

Moment-Based Density Approximants

Serge B. Provost

It is often the case that the exact moments of a statistic of the continuous type can be explicitly determined, while its density function either does not lend itself to numerical evaluation or proves to be mathematically intractable. The density approximants discussed in this article are based on the first n exact moments of the corresponding distributions. A unified semiparametric approach to density approximation is introduced. Then, it is shown that the resulting approximants are mathematically equivalent to those obtained by making use of certain orthogonal polynomials, such as the Legendre, Laguerre, Jacobi, and Hermite polynomials. Several examples illustrate the proposed methodology.

■ 1. Introduction

This article is concerned with the problem of approximating a density function from the moments (or cumulants) of a given distribution. Approximants of this type can be obtained, for example, by making use of Pearson or Johnson curves, see [1, 2, 3], or saddlepoint approximations as discussed in [4]. 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 in this article are expressed in terms of relatively simple formulae and apply to a very wide array of distributions. Moreover, their accuracy can be improved by 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, see [5, 6].

First, it should be noted that the b th moment of a statistic, $u(X_1, \dots, X_n)$, whose exact density is unknown, can be determined exactly or numerically by integrating the product $u(x_1, \dots, x_n)^b g(x_1, \dots, x_n)$ over the range of integration of the x_i 's, where $g(x_1, \dots, x_n)$ denotes the joint density of the X_i 's, $i = 1, 2, \dots, n$; for instance, this approach is used in Example 1. Alternatively, the moments of a random variable X can be obtained from the derivatives of its moment-generating function as is done in Example 4 or by making use of a relationship between the moments and the cumulants when the latter are known [7]. Moments can

also be derived recursively as is the case, for instance, in connection with certain queueing models. When the moments of a statistic uniquely determine its distribution and a sufficient number of moments are known, we can often approximate its density function in terms of sums involving orthogonal polynomials of a certain type. Conveniently, such polynomials are available as built-in *Mathematica* functions.

Density approximants based on Legendre and Laguerre polynomials are discussed in Sections 2 and 3, respectively, for random variables having finite and semi-infinite supports. The main formulae which allows for the *direct* evaluation of the density approximants are equations (15), derived in Section 2, and (29), obtained in Section 3. The approximant that is expressed in terms of Laguerre polynomials 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 are well known. As for distributions having compact supports, there are, for example, the Durbin-Watson statistic, Wilks' likelihood ratio criterion, the sample correlation coefficient, as well as many other useful statistics that can be expressed as the ratio of two quadratic forms, as discussed in [8].

In Section 4, we propose a unified density estimation methodology which only requires the moments of the distribution to be approximated and those of a suitable 'base density function'. As it turns out, this approach yields density approximants that are identical to those obtained from certain orthogonal polynomials—namely, the Legendre, Laguerre, Jacobi, and Hermite polynomials—whose associated weight function is proportional to the corresponding base density function. Several examples illustrate the various results. The *Mathematica* code utilized for implementing the main formulae and plotting the graphs is supplied in the Appendix.

For results in connection with the convergence of approximating sums that are expressed in terms of orthogonal polynomials, see [9, 10, 11, 12]. 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 representations of the approximants make them easy to report and amenable to complex calculations.

Until now, orthogonal polynomials have been scarcely discussed in the statistical literature in connection with the approximation of distributions. This might have been 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 be viewed as impediments any longer. It should be pointed out that the simple semiparametric technique proposed in Section 4 eliminates some of the complications associated with the use of orthogonal

polynomials while yielding identical density approximants. This article is self-contained, and the results presented herein potentially have a host of applications. Being that the subject matter of this article is density approximation as opposed to density estimation, it ought to be emphasized that the techniques presented herein are meant to be used in conjunction with exact moments rather than sample moments.

■ 2. Approximants Based on Legendre Polynomials

A polynomial density approximation formula which applies to any continuous distribution having a compact support is obtained in this section. This approximant is derived from an analytical result stated in [10], which is couched in statistical nomenclature in this section.

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

$$f_X[x] = \sum_{k=0}^{\infty} \lambda_k P_k[x], \tag{1}$$

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

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

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

$$\lambda_k = \frac{2k + 1}{2} \sum_{i=0}^{\text{Floor}[k/2]} (-1)^i 2^{-k} \frac{(2k - 2i)!}{i!(k - i)!(k - 2i)!} \mu_X[k - 2i] = \tag{3}$$

$$\frac{2k + 1}{2} P_k^*[\omega]$$

with $P_k^*[X] = P_k[X]$, wherein X^{k-2i} is replaced by the $(k - 2i)$ th moment of X :

$$\mu_X[k - 2i] = E(X^{k-2i}) = \int_{-1}^1 x^{k-2i} f_X[x] dx, \tag{4}$$

[13]. Legendre polynomials can also be obtained by means of a recurrence relationship, which is derived for instance in [9, 178]. Given the first n moments of X , $\mu_X[1], \dots, \mu_X[n]$, and setting $\mu_X[0] = 1$, the following truncated series denoted by $f_{X_n}[x]$ can be used as a polynomial approximation to $f_X[x]$:

$$f_{X_n}[x] = \sum_{k=0}^n \lambda_k P_k[x], \tag{5}$$

that is,

$$f_{X_n}[x] := \sum_{k=0}^n \left(\frac{2k+1}{2} \text{LegendreP}[k, X] /. X^j \rightarrow \mu_X[j] \right) \text{LegendreP}[k, x] \quad (6)$$

in *Mathematica* notation, where the pattern matching symbol \rightarrow (which is typed \Rightarrow in Input mode) conveniently replaces each occurrence of X^j in $\text{LegendreP}[k, X]$ with $\mu_X[j]$. It should be noted that expressions involving $:=$ (excluding the punctuation at the end of the formulae) can readily be used in a *Mathematica* notebook.

As explained in [14, 439], this polynomial turns out to be the least-squares approximating polynomial of degree n that minimizes the integrated squared error, that is, $\int_{-1}^1 (f_X[x] - f_{X_n}[x])^2 dx$. As stated in [15, 106], the moments of any continuous random variable whose support is a closed interval uniquely determine its distribution. Moreover, as shown by [10, 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 via a continuity modulus. It follows that 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 which is defined on the closed interval $[a, b]$. We denote its density function by $f_Y(y)$ and its k th moment by

$$\mu_Y[k] = E(Y^k) = \int_a^b y^k f_Y[y] dy, \quad k = 0, 1, \dots \quad (7)$$

As pointed out in the Introduction, alternative methods are available for evaluating the moments of a distribution when the exact density is unknown. On mapping Y onto X by means of the linear transformation

$$X = \frac{2Y - (a+b)}{b-a}, \quad (8)$$

we obtain the desired range for X , that is, the interval $[-1, 1]$. The j th moment of X , which is obtained as the expected value of the binomial expansion of $((2Y - (a+b))/(b-a))^j$ is given by

$$\mu_X[j] = \frac{1}{(b-a)^j} \sum_{k=0}^j \binom{j}{k} 2^k \mu_Y[k] (-1)^{j-k} (a+b)^{j-k}, \quad (9)$$

that is,

$$\mu_X[j_] := \text{Expand} \left[\left(\frac{2Y - (a+b)}{b-a} \right)^j \right] /. \text{Table}[Y^k \rightarrow \mu_Y[k], \{k, 0, j\}] \quad (10)$$

or equivalently

$$\mu_X[j_] := \text{Expand} \left[\left(\frac{2Y - (a+b)}{b-a} \right)^j \right] /. Y^j \rightarrow \mu_Y[j]. \quad (11)$$

Equation (6) 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 variable specified in equation (8) and noting that $\partial X / \partial Y = 2 / (b - a)$, we obtain the following approximate density function for Y :

$$f_{Y_n}[y] = \frac{2}{b-a} \sum_{k=0}^n \lambda_k P_k \left[\frac{2y - (a+b)}{b-a} \right], \tag{12}$$

that is,

$$f_{Y_n}[y_{\cdot}] := \sum_{k=0}^n \left(\frac{2k+1}{b-a} \text{LegendreP}[k, X] /. X^{j_{\cdot}} \rightarrow \mu_X[j] \right) \text{LegendreP} \left[k, \frac{2y - (a+b)}{b-a} \right] \tag{13}$$

in *Mathematica* notation. On combining equations (9) and (13), one obtains the following compact representation of the density approximant:

$$f_{Y_n}[y_{\cdot}] := \sum_{k=0}^n \frac{2k+1}{b-a} \left(\text{LegendreP}[k, X] /. X^{j_{\cdot}} \rightarrow \sum_{k=0}^j \frac{j! 2^k \mu_Y[k] (-a-b)^{j-k}}{(b-a)^j k! (j-k)!} \right) * \text{LegendreP} \left[k, \frac{2y - (a+b)}{b-a} \right]. \tag{14}$$

Now, observing that $\text{LegendreP}[k, X]$ with X^j replaced by $\mu_X[j]$ as given earlier, is equivalent to $\text{Legendre}[k, (2Y - (a+b))/(b-a)]$ with Y^j replaced by $\mu_Y[j]$, we also obtain

$$f_{Y_n}[y_{\cdot}] := \sum_{k=0}^n \frac{2k+1}{b-a} \left(\text{LegendreP} \left[k, \frac{2Y - (a+b)}{b-a} \right] /. Y^{j_{\cdot}} \rightarrow \mu_Y[j] \right) \text{LegendreP} \left[k, \frac{2y - (a+b)}{b-a} \right]. \tag{15}$$

Thus, given $\mu_Y[k]$, $k = 1, 2, \dots, n$, the first n moments of a random variable defined on the interval $[a, b]$, an n th-degree polynomial approximation of its density function can be directly obtained from equations (14) or (15).

