DISTRIBUTION CHARACTERIZATION IN A PRACTICAL MOMENT PROBLEM

H. C. JIMBO

Abstract. We investigate a problem connected with the evaluation of the asymptotic probability distribution function (APDFs) given from a set of finite order moments by applying the Gram-Schmidt process with the aid of computer algebra. By selecting weighting (discrete or continuous) function of similar shape to desired (APDFs), orthogonal polynomial series are obtained that are stable at high order and allow accurate approximation of tail probabilities.

1. Introduction

Many mathematicians have investigated the problem of moments, which consists of determining a probability density function from a set of its moments, since the pioneering work of Tchebycheff and Stieltjes during the 19^{th} century [13]. Since then a variety of techniques for approaching the problem have been developed. In 2000 [5], [6] have introduced a new technique "factorial behaviors based" to characterize the distribution of some classes of discrete functions or processes. In the present paper a method of extending the existing technique using orthogonal polynomial expansion is presented. This is of potential use to practitioners since it constitute part of an alternative approach to the Monte Carlo technique for probability risk analysis; it also has application as a method for predicting the probability of extreme events, for fitting distribution to large data sets. The need to develop alternative to Monte Carlo and other simulation technique has been noted by several authors [4], [2], [15]. While the simplicity of simulation techniques are desirables, there are lacking in terms of efficiency. It may take thousand of simulations and consequently many hours of computer time in order to achieve a high level of accuracy, especially in a fat tail distribution [13]. An alternative to the histogram produce by the Monte

Received September 11, 2003.
2000 *Mathematics Subject Classification.* Primary 62B05, 2B99, 37C75, 37D99.
Key words and phrases. Tail probabilities, problem of moments, probabilistic programming.

Carlo simulation is an accurate form of probability density function of frequency distribution. In general most probabilistic calculations are too complicated for exact PDF's to be determined. However for simple problems implicitly those that involve implicit calculations, it often possible to calculate the moment of the PDF. The remaining step is the determination of the PDF, which require solution to real or practical problem of moments. One of the difficulties with this step is that there exist no unique PDF for a given set of moments. However from a practical stance, this difficulty can be overcome just by increasing the constraint on the PDF by accounting a greater number of moments. Nevertheless most of the existing techniques generally only make use of a few lower order moments (often not more than four) in forming the probability function. The main contribution of this paper is the demonstration of a technique that incorporates informations given by high order moments by the use of discrete and continuous weigh functions. Many authors support the idea that a better approximation to a distribution can be archive by using a weighting function, which closely resembles to the desired PDF. But this is true only for the continuous case. If the chosen weighing function is discrete then the discussion is opened.

2. Background

In a strict sense, the problem of moments is concerned with the question of whether a set moment uniquely determines a PDF. Stuart and Ord [7] state that: "*in it full generality the problem of moments consider a set of constants and inquires whether they can be the moments of a distribution.*" However, it is added that for statistical purposes a more pertinent question is: "*given that the set of constants are the moments of a distribution, can any other distribution have the same set?*". An important result due to Carlman [2] is that a distribution on the range $[-\infty, +\infty]$ can be determined uniquely by its moments μ_r, if the series

$$\sum_{k=0}^{\infty} \frac{1}{(\mu_{2k})^{1/2}}$$

converges.

Similar results exist for functions on the range $[0, +\infty]$or alternatively in terms of absolute moments [15]. Godwin [3] roughly interprets these results as meaning that the moment must not be too large or the distribution must not be too spread out. Several examples of functions, which cannot be uniquely determined from their full set of moments, have been demonstrated in the literature. Many of these see little applications, although the lognormal distribution is an exception [2]. Nevertheless, there are at least two reasons why the closure problem may not be of practical application. Firstly, many randomly distributed quantities encountered in practice have finite range and a probability distribution with finite range can be uniquely determined from it full set of moments. Secondly, as Stuart and Ord point out, *"if two distributions have a certain number of moment in common, they will bear some resemblance to each other"*. In practice full (infinite) set of moment will never be attainable and the question of uniqueness does arises. However the infinite number of PDFs, which may be obtained from a finite set of moments become increasable similar as the number of given moments increases. Approximation of a distribution function by another function possessing even just the same four lower moments is often found remarquably good.

