CLOSED FORM SUMMATION FOR CLASSICAL DISTRIBUTIONS: VARIATIONS ON A THEME OF DE MOIVRE

BY

PERSI DIACONIS and SANDY ZABELL

TECHNICAL REPORT NO. 375

MAY 1991

PREPARED UNDER THE AUSPICES OF

NATIONAL SCIENCE FOUNDATION GRANT DMS89-05874

DEPARTMENT OF STATISTICS

STANFORD UNIVERSITY

STANFORD, CALIFORNIA
Closed Form Summation for Classical Distributions: Variations on a Theme of De Moivre

By

Persi Diaconis and Sandy Zabell

Technical Report No. 375
May 1991

Prepared Under the Auspices
Of
National Science Foundation Grant DMS89–05874

Department of Statistics
Stanford University
Stanford, California
Abstract

De Moivre gave a simple closed form expression for the mean absolute deviation of the binomial distribution. Later authors showed that similar closed form expressions hold for many of the other classical families. We review the history of these identities and extend them to obtain summation formulae for the restricted expectations of all polynomials orthogonal to the constants.
1. Introduction.

Let \( S_n \) denote the number of successes in \( n \) Bernoulli trials with chance \( p \) of success at each trial. Thus \( P\{S_n = k\} = \binom{n}{k}p^k(1-p)^{n-k} = b(k; n, p) \). In 1730, Abraham De Moivre gave a version of the surprising formula

\[
E\{|S_n - np|\} = 2\nu(1-p)b(\nu; n, p),
\]

where \( \nu \) is the modal term of the binomial; thus \( \nu \) is the unique integer such that \( np < \nu \leq np + 1 \). De Moivre's formula provides a simple closed form expression for the mean absolute deviation or \( L_1 \) distance of a binomial variate from its mean. The identity is surprising because the presence of the absolute value suggests that expressions for the tail sum \( \sum_{k\leq np} b(k; n, p) \) might be involved but, as is well known, there are no essential simplifications of such sums, see, e.g., Zeilberger (1989).

Dividing (1.1) by \( n \), and using the result that the modal term of a binomial tends to zero with increasing \( n \), it follows that

\[
E\left\{ \left| \frac{S_n}{n} - p \right| \right\} \to 0.
\]

De Moivre noted this form of the law of large numbers, and thought it could be employed to justify passing from sample frequencies to population proportions. As he put it, for the case \( p = \frac{1}{2} \) (De Moivre. 1756, p. 240):

**Corollary.** From the foregoing considerations it follows that if after taking a great number of experiments, it should be observed that the happenings or failings of an event have been very near a ratio of equality, it may safely be concluded that the probability of its happening at any one time assigned are very nearly equal.

Understanding the asymptotics of (1.2) in turn led De Moivre to his work on Stirling's formula. In Section 2 we discuss this history and argue that it was his work on this problem which ultimately led to his proof of the central limit theorem.

De Moivre's formula is at once easy enough to derive so that many people have subsequently rediscovered it, but also hard enough to have often been considered worth publishing, varying and generalizing. In Section 3 we review these later results and note
several applications, one to bounding binomial tail sums, one to the Bernstein polynomial version of the Weierstrass approximation theorem, and one to proving the monotonicity of convergence in (1.2).

In the second half of this paper, we offer a generalization along the following lines: De Moivre's result works because \( \sum_{a}^{b} (k - np)b(k, np) \) can be summed in closed form for any \( a \) and \( b \). The function \( x - np \) is the first orthogonal polynomial for the binomial distribution. We show that in fact all orthogonal polynomials (except the zeroth) admit similar closed form summation. The same result holds for many of the other standard families (normal, gamma, beta, and Poisson). There are a number of interesting applications of these results which we discuss, and in particular, there is a surprising connection with Stein's characterization of the normal and other classical distributions.

De Moivre's formula arose out of his attempt to answer a question of Sir Alexander Cuming. Cuming was a colorful character whose life is discussed in an Appendix.

**PART I: DE MOIVRE'S FORMULA AND ITS DESCENDANTS**

2. Cuming's Problem and De Moivre's \( L_1 \) Limit Theorem.

Abraham De Moivre wrote one of the first great books on probability, *The Doctrine of Chances*. First published in 1718, with important new editions in 1738 and 1756, it contains scores of important results, many in essentially their modern formulation. Most of the problems considered by De Moivre concern questions that arise naturally in the gambling context. Problem 72 of the third edition struck us somewhat differently:

A and B playing together and having an equal number of Chances to win one Game, engage themselves to a Spectator S that after an even number of Games is over, the Winner shall give him as many Pieces as he wins Games over and above one half the number of Games played, it is demanded how the Expectation of S is to be determined.

In modern notation, De Moivre is asking for the expectation \( E\{|S_n - n/2|\} \). In *The Doctrine of Chances*, De Moivre states that the answer to the question is \( \frac{n}{2} E/2^n \) where
$E$ is the middle term in the binomial expansion of $(1 + 1)^n$, that is, \( \binom{n}{n/2} \). De Moivre illustrates this result for the case $n = 6$ (when $E = 20$ and the expectation is $15/16$).

At the conclusion of Problem 72, De Moivre gives as a corollary the version of the law of large numbers quoted earlier. Problem 73 of The Doctrine of Chances essentially gives formula (1.1) for general values of $p$ (De Moivre worked with rational numbers). Immediately following this De Moivre moves on to the central limit theorem.

We were intrigued by De Moivre’s formula. Where had it come from? Problem 72, where it appears, is scarcely a question of natural interest to the gamblers De Moivre might have spoken to, unlike most of the preceding questions discussed in the Doctrine. And where had it gone? Its statement is certainly not one of the standard identities one learns today.

2.1 The Problem of Sir Alexander Cuming.

Neither the problem nor the formula appear in the 1718 edition of The Doctrine of Chances. They are first mentioned by De Moivre in his Miscellanea Analytica of 1730, a Latin work summarizing his mathematical research over the preceding decade (De Moivre 1730). De Moivre states there (p. 99) that the problem was initially posed to him in 1721 by Sir Alexander Cuming, a member of the Royal Society.

In the Miscellanea Analytica De Moivre gives the solution to Cuming’s problem (at pp. 99-101), including a proof of the formula in the symmetric case (given below in Section 2.3), but he contents himself with simply stating without proof the corresponding result for the asymmetric case. These two cases then appear as Problems 86 and 87 in the 1738 edition of the Doctrine of Chances, and Problems 72 and 73 in the 1756 edition.