It should be noted that the density approximants so obtained may be negative on certain subranges of the support of their distributions having low density. This will likely occur if an insufficient number of moments are being used. However, by mere inspection of the approximate density plot, we should be able to determine whether a higher degree polynomial ought to be used. Indeed, owing to the convergence of the approximant, the density function will converge everywhere to a nonnegative number as more moments are being used. If we wish to obtain a truly *bona fide* density function, we could always take a normalized function, $\zeta_{Y_n}[\cdot]$, which is initially defined as being equal to $f_{Y_n}[\cdot]$ except on subintervals where the latter is negative, wherein it is set equal to zero.

In the following application, a polynomial approximation is obtained for the density of V , the square of the distance between two points that are randomly distributed in the unit cube.

□ **Example 1: Exact and Approximate Density Functions of V**

Let $X = (X_1, X_2, X_3)$ and $Y = (Y_1, Y_2, Y_3)$ be two points in the unit cube whose coordinates, X_i and Y_i , $i = 1, 2, 3$, are all independently and uniformly distributed in the interval $[0, 1]$, and let V denote the square of the distance between these two random points, that is, $(Y_1 - X_1)^2 + (Y_2 - X_2)^2 + (Y_3 - X_3)^2$, whose support is the interval $[0, 3]$.

The closed-form representation of the density function of V that follows, which is believed to be original, was derived from the integral representations obtained by [16, Section 2.6.4] by making use of certain trigonometric identities as well as some of *Mathematica*'s algebraic simplification routines:

$$f_V(v) = d(v) I_{[0,1]}(v) + g(v) I_{[1,2]}(v) + s(v) I_{[2,3]}(v), \quad (16)$$

where the functions $d(v)$, $g(v)$, and $s(v)$ can be easily identified from the expression given in the Appendix for $f_V(v)$, and $I_A(v)$ denotes the indicator function which is equal to one whenever v belongs to the set A and zero otherwise.

The b th moment of V can be evaluated by integrating $v^b f_V(v)$ whenever $b \geq -1/2$. Alternatively, if the density function of V were not known, we could determine its b th moment from the following integral representation:

$$\int_0^1 \int_0^1 \int_0^1 \int_0^1 \int_0^1 \int_0^1 ((y_1 - x_1)^2 + (y_2 - x_2)^2 + (y_3 - x_3)^2)^b dx_1 dx_2 dx_3 dy_1 dy_2 dy_3, \quad (17)$$

whose evaluation can be handled by *Mathematica*. However, on noting that the density function of $\delta_j = (y_j - x_j)^2$ is $((1/\sqrt{\delta_j}) - 1) I_{[0,1]}(\delta_j)$, it is much more efficient to compute the b th moment of V as follows:

$$\int_0^1 \int_0^1 \int_0^1 (\delta_1 + \delta_2 + \delta_3)^b (\delta_1^{-1/2} - 1) (\delta_2^{-1/2} - 1) (\delta_3^{-1/2} - 1) d\delta_1 d\delta_2 d\delta_3. \quad (18)$$

Figure 1 shows the exact probability density function (PDF) of V obtained from equation (16) (solid line) superimposed on its thirteenth-degree polynomial approximation (dashed line) evaluated from equation (15) (or equivalently equation (13)) with $a = 0$ and $b = 3$. The exact and approximate cumulative distribution functions (CDFs) which can easily be evaluated by integration, appear in Figure 2, and their difference is plotted in Figure 3. The code that was used for plotting the graphs and evaluating the various functions is provided in the Appendix. In general, note that in order to avoid round-off errors, it is advisable to carry out the calculations with rational numbers. In this example, the moments already are in rational form. When this is not the case, the command `Rationalize[.]` can be used to obtain rational representations.

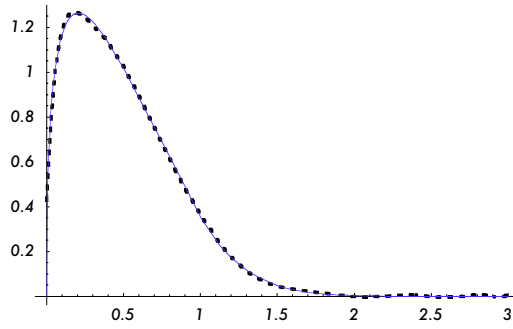


Figure 1. Exact and approximate (dashed line) PDFs. [Pq or Pq1 in the Appendix]

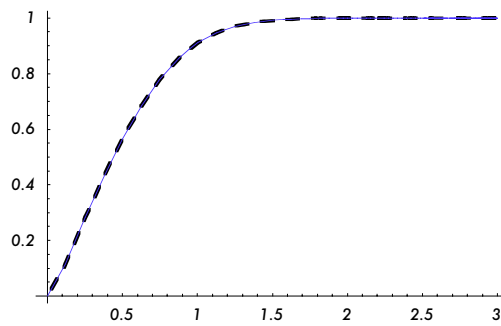


Figure 2. Exact and approximate (dashed line) CDFs. [PQ in the Appendix]

As Figure 3 indicates, the exact and approximate CDFs differ by less than 0.001 over the interval $[0, 3]$ in this case.

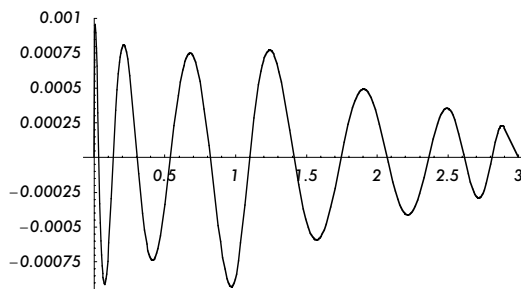


Figure 3. The difference between the exact and approximate CDFs. [Qd in the Appendix]

□ Example 2: 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 the compact formula given in equation (14). The exact density function of this mixture and its approximant, both plotted in Figure 4, are manifestly in close agreement. A glance at the *Mathematica* code that is provided in the Appendix for this example should convince the reader that very little programming is indeed required. Clearly, methodologies that are based on only a few moments would fail to provide satisfactory approximations in this case.

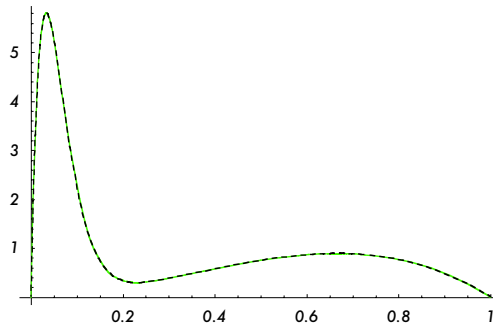


Figure 4. Exact and approximate (dashed line) PDFs. [Pb in the Appendix]

As specified in Section 4, approximants that are expressed in terms of Jacobi polynomials are ideally suited for approximating beta-type density functions. However, in the absence of prior knowledge about the shape of a density function, it is indicated to make use of approximants based on Legendre polynomials as they can theoretically accommodate any continuous distribution defined on a closed interval. It should be pointed out that if a density function turns out to be very irregular, a prohibitive number of moments might be required to approximate it satisfactorily. Thankfully, the majority of continuous distributions of interest are smooth and possess at most a few modes.

■ 3. Approximants Based on Laguerre Polynomials

As pointed out in the Introduction, 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 approximants should only be used when the underlying distribution possesses the tail behaviour of a gamma random variable. Fortunately, this is often the case for test statistics whose support is semi-infinite. Note that for other types of distributions defined on the positive half-line, such as the

lognormal which is considered in Example 3, the moments may not uniquely determine the distribution; see [15, 106] for conditions ensuring that they do.

Consider a random variable Y defined on the interval $[a, \infty)$, whose j th moment is denoted by $\mu_Y[j]$, $j = 0, 1, 2, \dots$, and let

$$c = \frac{\mu_Y[2] - \mu_Y[1]^2}{\mu_Y[1] - a}, \tag{19}$$

$$v = \frac{\mu_Y[1] - a}{c} - 1, \tag{20}$$

and

$$X = \frac{Y - a}{c}. \tag{21}$$

As explained in Remark 3.1, when the parameters c and v are so chosen, the leading term of the approximant is a shifted gamma density function whose mean and variance agree with those of Y . Although a can be any finite real number, it is in most cases of interest equal to zero. By definition, c belongs to \mathbb{R}^+ , the set of positive real numbers. Denoting the j th moment of X by

$$\mu_X[j] = E\left[\left(\frac{Y - a}{c}\right)^j\right], \tag{22}$$

that is,

$$\mu_X[j] := \sum_{b=0}^j \text{Expand}\left[\left(\frac{Y - a}{c}\right)^j\right] /. Y^{h_} \rightarrow \mu_Y[b], \tag{23}$$

