CHAPTER 2

DENSITY ESTIMATION AND ORTHOGONAL POLYNOMIALS

[*This chapter is based on the lectures of Professor Serge B. Provost of the Department of Statistical and Acturial Sciences, The University of Western Ontario, Canada.*]

2.0. Abstract

It is often the case that the exact moments of a continuous distribution can be explicitly determined, while its exact density function either does not lend itself to numerical evaluation or proves to be mathematically intractable. The density approximants proposed in this article are entirely specified by the first few moments of a given distribution. First, it is shown that the density functions of random variables confined to closed intervals can be approximated in terms of linear combinations of Legendre polynomials. In an application, the density function of a mixture of two beta distributions is approximated. It is also explained that the density functions of many statistics whose support is the positive half-line can be approximated by means of sums involving Laguerre polynomials; this approach is applied to a mixture of three gamma random variables. It is then shown that density approximants that are based on orthogonal polynomials such as Legendre, Laguerre, Jacobi and Hermite polynomials can be equivalently obtained by solving a linear system involving the moments of a so-called base density function.

2.1. Introduction

This lecture is concerned with the problem of approximating a density function from the theoretical moments (or cumulants) of the corresponding distribution. Approximants of this type can be obtained for instance by making use of Pearson or

Johnson curves [Solomon and Stephens 1978; Elderton and Johnson 1969], or saddlepoint approximations [Reid 1988]. These methodologies can provide adequate approximations in a variety of applications involving unimodal distributions. However, they may prove difficult to implement and their applicability can be subject to restrictive conditions. The approximants proposed here are expressed in terms of relatively simple formulae and apply to a very wide array of distributions; moreover, their accuracy can be improved by making use of additional moments. Interestingly, another technique called the inverse Mellin transform, which is based on the complex moments of certain distributions, provides representations of their exact density functions in terms of generalized hypergeometric functions; for theoretical considerations as well as various applications, the reader is referred to [Mathai and Saxena 1978] and [Provost and Rudiuk 1995].

First, it should be noted that the hth moment of a statistic, $u(x_1, \ldots, x_n)$, whose exact density is unknown, can be determined exactly or numerically by integrating the product $u(x_1, \ldots, x_n)^h \; g(x_1, \ldots, x_n)$ over the range of integration of the $x_i{}'s$ where $g(x_1, \ldots, x_n)$ denotes the joint density of the $x_i{}'s,\ n = 1, 2, \ldots$. Alternatively, the moments of a random variable x can be obtained from the derivatives of its moment-generating function or by making use of a relationship between the moments and the cumulants when the latter are known, see [Smith 1995]. Moments can also be derived recursively as for instance is the case in connection with certain queueing models. Once the moments of a statistic are available, one can often approximate its density function in terms of sums involving orthogonal polynomials. The approximant obtained for nonnegative random variables depends on two parameters that are determined so as to produce the best initial gamma approximation on the basis of the first two moments of the distribution. Furthermore, it was determined that for commonly encountered unimodal distributions, twelve moments usually suffice to produce reasonably accurate approximations.

The approximant proposed for distributions defined on semi-infinite intervals applies to a wide class of statistics which includes those whose asymptotic distribution is chi-square, such as $-2\,ln\,\lambda$ where λ denotes a likelihood ratio statistic, as well as those that are distributed as quadratic forms in normal variables, such as the sample serial covariance. It should be noted that an indefinite quadratic form can be expressed as the difference of two independent nonnegative definite quadratic forms whose cumulants, incidentally, are well-known. As for distributions having compact supports, one has for example the Durbin-Watson statistic, the sample correlation coefficient, as well as many other useful statistics that can be expressed as the ratio of two quadratic forms, as discussed for instance in [Provost and Cheong 2000].

In Section 2.4, we propose a unified approach for approximating density functions, which turns out to be mathematically equivalent to making use of orthogonal polynomials. This semiparametric methodology is also based on the moments of a distribution and only requires solving a linear system involving the moments of a so-called base density function.

Several illustrative examples are presented. For comparison purposes, each of them involves a distribution whose exact density function can be determined. First, the distribution of a mixture of two beta distributions is considered. The approximation technique presented in Section 2.3 is applied to a mixture of three gamma random variables. A mixture of three Gaussian random variables is considered in Section 2.4.