Several techniques have been developed for establishing a PDF from a set of moments. Most of the methods are discussed in detail by Elderton, Johnson and Ord [9], other insightful reviews are given by Wallace [16] and Springer [14]. The Pearson family of distributions often provides reasonable approximation to a PDF, base on just the first four moments. The traditional alternative is the use of series, which are expressed in term of Tchebycheff-Hermite polynomials; and its coefficients are determined by orthogonalization. Other orthogonal polynomials such as Jacobi or Laguerre polynomials have also been used to approximate PDF's. Other methods of approximation include Burr's [1] general system for fitting cumulative distribution functions; a step function method due to Von Mises [2] and a further distribution system proposed by Perk [10]. Of the methods describe above, no single one is indisputably recognized as being the best way o approximating a PDF. In general most of the method work reasonably well, although it has been found that difference between approximating functions are greater for more skewed PDF's [13].

Orthogonal polynomials have featured in several of the methods above and they have useful properties for approximating functions in general. According to Weistrass theorem, for any function on the range $[a, b]$ there exist an algebraic polynomial, which converges to that function as the order of approximation increases [12]. Orthogonal polynomial are usually are particularly useful since they are inexpensive to compute. They also possess the property of minimum least squares estimate of the ratio of the objective function of the square root of the weighting function at a given order. Nevertheless, there are some limitations to the use of orthogonal polynomials as approximating functions. Firstly, to obtain a very high degree of accuracy an excessive number of polynomials may sometime be required. Secondly, convergence is not guaranteed to be uniform. So while orthogonal polynomials may be used to give good approximation to PDF's, they may not necessary produce series that asymytotically approach the actual PDF. In the following work, it will be shown that these potential limitations may be overcome by the choice of the weighting function for the approximation of specific function.

3. Method

We want to determine the frequency distribution $f(x)$, which fits a given set of moments $\mu_f(k)$, $k = 1, \ldots, n$. Firstly a weighting function that can be discrete or continuous are established by fitting the lower moment using the Pearson distribution for example, then a family of orthogonal polynomial $\phi_k(x)$ is sought such that:

$$\int_{-\infty}^{+\infty} w(x)\phi_k(x)\phi_l(x) = 0, \qquad k \neq l \tag{1}$$

The equation (1) produces a set of n simultaneous equations formed from the moments $\mu_w(k)$. This set can be efficiently solved using Mat lab toolbox to give the coefficient of the polynomials. The squared norms h_n are also determined from $\mu_w(k)$ via: $h_n = \int_{-\infty}^{+\infty} w(x)\phi_n^2(x)dx$ and the desired PDF is given by the approximate solution of:

$$f(x) \approx \sum_{k=1}^{r} a_k w(x) \tag{2}$$

where the coefficients a_k are determined from the moments $\mu_f(k)$ by:

$$a_k = \frac{1}{h} \int_{-\infty}^{+\infty} f(x)\phi_k(x)dx \tag{3}$$

Order	Coefficient
1	0
2	0
3	0.80083
4	1.41367
5	1.62130
6	1.61819
7	1.57784
8	1.53637
9	1.41483
10	1.12148
11	0.76986
12	1.00375
13	−1.79600
14	0.69985
15	0.10281

Table 1. Coefficients of orthogonal series for approximating χ^2 distribution

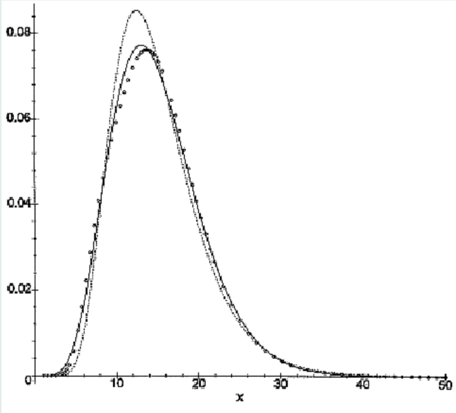

Figure 1. Approximation of a χ^2 distridution (solid line) by a 11th order series expansion based on the lognormal weighting function.