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

$$f_X[x] = x^v e^{-x} \sum_{j=0}^{\infty} \delta_j L_j(v, x), \tag{24}$$

where

$$L_j[v, x] = \sum_{k=0}^j (-1)^k \frac{\Gamma(v + j + 1) x^{j-k}}{k! (j - k)! \Gamma(v + j - k + 1)} \tag{25}$$

is a Laguerre polynomial of order j in x with parameter v , that is, $\text{LaguerreL}[j, v, x]$ in *Mathematica* notation and

$$\delta_j = \sum_{k=0}^j (-1)^k \frac{j!}{k! (j - k)! \Gamma(v + j - k + 1)} \mu_X[j - k], \tag{26}$$

which also can be represented by $j! / \Gamma(v + j + 1)$ times $L_j[v, X]$, wherein X^k is replaced with $\mu_X[k]$, [13, 17]. Then, on truncating the series given in

equation (24) and making the change of variable $Y = cX + a$, we obtain the following density approximant for Y :

$$f_{Y_n}[y] = \frac{(y-a)^v e^{-(y-a)/c}}{c^{v+1}} \sum_{j=0}^n \delta_j L_j\left(v, \frac{y-a}{c}\right), \quad (27)$$

that is,

$$f_{Y_n}[y] := \frac{(y-a)^v e^{-(y-a)/c}}{c^{v+1}} \sum_{j=0}^n \frac{j!}{\text{Gamma}[v+j+1]} * (\text{LaguerreL}[j, v, X] /. X^k \rightarrow \mu_X[k]) \text{LaguerreL}\left[j, v, \frac{y-a}{c}\right], \quad (28)$$

where $\text{Gamma}[\cdot] \equiv \Gamma[\cdot]$ denotes the gamma function or, observing that $\text{LaguerreL}[j, v, X]$ with X^k replaced by $\mu_X[k]$ as defined earlier, is equivalent to $\text{LaguerreL}[j, v, (Y-a)/c]$ with Y^k replaced by $\mu_Y[k]$, we obtain

$$f_{Y_n}[y] := \frac{(y-a)^v e^{-(y-a)/c}}{c^{v+1}} \sum_{j=0}^n \frac{j!}{(v+j)!} * (\text{LaguerreL}\left[j, v, \frac{Y-a}{c}\right] /. Y^k \rightarrow \mu_Y[k]) * \text{LaguerreL}\left[j, v, \frac{y-a}{c}\right], \quad (29)$$

where c and v are defined in equations (19) and (20), respectively. It should be noted that the representation of the approximant appearing in equation (29) does not require the evaluation of $\mu_X[k]$, $k = 0, 1, \dots, n$.

Remark 3.1 Note that $f_{Y_0}[y]$ is a shifted gamma density function with parameters $(\alpha \equiv v+1 = (\mu_Y[1]-a)^2)/(\mu_Y[2]-\mu_Y[1]^2)$ and $(\beta \equiv c = \mu_Y[2]-\mu_Y[1]^2)/(\mu_Y[1]-a)$ whose mean, $\alpha\beta + a = \mu_Y[1]$, and variance, $\alpha\beta^2 = \mu_Y[2]-\mu_Y[1]^2$, match those of Y and that, in light of equation (27), we can express $f_{Y_n}[y]$ as the product of an initial shifted gamma density approximation specified by $f_{Y_0}[y]$ times a polynomial adjustment. That is,

$$f_{Y_n}[y] = \frac{(y-a)^{\alpha-1} e^{-(y-a)/\beta}}{\beta^\alpha \Gamma[\alpha]} \sum_{j=0}^n \omega_j L_j\left(\alpha-1, \frac{y-a}{\beta}\right), \quad (30)$$

where $\omega_j = \Gamma[\alpha] \delta_j$.

□ Example 3: The Case of the Standard Lognormal Distribution

As pointed out at the beginning of this section, the proposed methodology is contraindicated when a distribution is not uniquely defined by its moments or when its tail behaviour is not that of a gamma random variable. A case in point is the lognormal distribution. As shown in Figure 5, if we employ the methodology outlined in this section, a very crude approximation of the CDF of the standard lognormal distribution is obtained on the basis of its first three moments. When additional moments are being used, the resulting density approximants turn out to be unusable.

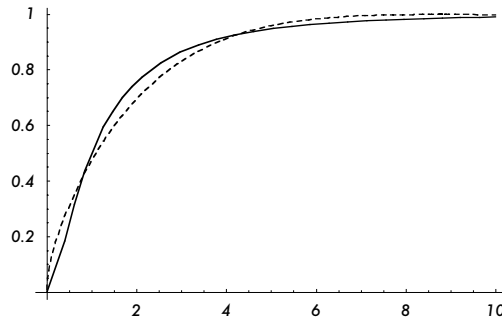


Figure 5. Exact and approximate (dashed line) CDFs. [CDFLN in the Appendix]

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, [18, Chapters 2, 7].

□ **Example 4: A Mixture of Gamma Random Variables**

Let the random variable Y be a mixture of three equally weighted, shifted gamma random variables defined on the interval $(5, \infty)$ with parameters $(\alpha_1 = 8, \beta_1)$, $(\alpha_2 = 16, \beta_2 = 1)$, and $(\alpha_3 = 64, \beta_3 = 1/2)$. The density and moment-generating functions of Y are given in the Appendix by $f_Y[y]$ and $\text{MGF}_Y[t]$, respectively. The b th moment of this distribution, denoted by $\mu_Y[b]$, is determined by evaluating the b th derivative of $\text{MGF}_Y[t]$ with respect to t at $t = 0$.

Figure 6 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 which make use of three or four moments could not capture adequately all the distinctive features of this particular distribution.

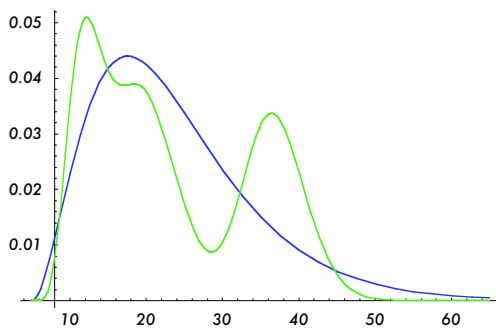


Figure 6. Exact density function and initial gamma approximant. [PGE in the Appendix]

The exact density function, $f_Y[y]$, and its approximant, $f_{Y_{60}}[y]$, evaluated from equation (29), are plotted in Figure 7. As pointed out in Remark 3.1, this density approximant results from a polynomial adjustment applied to the shifted gamma density specified by $f_{Y_0}[y]$. (Once such an approximant is obtained, a spline could be fitted in order to reduce the degree of precision that would be required in subsequent calculations.)

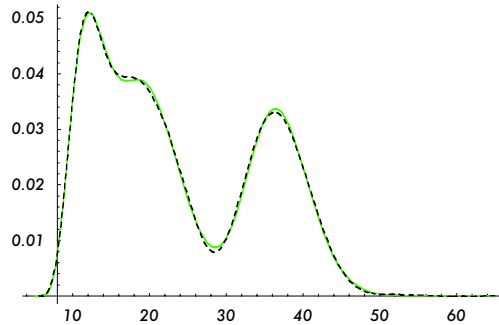


Figure 7. Exact and approximate (dashed line) PDFs. [PDEA in the Appendix]

This example illustrates that the proposed approximation formulae can also accommodate 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*.

■ 4. A Unified Methodology

Remark 3.1 suggests that the exact density function associated with a distribution whose first n moments are known can be approximated by means of the product of a base density function, whose parameters are determined by matching moments, and a polynomial of degree n , whose coefficients are obtained by making use of the method of moments as well. This general semiparametric approach to density approximation, which incidentally does not rely on orthogonal polynomials, is formally described in the following result.

Result 4.1 Let $f_Y[y]$ be the density function of a continuous random variable Y defined in the interval (a, b) , $E(Y^j) \equiv \mu_Y[j]$, $X = (Y - u)/s$, where $u \in \mathbb{R}$ and $s \in \mathbb{R}^+$, $a_0 = (a - u)/s$, $b_0 = (b - u)/s$, $f_X[x] = s f_Y[u + s x]$ denote the density function of X whose support is the interval (a_0, b_0) , $E(X^j) = E[((Y - u)/s)^j] \equiv \mu_X[j]$, and let the base density function $\psi_X[x] \equiv c_T w[x]$, where c_T is a positive normalizing constant, be an initial density approximant to $f_X[x]$ with $\int_{a_0}^{b_0} x^j \psi_X[x] dx \equiv m_X[j]$. Assuming that $\mu_X[i]$, $i = 0, 1, 2, \dots$, uniquely define the distribution of X , that $m_X[j]$ exists for

$j = 0, 1, \dots, 2n$, and that whenever $\psi_X[x]$ is a nontrivial function of x , its tail behaviour is congruent to that of $f_X[x]$, the latter can be approximated by

$$f_{X_n}[x] = \psi_X[x] \sum_{\ell=0}^n \xi_\ell x^\ell \tag{31}$$

with $(\xi_0, \dots, \xi_n)' = M^{-1}(\mu_X[0], \dots, \mu_X[n])'$, where M is an $(n + 1) \times (n + 1)$ matrix whose $(b + 1)$ th row is $(m_X[b], \dots, m_X[b + n])$, $b = 0, 1, \dots, n$. When $\psi_X[x]$ depends on r parameters, these are determined by equating $m_X[j]$ to $\mu_X[j]$, $j = 1, \dots, r$. The corresponding density approximant for Y is then

$$f_{Y_n}[y] = \psi_X\left[\frac{y-u}{s}\right] \sum_{\ell=0}^n \frac{\xi_\ell}{s} \left(\frac{y-u}{s}\right)^\ell. \tag{32}$$

This last formula can easily be coded as follows:

$$\begin{aligned} f_{Y_n}[y_] := & \psi_X\left[\frac{y-u}{s}\right] \\ & (\text{Inverse}[\text{Table}[m_X[b+i], \{b, 0, n\}, \{i, 0, n\}], \text{Table}[\mu_X[j], \{j, 0, n\}]] \\ & \frac{\text{Table}[(\frac{y-u}{s})^j, \{j, 0, n\}]}{s}. \end{aligned}$$

We now show that the polynomial coefficients ξ_ℓ can be determined by making use of the method of moments, that is, by equating the first n moments obtained from $f_{X_n}[x]$ to those of X :

$$\int_{a_0}^{b_0} x^b \psi_X[x] \sum_{i=0}^n \xi_i x^i dx = \int_{a_0}^{b_0} x^b f_X[x] dx, \quad b = 0, 1, \dots, n, \tag{33}$$

which is equivalent to $(m_X[b], \dots, m_X[b + n]) \cdot (\xi_0, \dots, \xi_n) = \mu_X[b]$, $b = 0, 1, \dots, n$, where

$$\mu_X[b] = \frac{1}{s^b} \sum_{k=0}^b \binom{b}{k} \mu_Y[k] (-u)^{b-k} = \text{Expand}\left[\left(\frac{Y-u}{s}\right)^b\right] /. Y^k \rightarrow \mu_Y[k], \tag{34}$$

or $M(\xi_0, \dots, \xi_n)' = (\mu_X[0], \dots, \mu_X[n])'$, that is, $(\xi_0, \dots, \xi_n)' = M^{-1}(\mu_X[0], \dots, \mu_X[n])'$, where M is as defined in Result 4.1.