For results on the convergence of approximating sums that are expressed in terms of orthogonal polynomials, the reader is referred to [Sansone 1959], [Alexits 1961], [Devroye and Györfi 1985] and [Jones and Ranga 1998]. Since the proposed methodology allows for the use of a large number of theoretical moments and the functions being approximated are nonnegative, the approximants can be regarded as nearly exact bona fide density functions, and quantiles can thereupon easily be estimated with great accuracy. As well, the polynomial representations of the approximants make them easy to report and amenable to complex calculations.

Up to now, orthogonal polynomials have been scarcely discussed in the statistical literature in connection with the approximation of distributions. This state of affairs might be due to difficulties encountered in deriving moments of high orders or in obtaining accurate results from high degree polynomials. In any case, given the powerful computational resources that are widely available these days, such complications can hardly any longer be viewed as impediments. It should be pointed out that the simple semiparametric technique proposed in Section 2.4 eliminates some of the complications associated with the use of orthogonal polynomials while yielding identical density approximants.

2.2. Approximants Based on Legendre Polynomials

A polynomial density approximation formula is obtained in this section for distributions having compact supports. This approximant is derived from an analytical result stated in [Alexits 1961], which is couched below in statistical nomenclature.

It should be pointed out that no a priori restrictions on the shape of the distribution need to be made in this case.

The density function of a random variable x that is defined on the interval $[-1, 1]$ can be expressed as follows:

$$f_x[x] = \sum_{k=0}^{\infty} \lambda_k P_k[x] \tag{2.2.1}$$

where $P_k[x]$ is a Legendre polynomial of degree k, that is,

$$P_k[x] = \frac{1}{2^k\, k!} \frac{\partial^k}{\partial x^k}\left(-1 + x^2\right)^k = \sum_{i=0}^{\text{Floor}[k/2]} \frac{(-1)^i\, x^{-2\,i+k}\, (-2\,i + 2\,k)!}{2^k\, i!\, (-2\,i + k)!\, (-i + k)!}, \tag{2.2.2}$$

$Floor[k/2]$ denoting the largest integer less than or equal to $k/2$, and

$$\begin{aligned} \lambda_k &= \frac{1 + 2\,k}{2} \sum_{i=0}^{\text{Floor}[k/2]} \frac{(-1)^i\, (-2\,i + 2\,k)!\, \mu_x(-2\,i + k)}{2^k\, i!\, (-2\,i + k)!\, (-i + k)!} \\ &= \frac{1 + 2\,k}{2} P^*{}_k(\omega) \end{aligned} \tag{2.2.3}$$

with $P^*{}_k(x) = P_k(x)$ wherein x^{k-2i} is replaced by the $(k - 2i)$th moment of x:

$$\mu_x[-2\,i + k] = E\,(x^{-2\,i+k}) = \int_{-1}^{1} x^{k-2i} f(x) dx, \tag{2.2.4}$$

see also [Devroye 1989]. Legendre polynomials can also be derived by means of a recurrence relationship, available for example in [Sansone 1959, p. 178]. Given the first n moments of $x, \mu[1], \ldots, \mu[n]$, and setting $\mu[0] = 1$, the following truncated sum denoted by $f_{x_n}(x)$ can be used as a polynomial approximation to $f_x(x)$:

$$f_{x_n}[x] = \sum_{k=0}^{n} \lambda_k P_k(x). \tag{2.2.5}$$

As explained in [Burden and Faires 1997, Chapter 8], this polynomial turns out to be the least-squares approximating polynomial that minimizes $\int_{-1}^{1}(f_x(x) - f_{x_n}(x))^2 dx$, the integrated squared error. As stated in [Rao 1965, p. 106], the moments of any continuous random variable whose support is a closed interval, uniquely determine its distribution, and as shown by [Alexits 1961, p. 304], the rate of convergence of the supremum of the absolute error, $|f_x(x) - f_{x_n}(x)|$, depends on

$f_x(x)$ and n, the degree of $f_{x_n}(x)$, via a continuity modulus. Therefore, more accurate approximants can always be obtained by making use of higher degree polynomials.