As De Moivre notes in the Doctrine of Chances (1756, pp. 240-241), the expectation of $S =: |S_n - np|$ increases with $n$, but decreases proportionately to $n$; thus he notes for
\[ p = \frac{1}{2} \] that

<table>
<thead>
<tr>
<th>( n )</th>
<th>( E[S] )</th>
<th>( E[S/n] )</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>0.9375</td>
<td>0.1563</td>
</tr>
<tr>
<td>12</td>
<td>1.3535</td>
<td>0.1128</td>
</tr>
<tr>
<td>100</td>
<td>3.9795</td>
<td>0.0398</td>
</tr>
<tr>
<td>200</td>
<td>5.63348</td>
<td>0.0282</td>
</tr>
<tr>
<td>300</td>
<td>6.9041</td>
<td>0.0230</td>
</tr>
<tr>
<td>400</td>
<td>7.97389</td>
<td>0.0199</td>
</tr>
<tr>
<td>500</td>
<td>8.91612</td>
<td>0.0178</td>
</tr>
<tr>
<td>700</td>
<td>10.5521</td>
<td>0.0151</td>
</tr>
<tr>
<td>800</td>
<td>11.280</td>
<td>0.0141</td>
</tr>
<tr>
<td>900</td>
<td>11.965</td>
<td>0.0133</td>
</tr>
</tbody>
</table>

For a proof that \( E[S] \) is increasing in \( n \), and \( E[S/n] \) is decreasing in \( n \), see Theorem 3 in Section 3.2 below. De Moivre does not give a proof in either the symmetric or asymmetric cases, and it is unclear whether he had one, or even whether he intended to assert monotonicity rather than simply limiting behavior.

Had De Moivre proceeded no further than this, his formula would have remained merely an interesting curiosity. But, as we shall now argue, a compelling case can be made that De Moivre’s work on Cuming’s problem led directly to his later breakthrough on the central limit problem; and here, too, the enigmatic Sir Alexander Cuming played a brief, but vital role.

2.2 “... the hardest Problem that can be proposed on the subject of Chance.”

After stating the Corollary quoted earlier, De Moivre noted that substantial fluctuations of \( S_n/n \) from \( p \), even if unlikely, were still possible, and that is was desirable, therefore, that “the Odds against so great a variation ... should be assigned”; a problem which he described as “the hardest Problem that can be proposed on the subject of Chance” (De Moivre, 1756, p. 240).

But initially, perhaps precisely because he viewed the problem as being so difficult, De Moivre had little interest in working on the questions raised by Bernoulli’s theorem. No discussion of Bernoulli’s work occurs in the first edition of the *Doctrine of Chances*; and in its preface De Moivre even states that despite the urging of both Montmort and Nicholas Bernoulli to do so, “I willing resign my share of that Task into better Hands” (De Moivre, 1718, p. xiv).
But now, only a few years later in 1721, as a result of his work on Cuming’s problem, De Moivre knew empirically that \( D_n = \frac{E[|S_n - p|]}{n} \) was decreasing in \( n \); and because of his formula, he also knew that in order to demonstrate this, at least asymptotically, all that was necessary was an understanding of the limiting behavior of the central term in the binomial expansion (unlike Bernoulli’s approach, which involved first approximating and then summing all terms in the tail). Thus motivated, De Moivre set to work, and later in the same year his efforts were crowned with success.

The asymptotic formula for the central term that De Moivre discovered in 1721 was 
\[
E_n \sim \frac{2}{B\sqrt{n}},
\]
where \( B \) is a constant, and the notation \( a_n \sim b_n \) indicates that \( a_n/b_n \to 1 \) as \( n \to \infty \) (De Moivre, 1730, pp. 125-129, De Moivre, 1756, pp. 243-244; see generally Stigler, 1986, pp. 70-77 and Hald, 1990, pp. 472-480). The constant \( B \), however, was a puzzlement. De Moivre was initially unable to obtain an exact value for it, but obtained instead the series “expansion”

\[
\log B = 1 - \frac{1}{12} + \frac{1}{360} - \frac{1}{1260} + \frac{1}{1680} - \ldots
\]

What De Moivre had run up against, of course, was a divergent series – here

\[
= 1 - \sum_{n=1}^{\infty} \frac{B_{2n}}{(2n - 1)(2n)},
\]

where \( B_k \) denotes the \( k \)-th Bernoulli number – arising from a valid asymptotic expansion; but more than a century and a half would have to elapse before such phenomena became well understood and, not surprisingly, De Moivre found himself unable to progress further.

It is instructive to read his comments 12 years later, in 1733:

When I first began that inquiry, I contented myself to determine at large the Value of \( B \), which was done by the addition of some Terms of the above-written Series; but as I perceived that it converged but slowly, and seeing at the same time that what I had done answered my purpose tolerably well, I desisted from proceeding farther till my worthy and learned Friend Mr. James Stirling, who had applied himself after me to that inquiry, found that the Quantity \( B \) did denote the Square-root of the Circumference of a Circle whose Radius is Unity, so that
if that Circumference be called \( c \), the Ratio of the middle Term to the Sum of all the Terms will be expressed by \( 2/\sqrt{\pi n} \). [DOC, p. 244]

De Moivre’s comment that he had “perceived that [the series] converged but slowly”, seems somewhat disingenuous: in the *Miscellanea analytica* De Moivre gave only the first four terms of the series and, as noted by Hald (1990, pp. 475-476), it is precisely at this point that the subsequent terms in the series begin to *increase*. That De Moivre was indeed aware of a serious difficulty is clear for, as he himself says, he “desisted from proceeding farther” until Stirling’s breakthrough nearly a decade later; and he turned to other lines of attack, which are also described in the *Miscellanea analytica*.

But what led James Stirling to take an interest in the problem? Here, once again, Sir Alexander Cuming briefly enters the scene. In 1725 a young James Stirling had returned to England after nearly a decade abroad; and shortly after his arrival in London his fellow Scotsman Cuming – no doubt knowing that De Moivre was stymied – succeeded in interesting Stirling in the problem of the evaluation of the binomial coefficients (as Stirling tells us in the preface to his *Methodus differentialis*). Stirling, who was an analyst of considerable power (he might be better known today had he not later turned to a highly successfully industrial career), soon made rapid progress on the problem; and learning this De Moivre – no doubt fearful of being “scooped” – moved quickly to publish his own work. This in turn appears to have led Stirling to publish his own results much earlier than he otherwise would; and the result was the joint appearance in 1730 of both De Moivre’s *Miscellanea analytica* and Stirling’s *Methodus differentialis*.