Remark 4.1 Note that it is not always necessary to transform the random variable Y . The transformation is, for instance, convenient for establishing that the proposed methodology yields density approximants identical to those obtained in terms of certain orthogonal polynomials. If there exists a base density function, $\psi_Y[y]$, whose support is the interval (a, b) and whose tail behaviour is congruent to that of Y when $\psi_Y[y]$ is a nontrivial function of y , then its parameters can be determined by equating $m_Y[j]$ to $\mu_Y[j]$ for $j = 1, \dots, r$, and $f_{Y_n}[y] = \psi_Y[y] \sum_{\ell=0}^n \xi_\ell y^\ell$ with $(\xi_0, \dots, \xi_n)' = M^{-1}(\mu_Y[0], \dots, \mu_Y[n])'$, where M is an $(n + 1) \times (n + 1)$ matrix whose $(b + 1)$ th row is $(m_Y[b], \dots, m_Y[b + n])$, $b = 0, 1, \dots, n$. Alternatively, in that case, one may set $u = 0$ and $s = 1$ in Result 4.1.

□ Connection to Approximants Expressed in Terms of Orthogonal Polynomials

We now show that the unified methodology described in Result 4.1 provides approximants that are mathematically equivalent to those obtained from orthogonal polynomials whose associated weight function is proportional to a base density function. In addition, an alternative representation of the ξ_i 's is given in terms of c_T , the moments of X , and quantities characterizing the type of orthogonal polynomials corresponding to the selected base density function. A general representation of the coefficients η_i in the linear combination of orthogonal polynomials T_i specified by equation (36) is also derived.

Let $\{T_i[x] = \sum_{k=0}^i \delta_{ik} x^k, i = 0, 1, \dots, n\}$ be a set of orthogonal polynomials defined on the interval (a_0, b_0) , which satisfy the following orthogonality property:

$$\int_{a_0}^{b_0} w[x] T_i[x] T_b[x] dx = \theta_b \text{ when } i = b, \quad (35)$$

$$b = 0, 1, \dots, n, \text{ and zero otherwise,}$$

where $w[x]$ is a weight function, and let c_T be a normalizing constant such that $c_T w[x] \equiv \psi_X[x]$ (the base density function defined in Result 4.1) integrates to one over the interval (a_0, b_0) . On noting that the orthogonal polynomials T_i are linearly independent [14, Corollary 8.7], we can write equation (31) as

$$f_{X_n}[x] = c_T w[x] \sum_{i=0}^n \eta_i T_i[x], \quad (36)$$

where, in light of equation (33) and the fact that orthogonal polynomials are linear combinations of powers of x , the η_i 's can be obtained from equating $\int_{a_0}^{b_0} T_b[x] f_{X_n}[x] dx$ to $\int_{a_0}^{b_0} T_b[x] f[x] dx$ for $b = 0, 1, \dots, n$. This yields the following equalities:

$$c_T \int_{a_0}^{b_0} T_b[x] w[x] \sum_{i=0}^n \eta_i T_i[x] dx = \int_{a_0}^{b_0} T_b[x] f_X[x] dx, \quad b = 0, 1, \dots, n. \quad (37)$$

Equivalently, we obtain

$$\sum_{i=0}^n \eta_i c_T \int_{a_0}^{b_0} w[x] T_i[x] T_b[x] dx = \sum_{k=0}^b \delta_{hk} \mu_X[k], \quad b = 0, 1, \dots, n, \quad (38)$$

where δ_{hk} is the coefficient of x^k in $T_b[x]$ or $\delta_{hk} = \text{CoefficientList}[T_b[x], x][[k + 1]]$. Thus, by virtue of the orthogonality property of the $T_i[\cdot]$'s specified by equation (35), we obtain the following general representation for the coefficients η in equation (36):

$$\eta_b = \frac{1}{c_T \theta_b} \sum_{k=0}^b \delta_{hk} \mu_X[k], \quad b = 0, 1, \dots, n, \quad (39)$$

and

$$f_{X_n}[x] = \psi[x] \sum_{i=0}^n \left(\frac{1}{c_T \theta_i} \sum_{k=0}^i \delta_{ik} \mu_X[k] \right) T_i[x]. \tag{40}$$

Now, since $T_i[x] = \sum_{\ell=0}^i \delta_{i\ell} x^\ell$ and $\sum_{i=0}^n \sum_{\ell=0}^i \equiv \sum_{\ell=0}^n \sum_{i=\ell}^n$, it follows that the coefficients ξ_ℓ in equation (31) correspond to the expression in parentheses in the following representation of $f_{X_n}[x]$:

$$f_{X_n}[x] = \psi[x] \sum_{\ell=0}^n \left(\sum_{i=\ell}^n \frac{\delta_{i\ell}}{c_T \theta_i} \sum_{k=0}^i \delta_{ik} \mu_X[k] \right) x^\ell. \tag{41}$$

Now, letting $Y = u + sX$, $Y_n = u + sX_n$, $a = u + sa_0$, $b = u + sb_0$, and denoting by $f_Y[y]$ and $f_{Y_n}[y]$ the density functions of Y and Y_n corresponding to those of X and X_n , respectively, we can approximate $f_Y[y]$ whose support is the interval (a, b) by

$$f_{Y_n}[y] = w \left[\frac{y-u}{s} \right] \sum_{i=0}^n \left(\frac{1}{s \theta_i} \sum_{k=0}^i \delta_{ik} \mu_X[k] \right) T_i \left[\frac{y-u}{s} \right], \tag{42}$$

where $\sum_{k=0}^i \delta_{ik} \mu_X[k] \equiv T_i[X] / X^{i-} \rightarrow \mu_X[j] \equiv T_i[(Y-u)/s] / Y^{i-} \rightarrow \mu_Y[j]$ or equivalently

$$f_{Y_n}[y_-] := \psi \left[\frac{y-u}{s} \right] \sum_{\ell=0}^n \left(\sum_{i=\ell}^n \frac{\delta_{i\ell}}{s c_T \theta_i} \sum_{k=0}^i \delta_{ik} \mu_X[k] \right) \left(\frac{y-u}{s} \right)^\ell, \tag{43}$$

which corresponds to the representation of $f_{Y_n}[y]$ given in equation (32).

It should be pointed out that Reinking proposed under a somewhat different setup general formulae for approximating density and distribution functions in terms of Laguerre, Jacobi, and Hermite polynomials [19]. Arguably, the approach proposed in Result 4.1 is not only conceptually simpler than that which is based on orthogonal polynomials, but it is also more general. The particular cases of density approximants expressed in terms of Laguerre, Legendre, Jacobi, and Hermite polynomials, which can all be equivalently obtained via Result 4.1, are individually considered in the remainder of this section.

Approximants Based on Laguerre Polynomials

Consider the approximants based on the Laguerre polynomials which were defined in Section 3. In that case, $u = a$, $s = c$, $Y = cX + a$, $a_0 = 0$, $b_0 = b = \infty$, $w[x] = x^v e^{-x}$, $c_T = 1/\text{Gamma}[v+1]$, $T_i[x]$ is a LaguerreL $[i, v, x]$ orthogonal polynomial which is defined on the interval $[0, \infty)$, and θ_b , as specified by equation (35), is equal to $\text{Gamma}[v+b+1]/b! = \text{Pochhammer}[b+1, v]$. It is then easily seen that the density expressions given in equations (42) and (28) 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 applying the transformation

$Y = cX + a$, the base density becomes a shifted gamma distribution with parameters $v + 1$ and c whose support is the interval $[a, \infty)$.

Alternatively, we can obtain an identical density approximant by making use of Result 4.1, wherein $\psi_X[x]$ is a gamma density function with parameters $v + 1$ and 1, whose j th moment $m_X[j]$ which is needed to determine the ξ_i 's, is given by $\text{Gamma}[v + 1 + j] / \text{Gamma}[v + 1] = \text{Pochhammer}[v + 1, j]$, $j = 0, 1, \dots, n$.

For instance, Laguerre series expansions were respectively derived by [20] and [21] for approximating the density functions of quadratic forms in normal variables and those of noncentral χ^2 and F random variables.

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

$$X = \frac{Y - u}{s} \quad (44)$$

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

Now, considering the approximants based on Legendre polynomials discussed in Section 2, which are defined on the interval $[-1, 1]$, we have $u = (a + b) / 2$, $s = (b - a) / 2$, $w[x] = 1$, $c_T = 1 / 2$, $T_i[x]$ is a LegendreP $[i, x]$ orthogonal polynomial, and $\theta_b = 2 / (2b + 1)$. It is then easily seen that $f_{Y_n}[y]$ as given in equation (42) yields the density function of Y_n specified in equation (13). The same density approximant can also be obtained by making use of Result 4.1. In this case, $m_X[j] = (1 - (-1)^{j+1}) / (2(j + 1))$, $j = 0, 1, \dots$.