We now turn our attention to the more general case of a continuous random variable y defined on the closed interval [a,b], whose kth moment is denoted by

$$\mu_y[k] = E(y^k) = \int_a^b y^k f_y(y)dy, \quad k = 0, 1, \ldots, \tag{2.2.6}$$

where $f_y(y)$ denotes the density function of y. As pointed out in Section 2.1, there exist several alternative methods for evaluating the moments of a distribution when the exact density is unknown. On mapping y onto x by means of the linear transformation

$$x = \frac{2y - (a + b)}{b - a}, \tag{2.2.7}$$

one has the desired range for x, that is, the interval $[-1, 1]$. The jth moment of x, expressed as the expected value of the binomial expansion of $((2y-(a+b))/(b-a))^j$ is then given by

$$\mu_x[j] = \frac{1}{(b-a)^j} \sum_{k=0}^{j} \binom{j}{k} 2^k \mu_y[k](-1)^{j-k} (a+b)^{j-k} \tag{2.2.8}$$

and (2.2.5) can then be used to provide an approximant to the density function of x. On transforming x back to y with the affine change of variables specified in (2.2.7) and noting that $\mathrm{d}x/\mathrm{d}y = 2/(b-a)$, one obtains the following polynomial approximation for the density function of y:

$$f_{y_n}[y] = [2/(b-a)] \sum_{k=0}^{n} \lambda_k P_k\left(\frac{2y-(a+b)}{b-a}\right). \tag{2.2.9}$$

Example 2.2.1. Approximate density of a mixture of beta random variables

Consider a mixture of two equally weighted beta distributions with parameters (3,2) and (2,30), respectively. A fifteenth-degree polynomial approximation was obtained from (2.2.9). The exact density function of this mixture and its approximant, both plotted in Figure 2.1, are manifestly in close agreement. Obviously, approaches that are based on a few moments would fail to provide satisfactory approximations in this case.

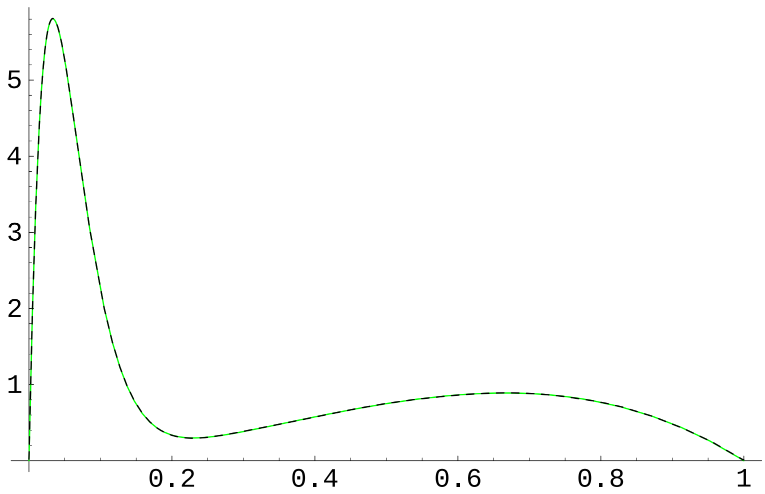

Figure 2.1: **Exact & Approximate (dashed) PDF's**

As will be mentioned in Section 2.4, beta-shaped density functions defined on closed intervals can be approximated in terms of Jacobi polynomials. However, it should be pointed out that approximants expressed in terms of Legendre polynomials can accomodate a much wider class of distributions defined on closed intervals.

2.3. Approximants Based on Laguerre Polynomials

As mentioned in Section 2.1, the density functions of numerous statistics distributed on the positive half-line can be approximated from their exact moments by means of sums involving Laguerre polynomials. It should be pointed out that such an approximant should only be used when the underlying distribution possesses the tail behaviour of a gamma random variable; thankfully, this is often the case for statistics whose support is semi-infinite. Note that for other types of distributions whose support is the positive half-line, such as the lognormal, the moments may not uniquely determine the distribution; see for instance [Rao 1965, p. 106] for conditions ensuring that they do.

Consider a random variable y defined on the interval (a, ∞), whose jth moment is denoted by $\mu_y[j],\ j = 0, 1, 2, \ldots$, and let

$$c = \frac{-\mu_y[1]^2 + \mu_y[2]}{-a + \mu_y[1]} \tag{2.3.1}$$

$$v = \frac{\mu_y[1] - a}{c} - 1 \tag{2.3.2}$$

and

$$x = \frac{y - a}{c}. \tag{2.3.3}$$

As will be explained later, when the parameters c and v are so chosen, the leading term of the resulting approximating sum will in fact be a gamma density function whose first and second moments agree with those of y. Note that although a can be any finite real number, it is often equal to zero. Denoting the jth moment of x by