This (happily friendly) rivalry was to have far-reaching consequences: the table of logarithms in the hastily published *Miscellanea analytica* contained serious errors, to which Stirling quickly alerted De Moivre (“... Dr. Stirling has admonished me ....”, noted a rueful De Moivre in a corrective supplement written shortly after.) Stirling was able to easily detect such errors, for he possessed an asymptotic expansion for \( \log_e n! \) which enabled him to compute it with great efficiency and precision. Informed of this by Stirling, De Moivre soon realized
that Stirling's approximation to $\log_e n!$ could be applied to the problem of understanding the asymptotic behavior of individual binomial terms, with sufficient control on the error of approximation to permit passing to an approximation for the sum of such terms in the binomial tail, and thence on - by 1733 - to the central limit theorem.

All this thus sprang from De Moivre's empirical observation that $M_n/n$ was decreasing, and that it provided an alternate approach to Bernoulli's law of large numbers. But there is one feature about De Moivre's conclusion that is puzzling. How did he make the leap from

$$E \left\{ \left| \frac{S_n}{n} - p \right| \right\} \to 0 \quad \text{to} \quad P \left\{ \left| \frac{S_n}{n} - p \right| > \epsilon \right\} \to 0?$$

De Moivre certainly knew the second statement from his work on the central limit theorem and from Bernoulli's earlier work on the law of large numbers. But more than 120 years would have to elapse before Chebychev's inequality would allow one to easily reach the second conclusion from the first.

Of course, the currently recognized modes of convergence were not well delineated in De Moivre's time. One can find him sliding between the weak and strong laws in several places. His statement of the corollary: "the happenings or failings of an event have been very near a ratio of equality" has a clear element of fluctuation in it. In contrast, even today $L_1$ convergence has a distant, mathematical flavor to it. It is intriguing that De Moivre seemed to give it such a direct interpretation.

2.3 De Moivre's Proof.

De Moivre's proof that $E[|S_n - n/2|] = (1/2)nE/2^n$ is simple but clever, impressive if only because of the notational infirmities of his day. Since it only appears in the Latin of the Miscellanea Analytica and is omitted from The Doctrine of Chances, we reproduce the argument here.

DE MOIVRE'S PROOF OF FORMULA (1.1), CASE $p = 1/2$: Let $E$ denote the "median term" (terminus medius) in the expansion of $(a + b)^n$, $D$ and $F$ the
coefficients on either side of this term, $C$ and $G$ the next pair on either side, and so on. Thus the terms are ..., $A, B, C, D, E, F, G, H, K, ...$ (where De Moivre omits the letter “$I$”, presumably to avoid confusion with the numeral “1”).

The expectation of the spectator after an even number of games is

$$E \times 0 + (D + F) \times 1 + (C + G) \times 2 + (B + H) \times 3 + (A + K) \times 4 + ...$$

Because the binomial coefficients at an equal distance from either side of the middle are equal, the expectation of the spectator reduces to

$$E \times 0 + 2D + 4C + 6B + 8A + ...$$

But owing to the properties of the coefficients, it follows that

$$(n + 2)D = nE$$

$$(n + 4)C = (n - 2)D$$

$$(n + 6)B = (n - 4)C$$

$$(n + 8)A = (n - 6)B$$

...$

Setting equal the sum of the two columns then yields

$$nD + nC + nB + nA + ... = nE + nD + nC + nB + nA + ...$$

$$+2D + 4C + 6B + 8A + ... = -2D - 4C - 6B - 8A - ...$$

Deleting equal terms from each side, and transposing the remainder, we have

$$4D + 8C + 12B + 16A + ... = nE$$

or

$$2D + 4C + 6B + 8A + ... = \frac{1}{2} nE.$$

Since the probabilities corresponding to each coefficient result from dividing by $(a + b)^n$, here $(1 + 1)^n = 2^n$, De Moivre’s theorem follows.
NOTE: For a mathematician of his stature, surprisingly little has been written about De Moivre. Helen Walker's brief article in *Scripta Mathematica* (Walker, 1934) gives the primary sources for the known details of De Moivre's life; other accounts include those of David (1962, pp. 161-178), Pearson (1978, pp. 141-146), and the *Dictionary of Scientific Biography*.

Ivo Schneider's detailed study (Schneider, 1968) provides a comprehensive survey of De Moivre's mathematical research. During the last two decades many books and papers have begun to appear on this history of probability and statistics, and a number of these provide extensive discussion and commentary on this specific aspect of De Moivre's work; these include, most notably, the books by Stigler (1986) and Hald (1990). Other useful discussions include those of Daw and Pearson (1972), Adams (1974), Pearson (1978, pp. 146-166), Hald (1984 and 1988), and Daston (1988, pp. 250-253).
3. Later proofs, applications, and extensions.

3.1 Later Proofs. De Moivre did not give a proof of his expression for the MAD in the case of the asymmetrical binomial (although he must have known one). This gap was filled by Todhunter (1865, pp. 182-183) who supplied a proof in his discussion of this portion of De Moivre’s work.

Todhunter’s proof proceeds by giving a closed form expression for a sum of terms in the expectation, where the sum is taken from the outside in. We abstract the key identity in modern notation.

**Lemma 1 (Todhunter's Formula).** For all integers $0 \leq \alpha < \beta \leq n$,

$$\sum_{k=\alpha}^{\beta} (k - np)b(k; n, p) = \alpha q b(\alpha; n, p) - (n - \beta)p b(\beta; n, p).$$

**Proof:** Because $p + q = 1$,

$$\sum_{k=\alpha}^{\beta} (k - np)b(k; n, p) = \sum_{k=\alpha}^{\beta} \{kq - (n - k)p\}b(k; n, p)$$

$$= \sum_{k=\alpha}^{\beta} kq b(k; n, p) - \sum_{k=\alpha}^{\beta} (n - k)p b(k; n, p),$$

but $(k + 1)q b(k + 1; n, p) = (n - k)p b(k; n, p)$; thus every term in the first sum (except the lead term) is canceled by the preceding term in the second sum, and the lemma follows.

□

We know of no proof for the $p \neq 1/2$ case prior to that given in Todhunter’s book. Todhunter had an encyclopedic knowledge of the literature, and it would have been consistent with his usual practice to mention further work on the subject if it existed. He (in effect) proved his formula by induction.

