Statistical Methodology 26 (2015) 61–71

Contents lists available at ScienceDirect

Statistical Methodology

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

Differentiated logdensity approximants

Serge B. Provost [a], Hyung-Tae Ha [b,*]

[a] Department of Statistical & Actuarial Sciences, The University of Western Ontario, London, Canada, N6A 5B7

[b] Department of Applied Statistics, Gachon University, Sung-Nam, 461-701, South Korea

ARTICLE INFO

Article history:
Received 24 June 2014
Received in revised form 25 December 2014
Accepted 25 February 2015
Available online 20 March 2015

Keywords:
Density approximation
Moments
Rational functions
Logdensity

ABSTRACT

A moment-based density approximation technique whereby the derivative of the logarithm of a density approximant is expressed as a rational function is introduced in this paper. Guidelines for the selection of the polynomial orders of the numerator and denominator are proposed. The coefficients are then determined by solving a system of linear equations. The resulting density approximation, referred to as a differentiated logdensity approximant or $\mathcal{DLA}$, satisfies a differential equation whose explicit solution is provided. It is shown that a unique solution exists when a polynomial is utilized in lieu of a rational function. The proposed methodology is successfully applied to two test statistics and several distributions. It is also explained that the same moment-matching technique can yield density estimates on the basis of sample moments. An example involving a widely analyzed data set illustrates this approach.

1. Introduction

Whereas the exact density functions of certain random quantities of interest are analytically intractable or difficult to obtain in closed forms, their moments can usually be determined with relative ease. Several distribution approximation techniques that are based on the moments or the cumulants of a random variable have been proposed in the statistical literature. These include: Pearson and Johnson curves, respectively described for instance in [25,7], which rely on the first few moments of a distribution; Edgeworth and Gram–Charlier series expansions, see [6,8], the saddlepoint

* Corresponding author. Tel.: +82 31 750 8884; fax: +82 31 753 8828.
E-mail addresses: provost@stats.uwo.ca (S.B. Provost), htha@gachon.ac.kr (H.-T. Ha).

http://dx.doi.org/10.1016/j.stamet.2015.02.005
1572-3127/

approximation, see [23] and the references therein, and alternative forms thereof such as those proposed by Lugannani and Rice [16] and Daniels [4], which are based on the cumulant generating function of a distribution. Related to the problem of approximating tail probabilities is the inverse problem of approximating the quantiles of a distribution, which was addressed for example by Fisher and Cornish [9], Hall [11] and Barndorff-Nielsen and Cox [2]. For a detailed review of such techniques and formal derivations, the reader is referred to Kolassa [14].

A moment-based density approximation technique that relies on orthogonal series expansions was used for instance by Durbin and Watson [5] to determine certain percentiles of their well-known test statistic and Cheah et al. [3] and Tiku [26] to approximate the density functions of noncentral χ^2 and F random variables. This approach was later generalized by Provost [20] who proposed a unifying methodology whereby a density function, $f(x)$, is approximated by an appropriately selected base density function, $b(x)$, that is multiplied by a polynomial whose coefficients are determined by solving a linear system involving the moments associated with $f(x)$ and $b(x)$. Interestingly, another methodology referred to as the inverse Mellin transform technique, which is based on the complex moments of certain types of distributions, provides, in many cases of interest, representations of their exact density functions in terms of generalized hypergeometric functions; for theoretical considerations as well as various applications, see for instance Mathai and Saxena [17] and Provost and Rudiuk [21].

In order to obtain a *bona fide* density approximant on the basis of the integer moments of a continuous distribution, one could reasonably posit a nonnegative representation involving polynomials. In an initial attempt, approximants of the form $k\, e^{p(x)}$ where $p(x)$ is a polynomial could be considered. However, this simple expression cannot even accommodate such basic cases as the beta or gamma densities, which can be respectively expressed as $k_1\, e^{(\alpha-1)\,\ln(x)+(\beta-1)\,\ln(1-x)}$ and $k_2\, e^{-x/\beta+(\alpha-1)\,\ln(x)}$. However, on noticing that the first derivatives of the exponents of e in these as well as other density functions such as the inverse Gaussian, Rayleigh and normal can be expressed as polynomials or ratios thereof, one is led to consider expressions of the form $e^{\int r(x)\,dx}$ as potentially viable representations of the density approximants, where $r(x)$ denotes a ratio of polynomials of degrees ν and δ whose coefficients can be determined by solving a linear system involving recursive relationships between the exact moments of the target distribution. It is established in Section 2 that when $\delta = 0$, the equation system has a unique solution. The resulting density approximants are obtained by solving a differential equation for which an explicit solution is given. Additionally, guidelines are provided for the selection of the degrees ν and δ. The proposed methodology accommodates multimodal distributions and produces accurate *bona fide* density approximations throughout the entire range of the distributions, which is not necessarily the case for certain aforementioned techniques.

It will be assumed that the distributions being approximated are such that they are uniquely determined by their moments. Conditions ensuring that this is the case (the so-called moment problem) are, for example, available in [22]. It should also be pointed out that whenever the cumulants of a distribution are known, its moments can be determined via a recursive relationship that is available, for instance, in [24].

The density approximation methodology that is advocated herein is applied to the sphericity and Wilks' L_{mvc} test statistics in Section 3. Illustrative examples involving a mixture of beta density functions, a triangular distribution and a mixture of normal density functions are presented in Section 4. It is shown in Section 5 that the proposed density approximation technique can also be applied in the context of density estimation; an example involving the Buffalo snowfall data set is presented. Certain computational aspects are discussed in Section 6 which includes some concluding remarks as well.

2. Differentiated logdensity approximants ($\mathcal{DLA}$)

Let the density function and integer moment of order h of a random variable X defined on the interval (α, β) be respectively denoted by $f(x)$ and $\mu_X(h) = E(X^h)$. It is shown in this section that, on representing the derivative of a logdensity approximant as a ratio of polynomials, one can obtain a set of recursive relationships between the moments, which yield the requisite polynomial coefficients.

The proposed density approximants, denoted by $f_{\nu,\delta}(x)$, are assumed to be expressible as

$$f_{\nu,\delta}(x) = k\, e^{\int_\alpha^x r(y)\,dy} \qquad (2.1)$$