4. Verification using the cases studied

The proposed method for fitting a PDF to a given set of moments may be first verified by using the moments from an exactly known distribution. As a demonstration, a χ^2 PDF with shape parameter $\upsilon = 15$, is chosen as the known PDF and an attempt shall be made to approximate this density function using the non classical orthogonal polynomials. A lognormal density function is used as weighting function for the orthogonal expansion since this family of functions is relatively similar to in shape to χ^2distribution. The most appropriate lognormal weighting function is that transformed from a normal function. This weighting function has the same first two moments as the χ^2 PDF. Using the Gram-Schmidt process a set of orthogonal polynomials is determined from the moments of the weighting function. Then using the specifique moments the coefficients a_k are determined. In this case the coefficients demonstrate that the orthogonal expansion is well behaved up to $11^t h$ order (Table 1). The 12^{th} and 13^{th} order coefficients show slight increases magnitude, although the associated series expansions are still good.

5. Example

To demonstrate the use of general orthogonal series expansion for approximating PDF's three further examples are considered where probabilities of extrem events are required given only limited data set. For the first example, a PDF is sought for the result of 40 compressions strength test s conducted by Mathur [8]. Values indicated that Pearson type 1 distribution is suitable as the weighting function for this case (Table 2). The type 1 PDF has the form:

$$w(x) = y_e \left(1 + \frac{x}{A_1}\right)^{m_1} \left(1 + \frac{x}{A_2}\right)^{m_2} \tag{4}$$

with parameters:

$$\begin{aligned} &y_e = 0.10892; \quad A_1 = 7.64336; \quad A_2 = 9.16439 \\ &m_1 = 1.59333; \quad m_2 = 2.10940. \end{aligned}$$

For the concrete strength data, the difference between the weighting function and series expansion is not so significant. Nevertheless, the eight order series does produce a slightly better fit as can be seen from the moments in Table 2. The second example is concerned with establishing a density function for a complex process; the incomplete Gauss process is used for that matter.

$$w_h(x) = \sum_{n=x+1}^{x+h} \Upsilon(n) \cdot \mathrm{e}^{2\pi \mathrm{i} \frac{an}{p}} \tag{5}$$

Moment	Data	Pearson type I	Eighth order series
1	0.25295×10^{2}	0.25295×10^{2}	0.25295×10^{2}
2	0.65029×10^{3}	0.65029×10^{3}	0.65029×10^{3}
3	0.16981×10^{5}	0.16981×10^{5}	0.16981×10^{5}
4	0.45018×10^{6}	0.45018×10^{6}	0.45018×10^{6}
5	0.12106×10^{8}	0.12106×10^{8}	0.12106×10^{8}
6	0.32997×10^{9}	0.32995×10^{9}	0.32997×10^{9}
7	0.91086×10^{10}	0.91071×10^{10}	0.91086×10^{10}
8	0.25444×10^{12}	0.25434×10^{12}	0.25443×10^{12}

Table 2. Moments about origine for for a concrete strenght data, Perason type 1 distribution, and 8th orderseries expansion.

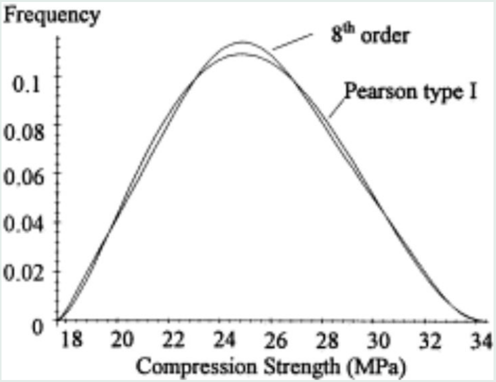

Figure 2. Approximation of a χ^2 distridution (solid line) by a 11th order series expansion based on the lognormal weighting function.

with parameters: $h = 10$; $p = 11$; $0 < x < 11$; $(a, p) = 1$; Υ: *Legendre symbol.*

In this case the PDF has a standard exponential form with parameter $\lambda = 1$ (see [5] for proof).

Let us now consider a much more complicated weighting function to fit our distribution.

For $s \leq a$, $Q = \prod_{k=1}^{s} p_k$, $Q \to \infty$, $h \to \infty$, $\frac{\log h}{\log Q} \to 0$, a discrete type of weighting function $w_t(x, p_1, \ldots, p_s) = \sum_{n=x+1}^{x+t} \left(\frac{n+a_1}{p_1}\right) \ldots \left(\frac{n+a_s}{p_s}\right)$ is proposed; $(\cdot)$ stands for the *Jacobi symbol.*

In this last case the PDF has a standard normal distribution. The theoretical proof of this part can be found in [6].

Some explanations about our proof.