Todhunter assumed, however, as did De Moivre, that $np$ is integral (although his proof does not really require this); and this restriction can also be found in Bertrand (1889, pp. 82-83). Bertrand noted that if

$$F(p, q) = \sum_{k > np} (n C_k) p^k q^{n-k},$$

10
then the mean absolute deviation could be expressed as \(2pq\left(\frac{\partial F}{\partial p} - \frac{\partial F}{\partial q}\right)\), and that term-by-term cancellation then leads to De Moivre's formula. The first discussion we know of giving the general formula without any restriction is in Poincaré's book (1896, pp. 56-60; 1912, pp. 79-83): if \(\nu\) is the first integer greater than \(np\), then the mean absolute deviation is given by \(2\nu qb(\nu; n, p)\). Poincaré's derivation is based on Bertrand's, but is a curiously fussy attempt to fill what he apparently viewed as logical lacunae in Bertrand's proof. It later appears in Uspensky's book as a problem (Uspensky, 1937, pp. 176-177), possibly by the route Poincaré (1896) \(\rightarrow\) Czuber (1914, pp. 146-147) \(\rightarrow\) Uspensky (1937).

De Moivre's identity has been rediscovered many times since. Frisch (1924, p. 161) gives the Todhunter formula and deduces the binomial MAD formula as an immediate consequence. It soon after entered the textbook literature (see, e.g., Feller (1968, p. 245)). This did not stem the flow of rediscovery, however. In 1930, Gruder (1930) rediscovered Todhunter's formula and in 1957, N.L. Johnson, citing Gruder, noted its application to the binomial MAD. Johnson's article triggered a series of generalizations. The MAD formula was also published in Frame (1945). None of these authors connected the identity to the law of large numbers so it remained a curious fact.

De Moivre's formula can be applied outside the realm of limit theorems. In a charming article, Blyth (1980) notes that the closed form expansion for the MAD has a number of interesting applications. If \(X\) is a binomial random variable with parameters \(n\) and \(p\), the deviation \(E|\frac{X}{n} - p|\) represents the risk of the maximum likelihood estimator under absolute value loss. A plot of this risk as a function of \(p\) has the appearance of the edge of a sea shell. As \(p\) varies between 0 and \(\frac{1}{2}\), the risk is roughly monotone but if \(n = 4, p = \frac{1}{4}\), the estimate does better than for nearby values of \(p\). Lehmann (1983, p. 58) gives De Moivre's identity with Blyth's application.

NOTES: 1. The formula for the mean absolute deviation of the binomial distribution can be expressed in several equivalent forms which are found in the literature. If \(\nu\) is the least integer greater than \(np\) and \(Y_{n,p}\) is the central term in
the expansion of \((p + q)^n\), then the formula for the mean absolute deviation can be written

\[ 2
\nu
q
b(\nu; n, p) \]  

(Poincaré, 1896; Frisch, 1924; Feller, 1968)

\[ 2npqb(\nu - 1; n - 1, p) \]  

(Uspensky, 1937)

\[ 2npqY_{n-1} \]  

(Frame, 1945)

\[ \nu(\nu C_\nu) p^n q^{n-\nu+1} \]  

(Johnson, 1957).

In his solution to Problem 73, De Moivre states that one should use the binomial term \(b(j; n, p)\) for which \(j/(n - j) = p/(1 - p)\); since this is equivalent to taking \(j = np\), the solution tacitly assumes that \(np\) is integral. In this case \(b(j; n, p) = b(j; n - 1, p)\) and \(j = \nu - 1\), hence

\[ 2npqb(j; n, p) = 2npqb(\nu - 1; n - 1; p); \]

thus the formula given by De Moivre agrees with the first of the standard forms.

### 3.2 Applications

1. As a first application we give a binomial version of Mills ratio for binomial tail probabilities.

**Theorem 1.** For \(\alpha > np\), all \(n \geq 1\) and all \(p \in (0, 1)\),

\[ \frac{\alpha}{n} \leq \frac{1}{b(\alpha, n, p)} \sum_{k=\alpha}^{n} b(k, n, p) \leq \frac{\alpha(1 - p)}{\alpha - np}. \]

**Proof:** For the upper bound, use Lemma 1 to see that

\[ \alpha \sum_{k=\alpha}^{n} b(k, n, p) \leq \sum_{k=\alpha}^{n} kb(k, n, p) = np \sum_{k=\alpha}^{n} b(k, n, p) + \alpha qb(\alpha, n, p). \]

The lower bound follows similarly.

**Remarks:** The upper bound is given in Feller (1968, p. 151). Feller gives a much cruder lower bound. Slightly stronger results follow from Markov's continued
fraction approach, see Uspensky (1937, pp. 52-56). As usual, this bound is poorest when \( \alpha \) is close to \( np \). For example, when \( p = \frac{1}{2} \), and \( \alpha = \left[ \frac{n}{2} \right] + 1 \), the ratio is of order \( \sqrt{n} \) while the lower bound is approximately \( \frac{1}{2} \) and the upper bound is approximately \( \frac{n}{4} \). The bound is useful in the tails. Similar bounds follow for the other families which admit a closed form expression for the mean absolute deviation.

2. De Moivre’s formula allows a simple evaluation of the error term in the Bernstein polynomial approximation to a continuous function. Lorentz (1986) or Feller (1971, Chap. 8) give the background to Bernstein’s approach.

Let \( f \) be a continuous function on \([0, 1]\). Bernstein’s proof of the Weierstrass approximation theorem approximates \( f(x) \) by the Bernstein polynomial

\[
B(x) = \sum_{i=0}^{n} f\left(\frac{i}{n}\right) \binom{n}{i} x^i (1-x)^{n-i}.
\]

(3.1)

The quality of approximation is often measured in terms of the modulus of continuity:

\[
\omega_f(\delta) = \sup_{|x-y| \leq \delta} |f(y) - f(x)|.
\]

With this notation, we can state

**Theorem 2.** Let \( f \) be a continuous function on \([0, 1]\). Then for any \( x \in [0, 1] \)

\[
|f(x) - B(x)| \leq \left(1 + \frac{1}{2\sqrt{n}}\right) \omega_f\left(\frac{1}{\sqrt{n}}\right).
\]

**Proof:** Clearly

\[
|f(x) - B(x)| \leq \sum_{i=0}^{n} |f(x) - f\left(\frac{i}{n}\right)| \binom{n}{i} x^i (1-x)^{n-i}.
\]

For any \( \delta \in (0, 1) \), dividing the interval between \( x \) and \( \frac{i}{n} \) into subintervals of length smaller than \( \delta \) shows

\[
|f(x) - f\left(\frac{i}{n}\right)| \leq \omega_f(\delta)(1 + \frac{|x - \frac{i}{n}|}{\delta}).
\]
Using this and De Moivre’s formula gives

\[
|f(x) - B(x)| \leq \omega_f(\delta) \left( 1 + \frac{2x(1-x)}{n\delta} \right) \left( \frac{n-1}{r-1} \right)^{r-1} (1-x)^{n-(r-1)}
\]

where \(nx \leq r \leq nx + 1\), and \(r\) is an integer depending on \(x\).