Approximants Based on Jacobi Polynomials

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

$$G_{n-}[\sigma_-, \tau_-, x_-] := n! \frac{\text{Gamma}[n + \sigma]}{\text{Gamma}[2n + \sigma]} \text{JacobiP}[n, \sigma - \tau, \tau - 1, 2x - 1] \quad (45)$$

defined on the interval $[0, 1]$, wherein $\text{JacobiP}[n, a_1, b_1, w]$ denotes an n th-degree Jacobi polynomial in the variable w with parameters a_1 and b_1 , $\sigma = \alpha + \beta + 1$, and $\tau = \alpha + 1$. In this case, the support of Y , the random variable of interest, is the finite interval $[a, b]$; thus $u = a$ and $s = b - a$ in the linear transformation leading to equation (42). The associated weight function is $w[x] = x^\alpha (1 - x)^\beta$ so that $c_T = 1 / \text{Beta}[\alpha + 1, \beta + 1]$, and the base density is that of a $\text{Beta}(\alpha + 1, \beta + 1)$ random variable, that is,

$$\psi_X[x] := \frac{1}{\text{Beta}[\alpha + 1, \beta + 1]} x^\alpha (1 - x)^\beta, \quad 0 < x < 1, \quad (46)$$

whose j th moment is given by

$$m_X[j] = \frac{\text{Gamma}[\alpha + \beta + 2] \text{Gamma}[\alpha + 1 + j]}{\text{Gamma}[\alpha + 1] \text{Gamma}[\alpha + j + \beta + 2]} = \frac{\text{Pochhammer}[\alpha + 1, j]}{\text{Pochhammer}[\alpha + \beta + 2, j]}, \quad j = 0, 1, \dots \tag{47}$$

The parameters α and β can be determined as follows:

$$\alpha = \mu_X[1] \frac{\mu_X[1] - \mu_X[2]}{\mu_X[2] - \mu_X[1]^2} - 1, \tag{48}$$

$$\beta = (1 - \mu_X[1]) \frac{\alpha + 1}{\mu_X[1]} - 1,$$

see [22, 44] and, in this case,

$$\theta_k^{-1} = \frac{(2k + \alpha + \beta + 1) \text{Gamma}[2k + \alpha + \beta + 1]^2}{k! \text{Gamma}[k + \alpha + 1] \text{Gamma}[k + \alpha + \beta + 1] \text{Gamma}[k + \beta + 1]}, \tag{49}$$

[19]. As illustrated in the next example, Result 4.1, as well as equations (42) and (43), yields identical density approximants.

Jacobi polynomial expansions were used by Durbin and Watson to approximate certain percentiles of their well-known test statistic [23].

□ **Example 5**

Example 1 is revisited by making use of a beta density as a base density function in Result 4.1 or equivalently by resorting to an approximant expressed in terms of Jacobi polynomials. Since the base density function already provides a good approximation in this case, fewer moments are required than in the case of the Legendre polynomial approximant (eight as opposed to 13) in order to achieve a similar degree of accuracy.

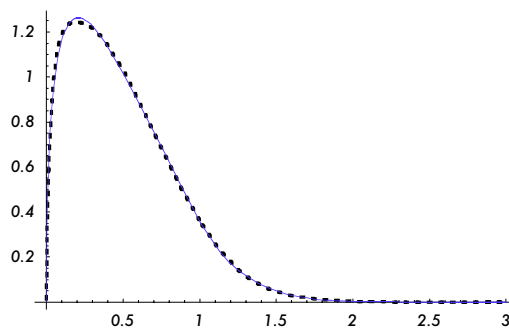


Figure 8. Exact and approximate (dashed line) PDFs. [S5 in the Appendix]

□ Example 6

Consider Wilks' likelihood ratio statistic, $\Lambda = |S_e|^{N/2} / |S_e + S_b|^{N/2}$, where S_e is the error sum of squares matrix and S_b is the hypothesis sum of squares matrix, which is used for testing linear hypotheses on regression coefficients on the basis of N p -dimensional observation vectors. Assuming standard normal theory, S_e and S_b are independent Wishart matrices. As shown in [24, Section 8.4], when the mean vectors are assumed to be of the form BZ_α , where $B = (B_1, B_2)$ with B_i of dimension $p \times q_i$, $i = 1, 2$, and the Z_α 's are given $(q_1 + q_2)$ -dimensional vectors, $\alpha = 1, \dots, N$, the b th moment of the statistic $U = \Lambda^{2/N}$ for testing the null hypothesis that B_1 is a given matrix is

$$\prod_{i=1}^p \frac{\text{Gamma}\left[\frac{N-q_1-q_2+1-i}{2} + b\right] \text{Gamma}\left[\frac{N-q_2+1-i}{2}\right]}{\text{Gamma}\left[\frac{N-q_1-q_2+1-i}{2}\right] \text{Gamma}\left[\frac{N-q_2+1-i}{2} + b\right]}. \quad (50)$$

Clearly, the support of Λ , and thus that of U , is the interval $(0, 1)$. As pointed out by Mathai [25] who obtained a general representation of the exact density function of U as an inverse Mellin transform, a wide array of techniques were used to determine the density of U for some particular values of N , q_1 , q_2 and p . Result 4.1 provides a means of obtaining approximants that can be viewed as exact for all intents and purposes. In any case, many of the so-called exact representations available in the literature are expressed in terms of integrals which have to be evaluated numerically or in terms of infinite series that have to be truncated.

Figures 9 and 10 show the approximate density and distribution functions of U for $N = 12$, $q_1 = 2$, $q_2 = 4$, and $p = 2$ when one lets $n = 2$ (dashed line) and $n = 20$ (solid line). Mathai [25] determined that for $N - q_1 - q_2 = 6$, $q_1 = 2$, and $p = 2$, the 95th and 99th percentiles of the distribution are respectively 0.87825 and 0.94719. As evaluated in the Appendix, $F1_{Y_{20}}[0.87825] = 0.950001$ and $F1_{Y_{20}}[0.94719] = 0.990001$, where $F1_{Y_n}[w]$ denotes a percentile corresponding to the point w , which is approximated on the basis of n moments. It is seen from the graph that an adequate approximant can be obtained from two moments in this case. In fact, $F1_{Y_2}[0.87825] = 0.9507$ and $F1_{Y_2}[0.94719] = 0.9903$.

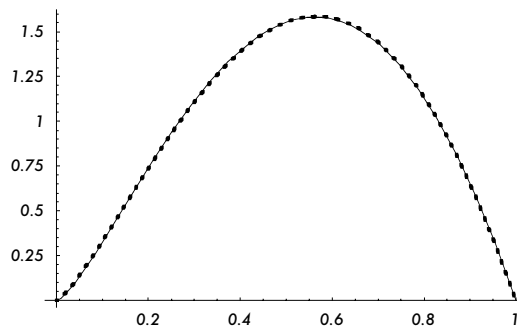


Figure 9. PDF of U for $n = 2$ (dashed line) and $n = 20$. [P6 in the Appendix]

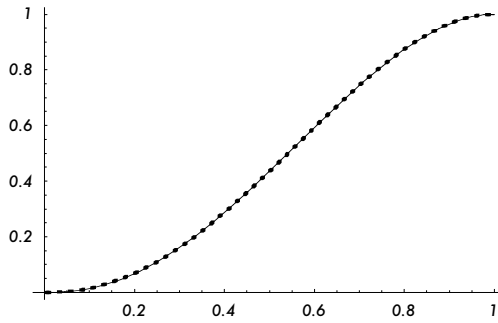


Figure 10. CDF of U for $n = 2$ (dashed line) and $n = 20$. [C6 in the Appendix]

Approximants Based on Hermite Polynomials

We can approximate densities whose tail behaviour is congruent to that of a normal density function by means of the modified Hermite polynomials given by

$$H^*_{k_}[x_]:=(-1)^k 2^{-k/2} \text{HermiteH}\left[k, \frac{x}{\sqrt{2}}\right] \tag{51}$$

and defined on the interval $(-\infty, \infty)$, where $\text{HermiteH}[k, w]$ denotes a k th-degree Hermite polynomial in the variable w in *Mathematica* notation. The weight function associated with modified Hermite polynomials which is $w[x] = e^{-x^2/2}$ is proportional to the density function of a standard Gaussian random variable. Thus, the requisite transformation is $X = (Y - u)/s$ with $u = \mu_Y[1]$ and $s = \sqrt{\mu_Y[2] - \mu_Y[1]^2}$, the normalizing constant is $c_T = 1/\sqrt{2\pi}$, and the base density function is that of a standard normal random variable, that is,

$$\psi_X[x] = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad -\infty < x < \infty, \tag{52}$$

whose j th moment is given by

$$m_X[j] = \frac{2^{1/2(-1+j)} (1 + (-1)^j) \text{Gamma}\left[\frac{1+j}{2}\right]}{\sqrt{2\pi}}, \quad j = 0, 1, \dots \tag{53}$$

Moreover, in this case,

$$\theta_k = \sqrt{2\pi} k!. \tag{54}$$

The same density approximants are obtained whether one makes use of equation (32) or (42). They are also known as (type-A) Gram-Charlier expansions. A methodology for determining the Hermite polynomial coefficients in the expansions in terms of the moments is presented in [3, Section 5.4], where the advantages and drawbacks of using such approximations are also discussed. Conditions ensuring the convergence of the approximants are available from Cramér [26].

□ Example 7

Consider an equally weighted mixture of two Gaussian distributions with parameters (3, 4) and (1, 1), whose density function is plotted as a solid line in Figure 11. On the basis of 30 moments, identical density approximants (represented by a dashed line) are obtained whether we make use of equation (42) or Result 4.1.