$$\mu_x[j] = E[((y - \mu)/c)^j], \tag{2.3.4}$$

the density function of the random variable x defined on the interval $(0, \infty)$ can be expressed as

$$f_x[x] = x^v e^{-x} \sum_{j=0}^{\infty} \delta_j L_j(v, x) \tag{2.3.5}$$

where

$$L_j[v, x] = \sum_{k=0}^{j} \frac{(-1)^k \, \Gamma(1 + j + v) \, x^{j-k}}{\Gamma(1 + j - k + v) \, (j - k)! \, k!} \tag{2.3.6}$$

is a Laguerre polynomial of order j with parameter v and

$$\delta_j = \sum_{k=0}^{j} \frac{(-1)^k \, j! \, \mu_x[j - k]}{\Gamma\,(1 + j - k + v) \, (j - k)! \, k!} \tag{2.3.7}$$

which also can be represented by $j!/\Gamma(v+j+1)$ times $L_j[v, x]$ wherein x^k is replaced with $\mu_x[k]$, see for example [Szegö 1959] and [Devroye 1989]. Then, on truncating the series given in (2.3.5) and making the change of variables $y = cx+a$, one obtains the following density approximant for y:

$$f_{y_n}[y\_] = \frac{(y - a)^v e^{-(y-a)/c}}{c^{v+1}} \sum_{j=0}^{n} \delta_j L_j(v, (y - a)/c).$$

Remark 2.3.1. On observing that $f_{y_0}(y)$ is a shifted gamma density function with parameters $\alpha \equiv v+1 = (\mu[1]-a)^2/(\mu[2]-\mu[1]^2)$ and $\beta \equiv c = (\mu[2]-\mu[1]^2)/(\mu[1]-a)$, one can express $f_{y_n}(x)$ as the product of an initial shifted gamma density approximation whose first two central moments, $\alpha\beta + a = \mu[[1]]$ and $\alpha\beta^2 = \mu[[2]] - \mu[[1]]^2$, match those of y, times a polynomial adjustment; that is,

$$f_{y_n}(y) = \frac{e^{\frac{a-y}{\beta}}\,(-a+y)^{-1+\alpha}}{\beta^{\alpha}\,\Gamma[\alpha]} \sum_{j=0}^{n} a_j\, L_j\, \Gamma(\alpha)(\alpha-1,(y-a)/c). \qquad (2.3.8)$$

The following example is relevant as nonnegative definite quadratic forms in normal variables which happen to be ubiquitous in Statistics can be expressed as mixtures of chi-square random variables, see for instance [Mathai and Provost 1992, Chapters 2 & 7].

Example 2.3.1. **Approximate density of a mixture of gamma random variables**

Let the random variable y be a mixture of three equally weighted shifted gamma random variables with parameters $(\alpha_1 = 8, \beta_1 = 1), (\alpha_2 = 16, \beta_2 = 1)$ and $(\alpha_3 = 64, \beta_3 = 1/2)$, all defined on the interval $(5, \infty)$. The hth moment of this distribution is determined by evaluating the hth derivative of its moment-generating function, $M_x(t)$, with respect to t at $t = 0$.

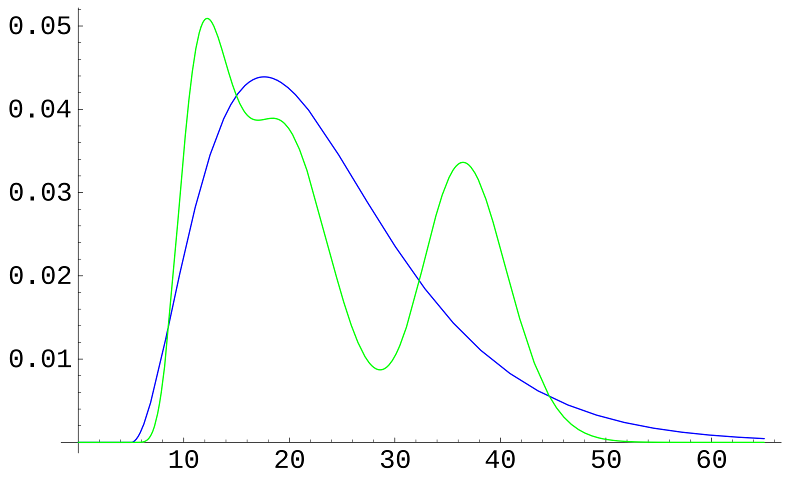

Figure 2.2: **Exact Density & Initial Gamma Approximant**

Figure 2.2 shows the exact density function of the mixture as well as the initial gamma density approximation given by $f_{y_0}[y]$. Clearly, traditional approximants such as those mentioned in Section 2.1, could not capture adequately all the distinctive features of this particular distribution.

