ j = 0, \ldots, r$. The vector $\boldsymbol{\alpha}$ is thus obtained by solving the linear system $\boldsymbol{m}_X = C\boldsymbol{\alpha}$ where $\boldsymbol{m}_X = (1, E(X), E(X^2), \ldots, E(X^r))'$ and the approximate density, defined to be zero outside the interval $[a, b]$, is given by $p(x) = \sum_{j=0}^r \alpha_j x^j$ for $a \le x \le b$, with $(\alpha_0, \ldots, \alpha_r)' = C^{-1}\boldsymbol{m}_X$. The approximate cumulative distribution function, $P(x)$, which is easily seen to be equal to 0 when $x < a$, $\sum_{j=1}^{r+1} \alpha_{j-1}\ (x^j - a^j)/j$ when $a \le x < b$, and 1 when $x \ge b$, could be used for instance to obtain confidence intervals or to generate random samples.

When the support of the distribution is the interval $[0, 1]$, the $(h+1, j+1)$th element of $C^{-1} \equiv \Xi$, as given in Pusey (1970, p. 450), is

$$\xi_{h+1,j+1} = \frac{(-1)^{j+h+2}\ (r+j+1)!\ (r+h+1)!}{(j+h+1)\ (j!\ h!)^2 (r-j)!\ (r-h)!}\ , \tag{2}$$

for $h = 0, \ldots, r$ and $j = 0, \ldots, r$. Hence, on letting $Y = (X-a)/(b-a)$ (whose support is the interval $[0, 1]$), and denoting the vector of moments of Y by $\boldsymbol{m}_Y = (1, E(Y), E(Y^2), \ldots, E(Y^r))'$, with

$$E(Y^\ell) = \frac{1}{(b-a)^\ell} \sum_{i=0}^{\ell} \binom{\ell}{i} (-a)^{\ell-i} E(X^i)\ , \tag{3}$$

$\ell = 1, \ldots, r$, one can directly evaluate the density approximant from the following sum:

$$p(x) = \frac{1}{b-a} \sum_{h=0}^{r} (\boldsymbol{\xi}'_{h+1}\,\boldsymbol{m}_Y) \Big(\frac{x-a}{b-a}\Big)^h\ , \tag{4}$$

where $\boldsymbol{\xi}'_{h+1} = (\xi_{h+1,1}, \ldots, \xi_{h+1,r+1})$ denotes the $(h+1)$th row of Ξ. Burden and Faires (1997, Ch. 8) show that $p(x)$ is the least-squares polynomial that minimizes the integrated squared error,

$$\int_a^b (f(x) - p(x))^2 \mathrm{d}x\,, \tag{5}$$

and that such a least-squares approximating polynomial can be expressed as a linear combination of Legendre polynomials. These orthogonal polynomials can be computed by means of a summation formula which is available for example in Devroye (1989), or a recurrence relationship which is given in Sansone (1959, p. 178). The representation of the least-squares approximants given in Equation 4, which is believed to be new, readily lends itself to numerical evaluation. As shown by Alexits (1961, p. 304), the rate of convergence of the supremum of the absolute error, $|f(x) - p(x)|$, depends on $f(x)$ and r, the degree of $p(x)$, via a continuity modulus. More accurate approximants can always be obtained by making use of higher degree polynomials. To illustrate the proposed approach, we will approximate the density function of the distance between two random points in a unit cube.

POLYNOMIALS AS DENSITY ESTIMATES AND APPLICATION CONSIDERATIONS

One can similarly obtain polynomial density estimates by making use of sample moments in lieu of exact moments. As explained below, an extended range (a, b) must be specified in this case. Hall (1982, Theorem 2) addresses the convergence of an equivalent density estimate expressed in terms of Legendre orthogonal polynomials. We apply our density estimation methodology to the set of 63 values of annual snowfall accumulation in Buffalo for the winters 1910/11 to 1972/73. For alternate approaches to density estimation and examples involving the snowfall data set, the reader is referred to Silverman (1986).

One should initially set r, the number of sample moments to be used to at least 15 in order to retain the structure of the distribution in the interval of interest. Then a and b, the end points of the preliminary support of the estimate, are determined by extending the observed range of the distribution. This is being done in order to relegate unwanted fluctuations or undesirable behavior occurring in the tail areas to portions of the extended range which are later discarded: letting α and β be respectively points in the neighborhoods of the minimum and maximum values of the data at which the polynomial estimate obtained from Equation 4 becomes negative, we define our density estimate to be zero outside the interval (α, β). Furthermore, if, for example, as a result of data sparseness, the polynomial estimate were to become negative on subintervals located within the interval (α, β), the density estimate would then be defined as the polynomial truncated to its positive parts. The resulting estimate would have to be normalized in order to obtain a *bona fide* density function, that is, a function that is always nonnegative and integrates to unity.

Just as a given data set can be represented by a variety of histograms by changing bin origin or bin width, the proposed approach can yield numerous polynomial estimates. The points a and b could be selected so that the resulting polynomial estimate shares many of the distinctive features of a histogram deemed to be representative of the underlying density. Some guidelines on bin width selection can be found in Izenman (1991). For example, Freedman and Diaconis (1981) suggested as approximate optimal bin width, twice the interquartile range of the sample divided by the cubic root of the sample size. More ample extensions of the range are indicated for smaller sample sizes and lead to smoother estimates that exhibit only the broader distributional features of the data.

AVERAGED SHIFTED HISTOGRAMS AS PRELIMINARY DENSITY ESTIMATES

To avoid some of the *ad hoc* features of the proposed density estimation methodology, one could for instance make use of the moments of an averaged shifted histogram (ASH) rather than the sample moments. An ASH is formed by averaging several (10 is usually enough) histograms of equal bin widths (chosen here to be approximately optimal) but different bin locations. As explained by Scott (1985), ASH's approximate triangular kernel estimators.

An adaptive histogram with equal bin count, which produces initial density estimates of a similar type will also be introduced. On the basis of the moments evaluated from such histograms, the estimation problem becomes a straightforward approximation problem, and the higher the degree of the approximating polynomial, the more faithfully the features of the preliminary density estimate will be reproduced. Of course, any *bona fide* density estimate could also serve as a preliminary estimate.

REFERENCES

Alexits, G. (1961). *Convergence Problems of Orthogonal Series.* New York: Pergamon Press.

Burden, R. L. and Faires, J. D. (1997). *Numerical Analysis, 6th Ed.* New York: Brooks/Cole.

Devroye, L. (1989). On random variate generation when only moments or Fourier coefficients are known. *Mathematics and Computers in Simulation* 31, 71–89.

Freedman, D. and Diaconis, P. (1981). On the histogram as a density estimator: L_2 theory. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebeite* 57, 453–476.

Hall, P. (1982). Comparison of two orthogonal series methods of estimating a density and its derivatives on an interval. *Journal of Multivariate Analysis* 12, 432–449.

Izenman, A. J. (1991). Recent development in nonparametric density estimation. *Journal of the American Statistical Association* 86, 205–224.

Pusey, M. (1970). Representations of one orthonormal set in terms of another. *Biometrika* 57, 445–450.

Sansone, G. (1959). *Orthogonal Functions.* New York: Interscience Publishers Inc.

Scott, D. W. (1985). Average shifted histograms: effective nonparametric density estimators in several dimensions. *The Annals of Statistics* 13, 1024–1040.

Silverman, B. (1986). *Density Estimation for Statistics and Data Analysis.* London: Chapman and Hall.