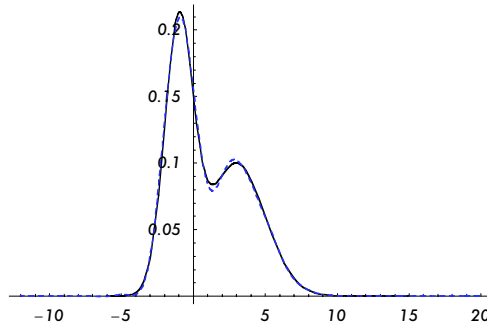


Figure 11. Exact and approximate (dashed line) PDFs. [S7, S71, or S7a in the Appendix]

■ Acknowledgments

I am very grateful to a reviewer for several helpful suggestions which improved the presentation of the material. Thanks are also due to the guest editor and a second reviewer for pointing out the work of Reinking [19]. Finally, the financial support of the Natural Sciences and Engineering Research Council of Canada is gratefully acknowledged.

■ Appendix

This appendix contains the *Mathematica* code that was used in connection with each of the seven examples. All the calculations are carried out with rational numbers so as to prevent any loss of precision. *It would be advisable to quit the kernel between examples.*

□ Code for Example 1

The end points of the support of the random variable are a and b . The degree of the polynomial approximation is n . In this example, the random variable of interest is V , which will be denoted by Y to conform to the general notation utilised in Section 4. Its b th moment is given by $\mu_Y[b]$ whereas the j th moment of X defined on the interval $[-1, 1]$ is denoted by $\mu_X[j]$. The exact density function is denoted by $f_Y[y]$ and the approximate density functions obtained

from equations (13) and (15) are given by $f_{Y_n}[y]$ and $f1_{Y_n}[y]$, respectively. The exact and approximate cumulative distribution functions obtained by integration are denoted by $Q_Y[u]$ and $Q_{Y_n}[u]$, respectively.

$$\begin{aligned}
 & \mathbf{a} = 0; \\
 & \mathbf{b} = 3; \\
 & \mathbf{n} = 13; \\
 & \mu_Y[\mathbf{h}_-] := \mu_Y[\mathbf{h}] = \\
 & \int_0^1 \int_0^1 \int_0^1 (\delta_1 + \delta_2 + \delta_3)^{\mathbf{h}} (\delta_1^{-1/2} - 1) (\delta_2^{-1/2} - 1) (\delta_3^{-1/2} - 1) \mathbf{d}\delta_1 \mathbf{d}\delta_2 \mathbf{d}\delta_3;
 \end{aligned}$$

$$\begin{aligned}
 & \mu_X[\mathbf{j}_-] := \\
 & \mu_X[\mathbf{j}] = \text{Expand}\left[\left(\frac{2 Y - (a + b)}{b - a}\right)^{\mathbf{j}}\right] /. \text{Table}[Y^i \rightarrow \mu_Y[i], \{i, 0, n\}];
 \end{aligned}$$

$$\begin{aligned}
 f_Y[y_-] := f_Y[y] = & \text{Which}\left[0 \leq y < 1, \pi (2 \sqrt{y} - 3 y) + 4 y^{3/2} - \frac{y^2}{2},\right. \\
 & 1 \leq y \leq 2, -\frac{1}{2} + \pi (3 - 4 \sqrt{y} + 6 y) + 3 y + y^2 - \\
 & 4 \sqrt{-1 + y} (1 + 2 y) - 12 y \text{ArcSin}\left[\frac{1}{\sqrt{y}}\right], 2 < y \leq 3, \\
 & 4 \sqrt{-2 + y} (1 + y) + \frac{1}{2} (-5 + \pi (2 - 6 y) - y (6 + y)) + \\
 & 4 (-1 + 3 y) \text{ArcCsc}\left[\sqrt{-1 + y}\right] + 8 \text{ArcTan}\left[\frac{3 - y}{2 \sqrt{-2 + y}}\right] - \\
 & \left. 4 \sqrt{y} \left(\text{ArcTan}\left[\frac{1}{\sqrt{(-2 + y) y}}\right] + \text{ArcTan}\left[\frac{1 - (-2 + y) y}{2 \sqrt{(-2 + y) y}}\right]\right)\right];
 \end{aligned}$$

$$f_{Y_n}[y_-] := \sum_{k=0}^n \left(\frac{2 k + 1}{b - a} \text{LegendreP}[k, X] /. X^{j_-} \rightarrow \mu_X[j] \right)$$

$$\text{LegendreP}\left[k, \frac{2 y - (a + b)}{b - a}\right];$$

$$f1_{Y_n}[y_-] := \sum_{k=0}^n \frac{2 k + 1}{b - a} \left(\text{LegendreP}\left[k, \frac{2 Y - (a + b)}{b - a}\right] /. Y^{j_-} \rightarrow \mu_Y[j] \right)$$

$$\text{LegendreP}\left[k, \frac{2 y - (a + b)}{b - a}\right];$$

$$\begin{aligned}
 \text{Pq} = & \text{Plot}\left[\{f_{Y_n}[y], f_Y[y]\}, \{y, 0, 3\},\right. \\
 & \text{PlotStyle} \rightarrow \{\{\text{Thickness}[0.00765], \text{Dashing}\{\{0.005, 0.019\}\}\}, \\
 & \left. \{\text{Thickness}[0.001], \text{Hue}[0.7]\}\};
 \end{aligned}$$

$$\begin{aligned}
 \text{Pq1} = & \text{Plot}\left[\{f1_{Y_n}[y], f_Y[y]\}, \{y, 0, 3\},\right. \\
 & \text{PlotStyle} \rightarrow \{\{\text{Thickness}[0.00765], \text{Dashing}\{\{0.005, 0.019\}\}\}, \\
 & \left. \{\text{Thickness}[0.001], \text{Hue}[0.7]\}\};
 \end{aligned}$$

$$Q_Y[u_] := \int_0^u f_Y[y] dy;$$

$$Q_{Y_n}[u_] := Q_{Y_n}[u] = \int_0^u f_{Y_n}[y] dy;$$

```
PQ = Plot[{Q_{Y_n}[u], Q_Y[u]}, {u, 0, 3}, PlotRange -> All,
  PlotStyle -> {{Thickness[0.00765'], Dashing[{0.025', 0.029'}]},
  {Thickness[0.001'], Hue[0.7']}}];
```

```
Qd = Plot[Q_{Y_n}[u] - Q_Y[u], {u, 0, 3}, PlotRange -> All];
```

□ Code for Example 2

The end points of the support of the random variable are a and b . The degree of the polynomial approximation is n . For a given sequence of moments, $\mu_Y[b]$, $b = 1, 2, \dots, n$, the approximate density, $f_{Y_n}[y]$, is obtained from equation (14) in this case.

```
ClearAll[a, b, n, μ, f];
a = 0;
b = 1;
n = 15;
```

```
<< Statistics`ContinuousDistributions`
```

$$\mu_Y[h_] := \mu_Y[h] = \int_0^1 x^h \frac{1}{2} (\text{PDF}[\text{BetaDistribution}[3, 2], x] + \text{PDF}[\text{BetaDistribution}[2, 30], x]) dx;$$

```
f_{Y_n}[y_] :=
```

$$\sum_{k=0}^n \frac{2k+1}{b-a} \left(\text{LegendreP}[k, X] / X^{j-1} \rightarrow \sum_{k=0}^j \frac{j! 2^k \mu_Y[k] (-a-b)^{j-k}}{(b-a)^j k! (j-k)!} \right) \text{LegendreP}\left[k, \frac{2y-(a+b)}{b-a}\right];$$

```
Pb = Plot[{{1/2 (PDF[BetaDistribution[3, 2], y] + PDF[BetaDistribution[2, 30], y]), f_{Y_n}[y]},
  {y, 0, 1}, PlotRange -> All, PlotStyle ->
  {{RGBColor[0, 1, 0]}, {Dashing[{0.01, 0.01, 0.01, 0.01}]}}];
```

□ Code for Example 3

The b th moment of the standard lognormal distribution whose support is the positive half-line is $\mu_Y[b] = e^{b^2/2}$. The exact and approximate density functions

are respectively given by $f_Y[y]$ and $f_{Y_n}[y]$, the latter being obtained from equation (28). $F_{Y_n}[w]$ denotes the distribution function corresponding to $f_{Y_n}[y]$.

```

a = 0;
n = 3;

<< Statistics`ContinuousDistributions`

f_Y [y_] := f_Y [y] = PDF[LogNormalDistribution[0, 1], y];
mu_Y [h_] := mu_Y [h] = ExpectedValue[y^h, LogNormalDistribution[0, 1], y];
c = (mu_Y [2] - mu_Y [1]^2) / (mu_Y [1] - a);
v = -1 + (mu_Y [1] - a) / c;

mu_X [j_] := mu_X [j] = Expand[(Y - a) / c]^j /. Y^h -> mu_Y [h];
f_Yn [y_] := f_Yn [y] = (y - a)^v e^(-(y-a)/c) / c^(v+1)
Sum[j! LaguerreL[j, v, (y-a)/c] / Gamma[v + j + 1], {j, 0, n}] (LaguerreL[j, v, X] /. X^k -> mu_X [k]);

PE = Plot[f_Y [y], {y, 0, 10},
  PlotRange -> All, PlotStyle -> RGBColor[0, 0, 1]];
PA = Plot[f_Yn [y], {y, 0.1, 10}, PlotRange -> All,
  PlotStyle -> Dashing[{0.01, 0.01, 0.01, 0.01}]];
PEA = Show[PE, PA];
CLN = Plot[CDF[LogNormalDistribution[0, 1], y], {y, 0, 10}];

F_Yn [w_] := F_Yn [w] = Integrate[f_Yn [y] dy, {y, 0, w}];
CLN3 = Plot[Evaluate[F_Yn [w]], {w, 0, 10}, PlotRange -> All,
  PlotStyle -> Dashing[{0.01, 0.01, 0.01, 0.01}]];
CDFLN = Show[CLN3, CLN];