The exact density function, $f_y[y]$ and its polynomial approximant, $f_{y_{60}}[y]$, are plotted in Figure 2.3. (Once such an approximant is obtained, one could for instance approximate it with a spline composed of third-degree polynomial arcs, in order to reduce the degree of precision required in further calculations.)

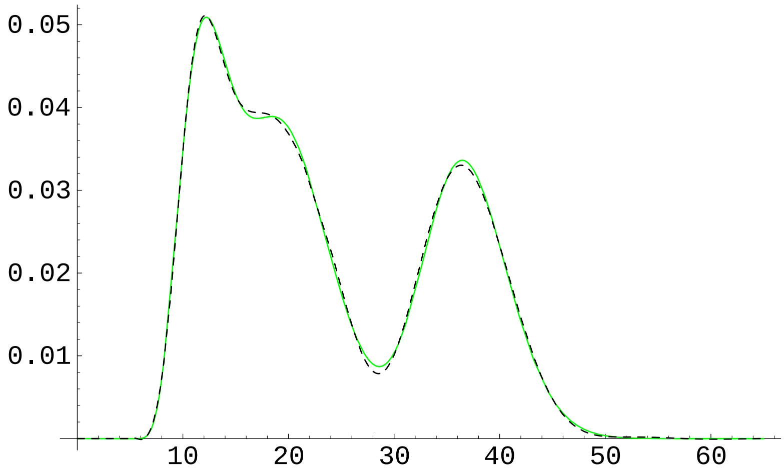

Figure 2.3: **Exact & Approximate (dashed line) PDF's**

This example illustrates that the proposed methodology can also accomodate multimodal distributions and that calculations involving high order Laguerre polynomials will readily produce remarkably accurate approximations when performed in an advanced computing environment such as that provided by Mathematica.

2.4. A Unified Methodology

The remark made in the previous section suggests the following general semi-parametric approach to density approximation, which consists of approximating the density function of a distribution whose first n moments are known by means of a base density function whose parameters are determined by matching moments, times a polynomial of degree n whose coefficients are also obtained by matching moments.

Result 2.4.1 Let $f_y[y]$ be the density function of a continuous random variable y defined on the interval (a,b), $E(y^j) \equiv \mu_y[j]$, $x = (y-u)/s$ where u and s are constants, $a_0 = (a-u)/s, b_0 = (b-u)/s$, $f_x[x] = |s| f_y[u+sx], E(x^j) = E[((y-u)/s)^j] \equiv \mu_x[j]$, and the base density function, $\psi_x[x]$, be an initial density approximant for x

defined on the interval $(a_0,\ b_0)$, whose jth moment, $m_x[j]$, exists for $j = 1, \cdots, 2n$. Assuming that the tail behavior of $\psi_x[x]$ is similar to that of $f_x[x]$, $f_x[x]$ can be approximated by

$$f_{x_n}[x] = \psi_x[x] \sum_{i=0}^{n} \xi_i x^i \tag{2.4.1}$$

with $(\xi_0, \ldots, \xi_n)' = M^{-1}(\mu[0], \ldots, \mu_x[n])'$ where M is an $(n+1)\times(n+1)$ matrix whose $(h+1)$th row is $(m_x[h], \ldots, m_x[h+n])'$, $h = 0, 1, \ldots, n$, and whenever $\xi_x[x]$ depends on r parameters, these are determined by matching $m_x[j]$ to $\mu_x[j]$ for $j = 1, \ldots, r$. The corresponding density for y is then

$$f^*_{y_n}[y] = f_{x_n}[(y-u)/s]/s. \tag{2.4.2}$$

The coefficients ξ_i, can easily be determined by equating the first n moments obtained from $f_{x_n}[x]$ to those of x :

$$\int_{a_0}^{b_0} x^h \psi_x[x] \sum_{i=0}^{n} \xi_i x^i dx = \int_{a_0}^{b_0} x^h f[x] dx,\ h = 0, 1, \ldots, n, \tag{2.4.3}$$

which is equivalent to

$$(m_x[h], \ldots, m_x[h+n]).(\xi_0, \ldots, \xi_n) = \mu_x[h],\ h = 0, 1, \ldots, n; \tag{2.4.4}$$

this linear system can be represented in matrix form as

$$M\,(\xi_0, \ldots, \xi_n)' = (\mu_x[0], \ldots, \mu_x[n])'$$

where M is as defined in Result 2.4.1.