The theorem follows on taking \(\delta = 1/\sqrt{n}\).

**Remarks:**

1) Lorentz (1986, pp. 20,21) gives \(|f(x) - B(x)| \leq \frac{5}{4} \omega_f\left(\frac{1}{\sqrt{n}}\right)\). Theorem 2 gives a slight improvement on this result for \(n > 4\). Lorentz shows that the function \(f(x) = |x - \frac{1}{2}|\) has \(|f(x) - B(x)| \geq \frac{1}{2} \omega_f\left(\frac{1}{\sqrt{n}}\right)\) so the \(\frac{1}{\sqrt{n}}\) rate is best possible.

2) Bernstein polynomials are useful in Bayesian statistics because of their interpretation as mixtures of beta distributions; see Dallal and Hall (1983), Diaconis and Ylvisaker (1985). The identities for other families presented in Section 4 can be employed to give similar bounds for mixtures of other families of conjugate priors.

3) The pointwise bound (3.2), or the simplification

\[
|f(x) - B(x)| \leq \omega_f(\delta)(1 + \frac{2x(1-x)}{n\delta})(1-x)
\]

have a direct appeal. Integrating gives the \(L_1\) bound:

\[
\|f - B\|_{L_1} \leq \omega_f(\delta)(1 + \frac{1}{3n\delta}) \quad \text{for all} \quad \delta \in (0,1).
\]

For “ordinary” functions \(\omega_f(\delta) \sim c\delta\) as \(\delta \to 0\) with \(c = \sup |f'(x)|\).

3. As a final application, we apply the general form of De Moivre’s formula (1.1) to show that the MAD of \(S_n\) is increasing in \(n\), but that the MAD of \(S_n/n\) is decreasing in \(n\). Recall that for \(S_n\), the modal term \(\nu_n = \lfloor np + 1 \rfloor = [np] + 1\), so that \(np < \nu_n \leq np + 1\).

**Theorem 3.** Let \(S_n \sim B(n,p)\) and \(M_n =: E[|S_n - np|]\). If \(p\) is fixed, then for every \(n \geq 1\),

\[
M_n \leq M_{n+1}, \text{ with equality precisely when } (n+1)p \text{ is integral;}
\]
(3.4) \( \frac{M_n}{n} \geq \frac{M_{n+1}}{n+1} \), with equality precisely when \( np \) is integral.

PROOF: It is necessary to consider two cases.

Case 1: \( \nu_n = \nu_{n+1} \). Then by the general form of De Moivre's formula,

\[
\frac{M_{n+1}}{M_n} = \frac{(n+1)q}{n+1 - \nu_n} \quad \text{and} \quad \frac{M_{n+1}/(n+1)}{M_n/n} = \frac{nq}{n+1 - \nu_n}.
\]

But \((n+1)p < [(n+1)p + 1] = \nu_{n+1} = \nu_n\), hence \( n + 1 - \nu_n < (n+1)q \), so that \( M_{n+1}/M_n > 1 \). Similarly, \( \nu_n \leq np + 1 \), hence \( nq \leq n + 1 - \nu_n \), and inequality (3.4) follows, with equality if and only if \( np + 1 \), hence \( np \), is integral.

Case 2: \( \nu_n < \nu_{n+1} \), clearly \( \nu_n = \nu_{n+1} - 1 = [(n+1)p] \leq (n+1)p \), and inequality (3.3) follows, with equality if and only if \((n+1)p\) is integral. Since \( np < \nu_n \), inequality (3.4) follows immediately, and the inequality is strict.

Since \( np \) integral implies \( \nu_n = \nu_{n+1} \), and \((n+1)p\) integral implies \( \nu_n < \nu_{n+1} \), the theorem follows. \( \square \)

3.3 Extensions to Other Families. De Moivre identity can be stated thus: For a binomial variate, the mean absolute deviation equals twice the variance times the density at the mode. It is natural to inquire whether such a simple relationship exists between the variance \( \sigma^2 \) and the mean absolute deviation \( \mu_1 \) for families other than the binomial. This simple question appears to have been first asked and answered in 1923 by Ladislaus von Bortkiewicz. If \( f(x) \) is the density function of a continuous distribution with expectation \( \mu \), von Bortkiewicz showed that the ratio \( R =: \mu_1/2\sigma^2 f(\mu) \) is unity for the gamma (“De Forestche”), normal (“Gaussche”), chi-squared (“Helmertscbe”), and exponential (“zufälligen Abstände massgebende”) distributions (“Fehlergesetz”); while it is \((\alpha + \beta + 1)/(\alpha + \beta)\) for the beta distribution (“Pearsonsche Fehlergesetz”) with parameters \( \alpha \) and \( \beta \).

Although von Bortkiewicz did not explain his motivation for considering this question, we have a conjecture. In 1914 the astronomer Arthur Eddington had asserted in passing the theoretical superiority of the sample mean absolute deviation \( \sigma_1 =: \sum |x_i - \bar{x}|/n \) over the sample standard deviation \( \sigma_2 =: \)
\[ \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}} \] in estimating the population standard deviation \( \sigma \) (Eddington, 1914). In 1921 a young R.A. Fisher countered Eddington's argument in a now famous paper in which the concept of sufficiency was first introduced (Fisher, 1921; see generally Stigler, 1973). Fisher's theoretical claim for the efficiency of \( \sigma_2 \) over \( \sigma_1 \) was based on the assumption that the underlying error structure was normal; and although in a footnote to Fisher's paper Eddington conceded Fisher's theoretical point, he also suggested that the departures from normality encountered in astronomical observations made the use of \( \sigma_1 \) preferable. Eddington's argument raises the interesting question of what the possible relationships are between \( \mu_1 \) and \( \sigma \), and an attempt to determine when the relationship that obtains in the normal case still holds in others may have led von Bortkiewicz to raise the question he did two years after Fisher's paper appeared.