```

□ Code for Example 4

The b th moment of the mixture of three shifted gamma random variables, which is denoted by $\mu_Y[b]$, is obtained by differentiation of the moment-generating function, $\text{MGF}_Y[t]$. The degree of the polynomial approximation is n . The exact and approximate density functions are respectively given by $f_Y[y]$ and $f_{Y_n}[y]$, the latter being determined from equation (29).

```

a = 5;
fY [y_] := If [y > a,
  (y - a)α1-1 e-(y-a)/β1 / (3 Gamma [α1] β1α1) +
  (y - a)α2-1 e-(y-a)/β2 / (3 Gamma [α2] β2α2) +
  (y - a)α3-1 e-(y-a)/β3 / (3 Gamma [α3] β3α3), 0];

α1 = 8; β1 = 1;
α2 = 16; β2 = 1;
α3 = 64; β3 = 1/2;

n = 60;
MGFY [t_] = ea t (1 - β1 t)-α1 / 3 + ea t (1 - β2 t)-α2 / 3 + ea t (1 - β3 t)-α3 / 3;
μY [h_] := μY [h] = D [MGFY [t], {t, h}] /. t -> 0;
c = (μY [2] - μY [1]2) / (μY [1] - a);
v = -1 + (μY [1] - a) / c;
fYn [y_] := fYn [y] = (y - a)v e-(y-a)/c / cv+1
  Sum [j! / Gamma [v + j + 1] (LaguerreL [j, v, (y - a) / c]) / j! Yk -> μY [k])
  * LaguerreL [j, v, (y - a) / c];
ψY [y_] := ψY [y] = (y - a)v e-(y-a)/c / (cv+1 Gamma [v + 1]);
PE = Plot [fY [y], {y, a, 65},
  PlotRange -> All, PlotStyle -> RGBColor [0, 1, 0]];
PA = Plot [fYn [y], {y, a, 65}, PlotRange -> All,
  PlotStyle -> Dashing [{0.01, 0.01, 0.01, 0.01}]];
PDEA = Show [PE, PA];
PG = Plot [ψY [y], {y, a, 65},
  PlotRange -> All, PlotStyle -> RGBColor [0, 0, 1]];
PGE = Show [PG, PE];

```

□ Code for Example 5

We use the notation introduced in Section 4. First, we approximate the density function of the random variable V as defined in Section 2 by making use of equation (42) with $T_i[x] \equiv G_i[\alpha + \beta + 1, \alpha + 1, x]$, a modified Jacobi polynomial. To conform to the general notation used in Section 4, we will again replace V with Y .

u = 0;
s = 3;
n = 8;
f_y[y_] :=

$$f_y[y] = \text{Which}\left[0 \leq y < 1, \pi(2\sqrt{y} - 3y) + 4y^{3/2} - \frac{y^2}{2}, 1 \leq y \leq 2, -\frac{1}{2} + \pi(3 - 4\sqrt{y} + 6y) + 3y + y^2 - 4\sqrt{-1+y}(1+2y) - 12y \text{ArcSin}\left[\frac{1}{\sqrt{y}}\right], 2 < y \leq 3, 4\sqrt{-2+y}(1+y) + \frac{1}{2}(-5 + \pi(2 - 6y) - y(6+y)) + 4(-1 + 3y) \text{ArcCsc}[\sqrt{-1+y}] + 8 \text{ArcTan}\left[\frac{3-y}{2\sqrt{-2+y}}\right] - 4\sqrt{y} \left(\text{ArcTan}\left[\frac{1}{\sqrt{-2+y}y}\right] + \text{ArcTan}\left[\frac{1 - (-2+y)y}{2\sqrt{-2+y}y}\right]\right)\right];$$

μ_y[h_] := μ_y[h] =

$$\int_0^1 \int_0^1 \int_0^1 (\delta_1 + \delta_2 + \delta_3)^h (\delta_1^{-1/2} - 1) (\delta_2^{-1/2} - 1) (\delta_3^{-1/2} - 1) d\delta_1 d\delta_2 d\delta_3;$$

μ_x[j_] := μ_x[j] = Expand[$\left(\frac{y-u}{s}\right)^j$] /. Y^k -> μ_y[k];

w[x_] := x^α (1 - x)^β;

$$\alpha = \mu_x[1] \frac{\mu_x[1] - \mu_x[2]}{\mu_x[2] - \mu_x[1]^2} - 1;$$

$$\beta = (1 - \mu_x[1]) \frac{\alpha + 1}{\mu_x[1]} - 1;$$

$$G_n[\sigma_, \tau_, x_] := m! \frac{\text{Gamma}[m + \sigma]}{\text{Gamma}[2m + \sigma]} \text{JacobiP}[m, \sigma - \tau, \tau - 1, 2x - 1];$$

$$\theta_{k_} := \theta_k = \frac{1}{\frac{(2k + \alpha + \beta + 1) \text{Gamma}[2k + \alpha + \beta + 1]^2}{k! \text{Gamma}[k + \alpha + 1] \text{Gamma}[k + \alpha + \beta + 1] \text{Gamma}[k + \beta + 1]}};$$

f_{y_n}[y_] := f_{y_n}[y] = w $\left[\frac{y-u}{s}\right]$

$$\sum_{i=0}^n \left(\frac{1}{s \theta_i} \sum_{k=0}^i (\text{CoefficientList}[G_i[\alpha + \beta + 1, \alpha + 1, X], X][[k + 1]] \mu_x[k]) \right.$$

$$\left. G_i\left[\alpha + \beta + 1, \alpha + 1, \frac{y-u}{s}\right]\right);$$

S5 = Plot[{Evaluate[f_{y_n}[y]], f_y[y]}, {y, 0, 3},
PlotStyle -> {{Thickness[0.00765], Dashing[{0.005, 0.019]}},
{Thickness[0.001], Hue[0.7]}}];

Identical density approximants denoted by $f1_{Y_n}[y]$ and $f2_{Y_n}[y]$ in the following code are obtained from Result 4.1 and equation (43), respectively.

$$\psi_x[x_] := \frac{1}{\text{Beta}[\alpha + 1, \beta + 1]} x^\alpha (1 - x)^\beta;$$

$$m_x[j_] := m_x[j] = \frac{\text{Pochhammer}[\alpha + 1, j]}{\text{Pochhammer}[\alpha + \beta + 2, j]};$$

$$\text{IMb}[x_] := (\text{Inverse}[\text{Table}[m_x[h + i], \{h, 0, n\}, \{i, 0, n\}]] \cdot \text{Table}[\mu_x[j], \{j, 0, n\}]) \cdot \text{Table}[x^j, \{j, 0, n\}];$$

$$f1_{Y_n}[y_] := \frac{\psi_x\left[\frac{y-u}{s}\right] \text{IMb}\left[\frac{y-u}{s}\right]}{s};$$

$$c_T = \frac{1}{\text{Beta}[\alpha + 1, \beta + 1]};$$

$$f2_{Y_n}[y_] := f2_{Y_n}[y] = \psi_x\left[\frac{y-u}{s}\right]$$

$$\sum_{l=0}^n \left(\sum_{i=l}^n \frac{1}{s c_T \theta_i} \text{CoefficientList}[G_i[\alpha + \beta + 1, \alpha + 1, X], X][[l + 1]] \right.$$

$$\left. \sum_{k=0}^i (\text{CoefficientList}[G_i[\alpha + \beta + 1, \alpha + 1, X], X][[k + 1]] \mu_x[k]) \right) \left(\frac{y-u}{s} \right)^l;$$

`Simplify[fYn[y] // N]`

`Simplify[f1Yn[y] // N]`

`Simplify[f2Yn[y] // N]`

□ Code for Example 6

We use Result 4.1 to approximate the density of Wilks' likelihood ratio criterion, which will be denoted by Y to conform to the general notation utilised in Section 4. In this case, both X and Y are defined on the interval $(0, 1)$. The transformation is included in the code so that it could be applied if needed. The symbol N being protected, it is replaced by N^* in the following code.

```

u = 0;
s = 1;
N* = 12;
q1 = 2;
q2 = 4;
p = 2;

```

$$\mu_Y[h_] := \mu_Y[h] = \prod_{i=1}^p \frac{\text{Gamma}\left[\frac{N^* - q_1 - q_2 + 1 - i}{2} + h\right] \text{Gamma}\left[\frac{N^* - q_2 + 1 - i}{2}\right]}{\text{Gamma}\left[\frac{N^* - q_1 - q_2 + 1 - i}{2}\right] \text{Gamma}\left[\frac{N^* - q_2 + 1 - i}{2} + h\right]};$$

$$\mu_X[j_] := \mu_X[j] = \text{Expand}\left[\left(\frac{Y - u}{s}\right)^j\right] /. Y^{k_} \rightarrow \mu_Y[k];$$

$$\alpha = \mu_X[1] \frac{\mu_X[1] - \mu_X[2]}{\mu_X[2] - \mu_X[1]^2} - 1;$$

$$\beta = (1 - \mu_X[1]) \frac{\alpha + 1}{\mu_X[1]} - 1;$$

$$\psi_X[x_] := \frac{1}{\text{Beta}[\alpha + 1, \beta + 1]} x^\alpha (1 - x)^\beta;$$

$$m_X[j_] := m_X[j] = \frac{\text{Pochhammer}[\alpha + 1, j]}{\text{Pochhammer}[\alpha + \beta + 2, j]};$$

```
IMb[x_, n_] :=
```

```

  IMb[x, n] = (Inverse[Table[m_X[h + i], {h, 0, n}, {i, 0, n}]] .
    Table[\mu_X[j], {j, 0, n}]).Table[x^j, {j, 0, n}];