2.5. Approximants Expressed in terms of Orthogonal Polynomials

By making use of the same notation, we now show that the unified approach described above provides approximants that are mathematically equivalent to those obtained from orthogonal polynomials whose weights are proportional to a certain base density function.

Let $\{T_i[x] = \sum_{k=0}^{i} \delta_{ik} x^k,\ i = 0, 1, \ldots, n\}$ be a set of orthogonal polynomials on the interval (a_0, b_0) such that

$$\int_{a_0}^{b_0} w[x]T_i[x]T_h[x]dx = \theta_h \text{ when } i = h,\ h = 0, 1, \ldots, n, \text{ and zero otherwise,} \tag{2.5.1}$$

where $w[x]$ is a weight function, and let c_T be a normalizing constant such that $c_T w[x] \equiv \psi_x[x]$ integrates to one over the interval (a_0, b_0). On noting that the orthogonal polynomials T_i are linearly independent (Burden and Faires (1993), Corollary 8.8), one can write (2.4.1) as

$$f_{x_n}[x] = c_T w[x] \sum_{i=0}^{n} \eta_i T_i p[x] \tag{2.5.2}$$

where the η_i's are obtained from equating $\int_{a_0}^{b_0} T_h[x] f_{x_n}[x]dx$ to $\int_{a_0}^{b_0} T_h[x] f[x]dx$ for $h = 0, 1, \ldots, n$, which yields the following linear system:

$$c_T \int_{a_0}^{b_0} T_h[x]w[x] \sum_{i=0}^{n} \eta_i T_i[x]dx = \int_{a_0}^{b_0} T_h[x] f[x]dx,\ h = 0, 1, \ldots, n, \tag{2.5.3}$$

which is equivalent to

$$\sum_{i=0}^{n} \eta_i c_T \int_{a_0}^{b_0} w[x]T_i[x]T_h[x]dx = \sum_{k=0}^{h} \delta_{hk}\mu_x[k],\ h = 0, 1, \ldots, n, \tag{2.5.4}$$

where δ_{hk} is the coefficient of x^k in T_h. Thus, by virtue of the orthogonality property given in (2.5.1), one has

$$\eta_h = \frac{1}{c_T \theta_h} \sum_{k=0}^{h} \delta_{hk}\mu_x[k],\ h = 0, 1, \ldots, n, \tag{2.5.5}$$

and

$$f_{x_n}[x] = \psi[x] \sum_{i=0}^{n} \left(\frac{1}{c_T \theta_i} \sum_{k=0}^{i} \delta_{ik}\mu_x[k]\right)T_i[x]. \tag{2.5.6}$$

Now, letting $y = u + sx$, $a = u + sa_0$, $b = u + sb_0$, and denoting the density functions of y and y_n corresponding to those of x and x_n by $f_y[y]$ and $f_{y_n}[y]$, respectively, $f_y[y]$ whose support is the interval (a,b) can be approximated by

$$f_{y_n}[y] = w[(y - u)/s] \sum_{i=0}^{n} \left(\frac{1}{s\theta_i} \sum_{k=0}^{i} \delta_{ik}\mu_x[k]\right)T_i[(y - u)/s]. \tag{2.5.7}$$

It is apparent that several complications associated with the use of orthogonal polynomials can be avoided by resorting to the direct approach described in Result 2.4.1. Density approximants expressed in terms of Laguerre, Legendre, Jacobi and Hermite polynomials are discussed below.

2.5.1. Approximants based on Laguerre polynomials

Consider the approximants based on Laguerre polynomials discussed in Section 2.3. In that case, $y = cx + a$, so that $u = a, s = c, a_0 = 0, b_0 = \infty, w[x] = x^v e^{-x}, T_i[x]$ is the Laguerre $L_i[v, x]$ orthogonal polynomial and $\theta_h = Gamma[v + h + 1]/h!$. It is easily seen that the density expressions given in (2.5.7) and (2.3.8) coincide.