Shortly after von Bortkiewicz's paper appeared, Karl Pearson noted that the continuous examples considered by von Bortkiewicz could be treated in a unified fashion by observing that they were all members of the Pearson family of curves (Pearson, 1924). If \( f(x) \) is the density function of a continuous distribution, then \( f(x) \) is a member of this family if it satisfies the differential equation

\begin{equation}
\frac{f'(x)}{f(x)} = \frac{x + a}{b_0 + b_1 x + b_2 x^2}.
\end{equation}

Then, letting \( p(x) = b_0 + b_1 x + b_2 x^2 \), it follows that

\[ (fp)'(x) = f(x)((1 + 2b_2)x + (a + b_1)). \]

If \( b_2 \neq -\frac{1}{2} \), and \( f(x)p(x) \to 0 \) as \( x \to \pm \infty \), then integrating from \( -\infty \) to \( \infty \) yields \( \mu = -(a + b_1)/(1 + 2b_2) \), so that

\[ f(x\{x - \mu\} = \frac{(fp)'(x)}{1 + 2b_2} \]

and

\begin{equation}
\int_{-\infty}^{t} (x - \mu)f(x)dx = \frac{f(t)p(t)}{1 + 2b_2}.
\end{equation}

This gives the following result.
Proposition 1. If \( f \) is a density from the Pearson family (3.3) with mean \( \mu \), then
\[
\int_{-\infty}^{\infty} |x - \mu| f(x) \, dx = -2f(\mu) \left\{ \frac{b_0 + b_1 \mu + b_2 \mu^2}{1 + 2b_2} \right\}.
\]

Remark: If \( \beta_1 =: \mu_3^2/\mu_2^3 \) and \( \beta_2 =: \mu_4/\mu_2^2 \) denote the coefficients of skewness and kurtosis, then, as Pearson showed, this last expression may be re-expressed as
\[
\mu_1 = C \{2 \sigma^2 f(\mu)\}, \quad \text{where} \quad C = \left\{ \frac{4\beta_2 - 3\beta_1}{6(\beta_2 - \beta_1 - 1)} \right\}.
\]
The constant \( C = 1 \iff 2\beta_2 - 3\beta_1 - 6 = 0 \), which is the case when the underlying distribution is normal or Type III (gamma). We give further results for Pearson curves in the next section.

Just as with De Moivre's calculation of the MAD for the binomial, the von Bortkiewicz-Pearson formulas were promptly forgotten and later rediscovered. Ironically, this would happen in Pearson's own journal. After the appearance in 1957 of N.L. Johnson's paper in *Biometrika*, a series of further papers appeared over the next decade which in turn rediscovered the results of von Bortkiewicz and Pearson: Ramasubban (1958) in the case of the Poisson distribution; Kamat (1965, 1966a) in the case of the Pearson family; see also the articles by Bardwell (1960), and Kamat (1966b).
PART II: CLOSED FORM SUMMATION FOR CLASSICAL DISTRIBUTIONS


De Moivre's identity follows from a closed form expression for the partial sum
\[ \sum_{k=0}^{a} (k - np)b(k; n, p). \]
The function \( k \mapsto k - np \) is the first orthogonal polynomial for the binomial distribution. In this section we show that all of the orthogonal polynomials, except the zeroth, admit similar closed form partial sums and that the same holds true for the other classical distributions as well.

Passage to the limit shows such identities must hold for the orthogonal polynomials associated to the normal distribution (the so-called "Hermite polynomials"). The arguments are clearest here so we begin with this case in Section 4.1. A variety of applications are presented. Most notably, the identities give a singular value decomposition for an operator associated to "Stein's method" for proving limit theorems and finding unbiased estimates of risk. In Sections 4.2 and 4.3 we then show how very similar arguments permit the derivation of corresponding results in the case of the gamma and beta distributions, where the appropriate orthogonal polynomials are the Laguerre and Jacobi polynomials, respectively.

The occurrence of these three special families of orthogonal polynomials and distributions is not an accident. The Hermite, Laguerre, and Jacobi polynomials form the three classical families of orthogonal polynomials, known to satisfy many important and special properties; and the normal, gamma, and beta families are precisely those members of the Pearson family for which orthogonal polynomials of all orders exist. This connection is spelled out in Section 5.

Finally, corresponding results are derived for two families of discrete distributions: in Section 6.1 we discuss the Poisson distribution, and then in Section 6.2, we finally return to where we began: the binomial.

4.1 The Normal Density and Hermite Polynomials. A familiar theorem
says that the integral
\[
\int_{-\infty}^{a} e^{-x^2/2} dx
\]
cannot be written as an elementary function of \(a\). Rosenlicht (1972) gives an accessible account in modern language. Of course certain indefinite normal integrals can be simply evaluated, \(\int_{-\infty}^{a} xe^{-x^2/2} dx\), for example. The following lemma determines all polynomials whose integral can be so evaluated.

Recall first that the Hermite polynomials are the orthogonal polynomials on \(\mathbb{R}\) with respect to the kernel \(e^{-x^2/2} dx\). They are given by the explicit formula

\[
(4.1) \quad H_n(x) = n! \sum_{k=0}^{[\frac{n}{2}]} \frac{(-1)^k (\sqrt{2})^{n-2k}}{(n-2k)! k!}, \quad n = 0, 1, 2, \ldots
\]

Thus \(H_0 = 1, H_1 = \sqrt{2}x, H_2 = 2(x^2 - 1), H_3 = 2\sqrt{2}(x^3 - 3x), \ldots\). They satisfy the relation

\[
(4.2) \quad \int_{-\infty}^{\infty} H_r(x)H_s(x)e^{-x^2/2} = \sqrt{2\pi}r!r\delta_{rs}.
\]

Background and standard properties of Hermite polynomials can be found in Chihara (1978), an excellent introduction to the subject of orthogonal polynomials.