1. First we compute the 2r-th moment $[A_{p_1 \ldots p_s}(2r)]$ of our weighting function

$$w_t(x, p_1, \ldots, p_s) = \sum_{n=x+1}^{x+t} \left(\frac{n+a_1}{p_1}\right) \ldots \left(\frac{n+a_s}{p_s}\right).$$

2. Use the Chinese remainder theorem: $x = Qp_1^{-1}x_1 + \ldots + Qp_s^{-1}x_1 \pmod{Q}$; $Q = p_1 \ldots p_s$.

3. Range x over the complete set of residues modulo p; range x_s over the set of residues modulo p_s and rewrite $w_t(x, p_1, \ldots, p_s)$.

4. Partition of $A_{p_1 \ldots p_s}(2r) = B_1 + B_2$ depending on whether r is odd or even.

5. Moments estimation : $\begin{cases} A_{p_1 \ldots p_s}(2r) = 1 \cdot 3 \ldots \cdot (2r-1) + \mathrm{O}(h^{-1}) \\ A_{p_1 \ldots p_s}(2r-1) \ll h^r Q^{\frac{1}{2a}} \end{cases}$

6. Conclusion

A method of producing probability density functions by choosing appropriate weighting functions on the one hand and by solving the practical moment problem using non-classical orthogonal polynomials on the other hand is presented. The approach which utilizes orthogonal process has potential applications in many area of probability theory, for example: determining frequency distribution from large large data sets, for estimating the probability distributions of extreme events given limited data. The presented examples have shown that stable high order series expansion may be obtained with moments that closely match those which are prescribed and a proper choose of the weighting function highly recommended.

Acknowledgment. H. C. Jimbo gratefully acknowledges the support by the Japan Society for the promotion of Science (JSPS), The Institute of Statistical Mathematics (ISM), The Moscow State University (MGU) and Prof. V. N. Chubarikov for useful advice and discussion. Finally I would like to express all my thanks to the anonymous referees for helpful comments from which some sections of this paper were improved.

1. Burr I. W. *Cumulative frequency functions*. Ann. Math. Stat. **13** (1942), 215.
2. Carlman T., *Les functions quasi-analytiques*, Paris: Gauthier-Villars 1925analytiques.
3. Godwin H. J., *Inequalities on distribution functions*. Griffin, London 1964.
4. Haimes Y. Y, Barry and Lambert J.H. *When and how can you specify a probability distribution when you don't know much?* Risk Analysis **14** (5) (1994), 661–707.
5. Jimbo H. C., *Distribution of absolute values of incomplete Gauss sums*. Moscow University Mathematic Bulletin **56** (2) (2001).
6. Jimbo H. C. and Chubarikov V. N., *On asymptotic distribution of values of arithmetic functions*, Doklady Mathematic, **63** (2), (2001).
7. Leipnik R. *The lognormal distribution and strong non-uniqueness of moment problem*. Theory Prob Applic. **26** (1981), p. 850.
8. Marthur A. K., *Statistical Analysis of concrete compression results: A comparative study of different codal provision on National seminar on Modern Trends in Building Materials*. Design and construction, Institute of Engineers, Allahabad, India, 1987.
9. Ord J. K., *Families of frequency distributions*. Griffin, London 1972.

10. Peck A., Corelick S., de Marsily G., Foster S. and Kavalevsky V., *Consequencees of spatial variability in aquifer properties and data limitation for ground water modeling practice*, IAHS publication No. 175, Oxfordshire UK: IAHS Press 1988.
11. Perk W., *Some observations on inverse probability including a new difference rule*. J. Inst. Actuar **73** (1947), 285–334.
12. Sank A. *An improved Monte Carlo Technique in structural reliability*. Reliability Engng. System safety **24** (1989), 275–92.
13. Shohat J.A. and Tarmakin J.D., *The problem of moments*, Americain Mathematical Society, 1949.
14. Springer M. D., *The algebra on random variables*. Wiley, New York 1979.
15. Stuart A., and Ord J. K., Distribution Theory. Kendell" *Advanced theory of statistics* Vol. 1, 6^{th} ed. Edward Holder, London 1994.
16. Wallace D.C. *Asymptotic approximations to distributions*. Ann Math. Stat **29** (1958), 634–635.

H. C. Jimbo, Institute of Statistical mathematics 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106-8569, Japan, *e-mail*: `jimboc@ism.ac.jp`