In this case, the base density function, $\psi_x[x]$ is that of a gamma random variable with parameters $v + 1$ and 1. Note that after the transformation, our base density is a shifted gamma distribution with parameters $v + 1$ and c, whose support is the interval (a, ∞).

Alternatively, one can obtain an identical density approximant by making use of Result 2.4.1 where in $\psi_x[x]$ is a $Gamma(\lambda + 1, 1)$ density function whose jth moment, $m_x[j]$, which is needed to determine the ξ's, is given by $Gamma[v + 1 + j]/Gamma[v + 1]$, $j = 0, 1, \ldots, 2n$.

2.5.2. Approximants based on Legendre polynomials

First, we note that whenever the finite interval (a, b) is mapped onto the interval (a_0, b_0), the requisite affine transformation is

$$x = \frac{y - u}{s} \tag{2.5.8}$$

with $u = (ab_0 - a_0 b)/(b_0 - a_0)$ and $s = (b - a)/(b_0 - a_0)$.

Consider the approximants based on Legendre polynomials discussed in Section 2.2, which are defined on the interval $(-1, 1)$. In that case, $u = \frac{(a+b)}{2}$, $s = (b - a)/2$, $w[x] = 1$, $T_i[x]$ is the Legendre orthogonal polynomial $P_i[x]$ and $\theta_h = \frac{2}{(2h+1)}$. It is easily seen that $f_{y_n}[y]$ given in (2.5.7) yields $f_{y_n}[y]$ of (2.2.9).

2.5.3. Approximants based on Jacobi polynomials

In order to approximate densities for which a beta type density is suitable as a base density, we shall make use of the following alternative form of the Jacobi polynomials

$$G_n[\alpha,\beta,x] = n!\frac{Gamma[n+\alpha]}{Gamma[2n+\alpha]}JacobiP[n,\alpha-\beta,\beta-1,2x-1] \tag{2.5.9}$$

defined on the interval (0, 1), where $JacobiP[n,a_1,b_1,z]$ denotes a standard Jacobi polynomial of order n in z with parameters a_1 and b_1. In this case, the weight function is $x^\alpha(1-x)^\beta$ and the base density is that of a $Beta(\alpha+1,\beta+1)$ random variable, that is,

$$\psi_x[x] = \frac{1}{Beta[\alpha+1,\beta+1]}x^\alpha(1-x)^\beta,\ 0<x<1, \tag{2.5.10}$$

whose jth moment is given by

$$m_x[j] = \frac{Gamma[\alpha+\beta+2]Gamma[\alpha+1+j]}{Gamma[\alpha+1]Gamma[\alpha+j+\beta+2]}. \tag{2.5.11}$$

The parameters α and β can be determined as follows:

$$\begin{aligned}\alpha &= \mu_x[1](\mu_x[1]-\mu_x[2])/(\mu_x[2]-\mu_x[1]^2)-1,\\ \beta &= (1-\mu_x[1])(\alpha+1)/(\mu_x[1]-1),\end{aligned} \tag{2.5.12}$$

see Johnson and Kotz (1970). Moreover, in this case,

$$\theta_k^{-1} = \frac{(2k+a+b+1)Gamma[2k+a+b+1]^2}{k!Gamma[k+a+1]Gamma[k+a+b+1]Gamma[k+b+1]}. \tag{2.5.13}$$

2.5.4. Approximants based on Hermite polynomials

Densities of random variables for which a normal density can provide a reasonable initial approximation can be expressed in terms of the modified Hermite polynomials given by

$$H_k^*[x] = (-1)^k 2^{-k/2} HermiteH[k, x\sqrt{2}] \tag{2.5.14}$$

where $HermiteH[k,z]$ denotes a standard Hermite polynomial of order k in z. $H_k^*[x]$ is also defined on the interval $(-\infty,\infty)$, and its associated weight function is $w[x] =$

$e^{-x^2/2}$. Clearly, $x = (y - u)/s$ with $u = \mu_y[1]$ and $s = \sqrt{\mu_y[2] - \mu_y[1]^2}$. In this case, the base density is that of a standard normal random variable, that is,

$$\psi_x[x] = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \qquad -\infty < x < \infty, \tag{2.5.15}$$

whose jth moment is given by

$$m_x[j] = \frac{2^{\frac{1}{2}(-1+j)}(1 + (-1)^j) Gamma[\frac{1+j}{2}]}{\sqrt{2\pi}}, \quad j = 0, 1, \ldots, \tag{2.5.16}$$

and

$$\theta_k = \sqrt{2\pi}\, k!. \tag{2.5.17}$$

Example 2.5.1. Consider an equally weighted mixture of a $N(3, 4)$ and a $N(1, 1)$ distributions. The exact and approximate densities are shown in Figure 2.4.