The basic identity needed is the following:

**Lemma 1.** For \(n \geq 1\) and any real \(a\),

\[
\int_{-\infty}^{a} H_n(x)e^{-x^2/2} dx = -\sqrt{2}H_{n-1}(a)e^{-a^2/2}.
\]

**Proof:** The Hermite polynomials can be represented by a Rodrigues type formula as

\[
H_n(x) = (-\sqrt{2})^n e^{x^2/2} D^n e^{-x^2/2}
\]

with \(D^n\) denoting \(n\)-fold differentiation. From this

\[
\int_{-\infty}^{a} H_n(x)e^{-x^2/2} dx = (-\sqrt{2})^n \int_{-\infty}^{a} D^n e^{-x^2/2} dx = -\sqrt{2}H_{n-1}(x) e^{-x^2/2}\bigg|_{-\infty}^{a}.
\]

\(\square\)
Corollary 1. A polynomial $p(x)$ can be integrated against $e^{-x^2/2}$ in finite terms if and only if $p(x)$ is orthogonal to the constants in $L^2(e^{-x^2/2})$.

Example 1: The analog of De Moivre's identity for the standard normal distribution takes the form

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} |x|e^{-x^2/2} \, dx = \frac{2}{\sqrt{2\pi}};$$

it follows as a special case of Lemma 4.1.1, with $n = 1$ and $a = 0$.

Example 2: In deriving a total variation bound on the binomial approximation to the hypergeometric distribution, Diaconis and Freedman (1980) encountered the identity

$$\int_{-\infty}^{\infty} |x^2 - 1|e^{-x^2/2} \, dx = 4e^{-2}.$$

It seemed surprising that such a normal integral could be evaluated. The Corollary clarifies the reason why such a formula exists: the polynomial part of the integrand involves $H_2(x)/2$.

Example 3 (Stein's Method): Charles Stein has used the following characterization of the normal distribution as part of his approach to proving limit theorems and in deriving his unbiased estimate of risk in statistics.

Lemma (Stein, 1986). A random variable $Z$ has a standard normal distribution if and only if

$$E(f'(Z)) = E(Zf(Z))$$

for every smooth function $f$ of compact support.

Stein's argument involves study of the differential equation

$$g(x) = f'(x) - xf(x).$$

Here $g$ is given with mean 0: $\int_{-\infty}^{\infty} g(x)e^{-x^2/2} = 0$ and $f$ is sought. The equation is solved in closed form by

$$\frac{d}{dx} \left( x \int_{-\infty}^{x} g(t)e^{-t^2/2} \, dt \right) = 0,$$

$$g(x) = f(x) = e^{x^2/2} \int_{-\infty}^{x} g(t)e^{-t^2/2} \, dt.$$
The operator $U$ mapping $g$ to $f$ sends $L^2_0(e^{-\nu^2/2})$ into $L^2(e^{-\nu^2/2})$, where the subscript $0$ denotes that part of $L^2$ orthogonal to the constants. Lemma 1 can be employed to give a singular value decomposition for the operator $U$. By (4.2), the functions $e_n(x) = H_n(x)/(\sqrt{2\pi}2^n n!)^{1/2}$ are an orthonormal basis for $L^2(e^{-\nu^2/2})$; Lemma 1 then yields

**Corollary 2.** The operator $U$ defined by (4.3) is a bounded linear operator from $L^2_0(e^{-\nu^2/2})$ into $L^2(e^{-\nu^2/2})$. If $\{e_n\}_{n=1}^\infty$ and $\{e_n\}_{n=0}^\infty$ are taken as orthonormal bases of these spaces then $U$ satisfies

$$Ue_n = \frac{-1}{\sqrt{n}}e_{n-1}.$$ 

**Remark:** The bases $\{e_n\}_{n \geq 1}$ and $\{e_n\}_{n \geq 0}$ give a singular value decomposition of $U$ with singular values $-\frac{1}{\sqrt{n}}$. We hope to use this decomposition to study the stability of Stein's characterization of the normal distribution. Corollary 2 may be used in conjunction with Stein's approach to give bounds and approximations for moments. Indeed, Stein (1986, p. 13, Eq. 33) gives the expression

$$Eh = E_0h + E((T \circ \alpha - iT_0) \circ U_0)h.$$ 

With the expectation on the left being the basic object of study, $E_0$ the normal expectation, and $(T \circ \alpha - iT_0)$ a simple operator. The operator $U_0$ is our $U$ above. Taking $h$ as the Hermite polynomials $e_n$ gives explicit identities for moments. As will be seen shortly, virtually identical interpretations hold for the characterizations of the other classical distributions.

**4.2 The Gamma density and Laguerre polynomials.**

For $\alpha > 0$, the gamma distribution with parameter $\alpha$ has density $\nu_\alpha(x) = e^{-\nu}x^{\alpha-1}/\Gamma(\alpha)$ on $(0, \infty)$. The orthogonal polynomials for this density are called Laguerre polynomials. They have the explicit representation

$$L^{\alpha-1}(x) = \sum_{k=0}^{n} \binom{n+\alpha-1}{n-k} (-x)^k.$$ 

Thus $L_0^{\alpha-1} = 1, L_1^{\alpha-1} = \alpha - x, L_2^{\alpha-1} = \frac{1}{2}(\alpha+1)\alpha - (\alpha + 1)x + \frac{1}{2}x^2$. The identity here is
Lemma 1. Let $L_n$ be defined by (4.4). Then, for $n \geq 1$ and $a > 0$,

$$
\int_0^a L_n^{\alpha-1}(x) \gamma_{\alpha}(x) dx = \frac{x}{n} \gamma_{\alpha+1}(a) L_{n-1}^{\alpha}(a).
$$

Proof: Chihara (1978, pg. 145) gives the Rodrigues type formula

$$
L_n^{\alpha} = \frac{1}{n!} x^{-\alpha} e^x D^n[x^{n+\alpha} e^{-x}].
$$

Remark: For integer values of $\alpha$ the integrals can be evaluated by elementary techniques, even when $n = 0$.

Example 1: The analog of De Moivre's identity is

$$
\frac{1}{\Gamma(\alpha)} \int_0^\infty |x - \alpha x^{\alpha-1} e^{-x} dx = 2\alpha \gamma_{\alpha+1}(\alpha).
$$

Here, as earlier, the mean absolute deviation is twice the variance times the density at its mode.