```

$$f1_{Y_n}[y_] := \frac{\psi_X\left[\frac{y-u}{s}\right] \text{IMb}\left[\frac{y-u}{s}, n\right]}{s};$$

```

P6 = Plot[{Evaluate[f1_{Y_2}[y]], Evaluate[f1_{Y_{20}}[y]]}, {y, u, u + s},
  PlotStyle -> {{Thickness[0.00765], Dashing[{0.005, 0.019]}],
    {Thickness[0.001]}}];

```

$$F1_{Y_n}[w_] := F1_{Y_n}[w] = \int_u^w f1_{Y_n}[y] dy;$$

```

C6 = Plot[{Evaluate[F1_{Y_2}[w]], Evaluate[F1_{Y_{20}}[w]]}, {w, u, u + s},
  PlotStyle -> {{Thickness[0.00765], Dashing[{0.005, 0.019]}],
    {Thickness[0.001]}}];

```

```
F1_{Y_{20}}[0.87825] // N
```

```
F1_{Y_{20}}[0.94719] // N
```

```
F1_{Y_2}[0.87825] // N
```

```
F1_{Y_2}[0.94719] // N
```

□ Code for Example 7

We use the notation introduced in Section 4. First, we approximate the density function of the Gaussian mixture with $f_{Y_n}[y]$ by making use of an alternate representation of equation (42) with $T_i[x] \equiv H^*_i[x]$, a modified Hermite polynomial. The function $f1_{Y_n}[y]$ provides an approximant which is directly expressed in terms of the moments of Y .

```

fY[y_] := fY[y] =  $\frac{1}{2} \frac{1}{\sqrt{2\pi}} \frac{1}{2} (e^{-(y-3)^2/8}) + \frac{1}{2} \frac{1}{\sqrt{2\pi}} (e^{-(y+1)^2/2});$ 
μY[h_] :=
  μY[h] =  $\frac{1}{2} \int_{-\infty}^{\infty} y^h \frac{1}{\sqrt{2\pi}} \frac{1}{2} (e^{-(y-3)^2/8}) dy + \frac{1}{2} \int_{-\infty}^{\infty} y^h \frac{1}{\sqrt{2\pi}} (e^{-(y+1)^2/2}) dy;$ 
n = 30;
Table[μY[h], {h, 0, n}];
u = μY[1];
s =  $\sqrt{\mu_Y[2] - \mu_Y[1]^2};$ 
μX[j_] := μX[j] = Expand[ $(\frac{Y-u}{s})^j$ ] /. Yh -> μY[h];
w[x_] := e-x2/2;

H*k[x_] := (-1)k 2-k/2 HermiteH[k,  $\frac{x}{\sqrt{2}}$ ];
θk := θk =  $\sqrt{2\pi} k!$ ;
fYn[y_] := w[ $\frac{y-u}{s}$ ]  $\sum_{i=0}^n \left( \frac{1}{s \theta_i} (H^*_i[x] /. x^j ->$ 
  μX[j]) H*i[ $\frac{y-u}{s}$ ]  $\right)$ ;
f1Yn[y_] := w[ $\frac{y-u}{s}$ ]  $\sum_{i=0}^n \left( \frac{1}{s \theta_i} (H^*_i[x] /. x^j ->$ 
  Expand[ $(\frac{Y-u}{s})^j$ ] /. Yh -> μY[h]) H*i[ $\frac{y-u}{s}$ ]  $\right)$ ;

S7 = Show[Plot[fY[y], {y, -6, 15}, PlotRange -> All],
  Plot[Evaluate[fYn[x]], {x, -12, 20}, PlotRange -> All,
  PlotStyle -> {Dashing[{0.01, 0.01}], RGBColor[0, 0, 1]}];
S71 = Show[Plot[fY[y], {y, -6, 15}, PlotRange -> All],
  Plot[Evaluate[f1Yn[x]], {x, -12, 20}, PlotRange -> All,
  PlotStyle -> {Dashing[{0.01, 0.01}], RGBColor[0, 0, 1]}];

```

The same density approximant is obtained from equation (32).

```

n = 30;
ψx[x_] :=  $\frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ ;
mx[j_] := mx[j] =  $\frac{2^{1/2(-1+j)} (1 + (-1)^j) \text{Gamma}\left[\frac{1+j}{2}\right]}{\sqrt{2\pi}}$ ;
IMb[x_] := (Inverse[Table[mx[h+i], {h, 0, n}], {i, 0, n}]] .
Table[μx[j], {j, 0, n}].Table[xj, {j, 0, n}];
f2vn[y_] :=  $\frac{\psi_x\left[\frac{y-u}{s}\right] \text{IMb}\left[\frac{y-u}{s}\right]}{s}$ ;
S7a = Show[Plot[fy[y], {y, -6, 15}, PlotRange → All],
Plot[Evaluate[f2vn[x]], {x, -12, 20}, PlotRange → All,
PlotStyle → {Dashing[{0.01, 0.01}], RGBColor[0, 0, 1]}]];

```

■ References

- [1] H. Solomon and M. Stephens, "Approximations to Density Functions Using Pearson Curves," *Journal of the American Statistical Association*, **73**(361), 1978 pp. 153–160.
- [2] W. P. Elderton and N. L. Johnson, *Systems of Frequency Curves*, Oxford: Cambridge University Press, 1969.
- [3] C. Rose and M. D. Smith, *Mathematical Statistics with Mathematica*, New York: Springer-Verlag, 2002.
- [4] N. Reid, "Saddlepoint Methods and Statistical Inference," *Statistical Science*, **3**(2), 1988 pp. 213–238.
- [5] A. M. Mathai and R. K. Saxena, *The H-function with Applications in Statistics and Other Disciplines*, New York: John Wiley & Sons, 1978.
- [6] S. B. Provost and E. M. Rudiuk, "Moments and Densities of Test Statistics for Covariance Structures," *International Journal of Mathematical and Statistical Sciences*, **4**(1), 1995 pp. 85–104.
- [7] P. J. Smith, "A Recursive Formulation of the Old Problem of Obtaining Moments from Cumulants and Vice Versa," *The American Statistician*, **49**, 1995 pp. 217–219.
- [8] S. B. Provost and Y.-H. Cheong, "On the Distribution of Linear Combinations of the Components of a Dirichlet Random Vector," *The Canadian Journal of Statistics*, **28**, 2000 pp. 417–425.
- [9] G. Sansone, *Orthogonal Functions*, Pure and Applied Mathematics, Vol. 9, New York: Interscience Publishers, 1959.
- [10] G. Alexits, *Convergence Problems of Orthogonal Series*, New York: Pergamon Press, 1961.
- [11] L. Devroye and L. Györfi, *Nonparametric Density Estimation: The L₁ View*, New York: John Wiley & Sons, 1985.
- [12] W. B. Jones and A. S. Ranga, eds., *Orthogonal Functions, Moment Theory, and Continued Fractions, Theory and Applications*, Lecture Notes in Pure and Applied Mathematics, New York: Marcel Dekker, 1998.

- [13] L. Devroye, "On Random Variate Generation When Only Moments or Fourier Coefficients Are Known," *Mathematics and Computers in Simulation*, **31**, 1989 pp. 71–89.
- [14] R. L. Burden and J. D. Faires, *Numerical Analysis*, 4th ed., Boston: PWS-Kent, 1989.
- [15] C. R. Rao, *Linear Statistical Inference and Its Applications*, New York: John Wiley & Sons, 1965.
- [16] A. M. Mathai, *An Introduction to Geometric Probability: Distributional Aspects with Applications*, The Netherlands: Gordon and Breach, 1999.
- [17] G. Szegő, *Orthogonal Polynomials*, Providence, RI: American Mathematical Society, 1959.
- [18] A. M. Mathai and S. B. Provost, *Quadratic Forms in Random Variables, Theory and Applications*, New York: Marcel Dekker, 1992.
- [19] J. T. Reinking, "Orthogonal Series Representations of Probability Density and Distribution Functions," *The Mathematica Journal*, **8**(3), 2003 pp. 473–483.
- [20] J. Gurland, "Distribution of Definite and of Indefinite Quadratic Forms," *The Annals of Mathematical Statistics*, **26**, 1955 pp. 122–127.
- [21] M. L. Tiku, "Laguerre Series Forms of Non-central χ^2 and F Distributions," *Biometrika*, **52**(3), 1965 pp. 415–427.
- [22] N. L. Johnson and S. Kotz, *Distributions in Statistics: Continuous Univariate Distributions*, Vol. 2, New York: Wiley, 1970.
- [23] J. Durbin and G. S. Watson, "Testing for Serial Correlation in Least-Squares Regression II," *Biometrika*, **38**(1–2), 1951 pp. 159–178.
- [24] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, 3rd ed., New York: John Wiley & Sons, 2003.
- [25] A. M. Mathai, "On the Distribution of the Likelihood Ratio Criterion for Testing Linear Hypotheses on Regression Coefficients," *Annals of the Institute of Statistical Mathematics*, **23**, 1971 pp. 181–197.
- [26] H. Cramér, "On Some Classes of Series Used in Mathematical Statistics," *Proceedings of the Sixth Scandinavian Congress of Mathematicians*, Copenhagen, 1925 pp. 399–425.

About the Author

Serge B. Provost is a professor of statistics in the Department of Statistical and Actuarial Sciences at The University of Western Ontario. He holds a Ph.D. in mathematics and statistics from McGill University. He is a fellow of the Royal Statistical Society and an associate of the Society of Actuaries. He co-authored two research monographs on quadratic and bilinear forms in random variables. His research interests include distribution theory, multivariate analysis, order statistics, time series, and computational statistics.

Serge B. Provost

Department of Statistical and Actuarial Sciences
The University of Western Ontario
London, Ontario, Canada N6A 5B7
sp@uwo.ca