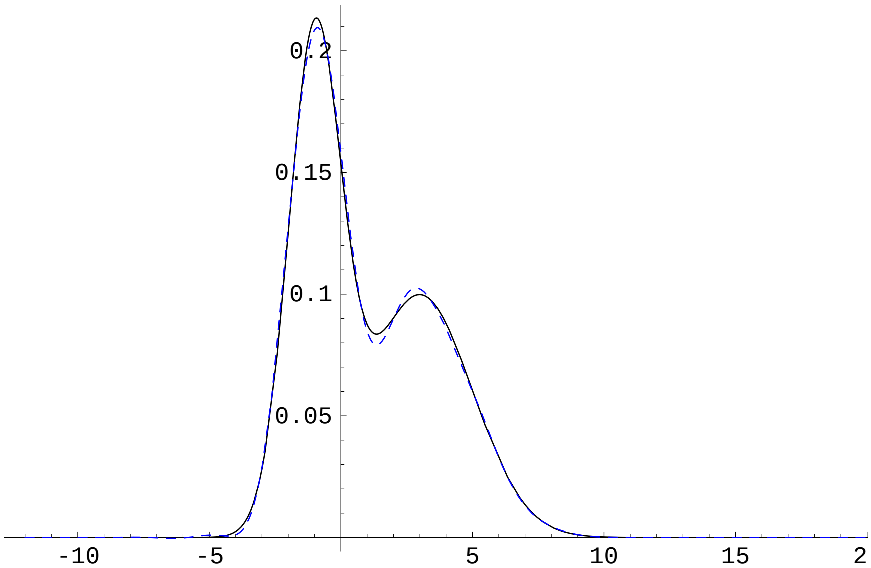

Figure 2.4: **Exact & Approximate (dashed line) PDF's**

References

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

Barndorff-Nielsen, O. and Cox, D. R., 1979. Edgeworth and Saddlepoint Approximations with Statistical Applications. *Journal of the Royal Statistical Society, Series B*, **41,** 279-312.

Burden, R. L. and Faires, J. D., 1993. *Numerical Analysis*, 5th Ed. Brooks/ Cole, New York.

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

Devroye, L. and Györfi, L., 1985. *Nonparametric Density Estimation*, The L_1 View. Wiley, New York.

Elderton, W. P. and Johnson, N. L., 1969. *Systems of Frequency Curves*. Cambridge University Press.

Johnson, N. L. and Kotz, S., 1970. *Distributions in Statistics: Continuous Univariate Distribu tions-2,* Wiley, New York

Jones, W. B. and Ranga, A. S., Eds., 1998. *Orthogonal Functions, Moment Theory, and Continued Fractions, Theory and Applications*. Marcel Dekker, New York.

Mathai, A. M. and Provost, S. B., 1992. *Quadratic Forms in Random Variables, Theory and Applications*. Marcel Dekker, New York.

Mathai, A. M. and Saxena, R. K., 1978. *The H-function with Applications in Statistics and Other Disciplines*. John Wiley and Sons, New York.

Provost, S. B. and Rudiuk, E. M., 1995. Moments and densities of tests statistics for covariance structures. *International Journal of Mathematical and Statistical Sciences*, **4,** 85-104.

Provost, S. B. and Cheong, Y.-H., 2000. On the distribution of linear combinations of the components of a Dirichlet random vector. *sl The Canadian Journal of Statistics*, **28,** 417-425. Rao, C. R., 1965. *Linear Statistical Inference and Its*

Applications. Wiley, New York.

Reid, N., 1988. Saddlepoint methods and statistical inference. *Statistical Science*, **3,** 213-227.

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

Smith, P. J., 1995. A recursive formulation of the old problem of obtaining moments from cumulants and vice versa. *The American Statistician*, **49,** 217-219.

Solomon, H. and Stephens, M., 1978. Approximations to density functions using Pearson curves. *Journal of the American Statistical Association*, **73,** 153-160.

Szegö, G., 1959. *Orthogonal Polynomials*. American Mathematical Society, New York.