Example 2 (Stein's Method): The gamma density can be characterized as follows: a random variable $X$ has a $\gamma_{\alpha}$ density if and only if

$$
E(X f'(X)) = E((X - \alpha)f(X))
$$

for every smooth function $f$ of compact support. This can be used to prove limit theorems for exponential and chi-squared variables via analogs of Stein's method. The formalism involves a study of the equation

$$
x f'(x) + (\alpha - x)f(x) = g(x).
$$

This can be solved explicitly as

$$
Ug(x) = x^{-\alpha} e^x \int_0^x g(t)t^{\alpha-1} e^{-t} dt.
$$

Lemma 1 gives

21
COROLLARY 1. The operator $U$ defined in (4.5) is a bounded linear operator from $L^0_0(\gamma_0)$ into $L^2(\gamma_0+1)$. If $\{L^{-1}_n\}_{n=1}^\infty$ and $\{L_1^\infty\}_{n=0}^\infty$ are taken as orthogonal bases of these spaces, then

$$U(L^{-1}_n) = nL^\alpha_{n-1}.$$

4.3 The beta distribution and Jacobi polynomials.

For $\alpha, \beta > 0$, the beta distribution with parameters $\alpha, \beta$ has density

$$\beta(x; \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1} \quad \text{on} \quad [0, 1].$$

The corresponding orthogonal polynomials are called Jacobi polynomials. They are given explicitly as

$$p^{\alpha-1,\beta-1}_n(x) = \sum_{k=0}^{n} \binom{n + \alpha - 1}{k} \binom{n + \beta - 1}{n - k} (x - 1)^k x^{n-k}. \quad (4.6)$$

Well known special cases include the Legendre polynomials (orthogonal polynomials for the uniform distribution on $[0, 1]$) and the Chebyshev polynomials of the first and second kind. The identity here becomes

LEMMA 1. Let $p_n$ be defined by (4.6). For $n \geq 1$,

$$\int_0^a p^{\alpha-1,\beta-1}_n(x)\beta(x; \alpha, \beta)dx = \frac{-\alpha\beta}{n(\alpha + \beta + 1)(\alpha + \beta)} p^{\alpha,\beta}_n(a)\beta(\alpha + 1, \beta + 1).$$

PROOF: Chihara (1978, pg. 143) gives a Rodrigues type formula which may be rewritten as

$$p^{\alpha,\beta}_n(x) = \frac{(-1)^n}{n!}x^{-\alpha}(1-x)^{-\beta} D^n(x^{n+\alpha}(1-x)^{n+\beta}).$$

The result follows after elementary manipulation. \qed

EXAMPLE 1: (The Dirichlet distribution). For $m \geq 1$, the standard $m$-simplex is denoted $\Delta_m = \{x \in \mathbb{R}^m : x_i \geq 0, \ x_1 + \cdots + x_m = 1\}$. The symmetric Dirichlet
distribution on $\Delta_m$ has density $D_k(x_1, \cdots, x_m) = \frac{\Gamma(km)}{\Gamma(k)^m} \prod_{i=1}^{m} x_i^{k-1}$. This has been extensively used as a prior density for Bayesian calculations by I.J. Good.

For $k$ large, $D_k$ converges to a point mass at the center of the simplex $x^* = (\frac{1}{m}, \frac{1}{m}, \cdots, \frac{1}{m})$. The rate of convergence of $D_k$ to $x^*$ can be studied as an application of Lemma 1. If $P$ is a random choice from $D_k$, let $E_k = E_k\|P - x^*\|$ where $\|P - x^*\| = \frac{1}{2} \sum_{i=1}^{m} |P_i - \frac{1}{m}|$ denotes total variation. Thus $E_k$ is a subjectivist measure of the expected distance of a typical pick from $D_k$ to the uniform measure $x^*$.

**COROLLARY 1.** $E_k = \frac{1}{k} \frac{\Gamma(mk)}{\Gamma(k)^m \Gamma(k(m-1))} (\frac{1}{m})^k (1 - \frac{1}{m})^{k(m-1)}$.

The proof follows from linearity using the mean absolute deviation formula for the beta: If $X \in [0, 1]$ has a $\beta(\nu; \alpha, \beta)$ density with mean $\mu = \frac{\alpha}{\alpha + \beta}$ then

$$E_{\alpha, \beta}|X - \mu| = \frac{2\mu(1 - \mu)}{(\alpha + \beta)} \beta(\mu; \alpha, \beta).$$

This in turn follows easily from Lemma 1. □

**REMARKS:** If $k = 1$, $D_k$ becomes the uniform distribution on $\Delta_m$. Then,

$$E_1 = \frac{m-1}{m} (1 - \frac{1}{m})^{m-1} > \frac{1}{e}.$$

Thus a point chosen at random on $\Delta_m$ is not too close to the uniform distribution if $m$ is large.

For $m$ fixed, as $k \to \infty$,

$$E_k \sim \frac{1}{\sqrt{2\pi k}} (1 - \frac{1}{m})^{3/2}.$$

Thus for $k$ large, a typical pick from $D_k$ is close to the center in total variation. Using Markov's inequality gives convergence of $D_k$ to $x^*$ in probability.

**5. The Pearson family of curves.**

The results derived for the normal, gamma, and beta distributions can be generalized to other members of the Pearson family. Moreover, in a sense to
be made precise, Pearson families are the only families of continuous proba-

bility densities for which the particular argument employed works. In this section

we present background, show that the orthogonal polynomials associated to a

Pearson family admit closed form integrals, and prove that this characterizes the

Pearson families.

A) PEARSON CURVES:

In 1895, the English statistician Karl Pearson introduced his famous family

of frequency curves. Pearson generated the elements of his family by considering

the possible solutions to the differential equation

\[ \frac{f'(x)}{f(x)} = \frac{a_0 + a_1 x}{b_0 + b_1 x + b_2 x^2} = \frac{q(x)}{p(x)}. \]

(Strictly speaking, Pearson took \(a_1 = 1\), but it is more natural to include the

coefficient and permit the possibility \(a_1 = 0\).) The Pearson family has a simple

structure. There are in essence five basic solutions, depending on whether the

polynomial \(p(x)\) in the denominator is constant, linear, or quadratic; and in the

latter case on whether the discriminant of \(p(x)\) is positive, negative, or zero.

It is easy to show that the Pearson family is closed under location and scale

change. Thus the study of the family can be reduced to the differential equations

that result after an affine transformation of the independent variate.

If \(\deg p(x) = 0\), then after change of variable the differential equation reduces
to \(f'(x)/f(x) = \pm x\); if \(f(x)\) is assumed to be defined on the maximal interval
possible (here \(-\infty < x < \infty\)), then in order for \(f(x)\) to be integrable, only the
negative sign is permissible and \(f(x)\) is seen to be the standard normal density.
If \(\deg p(x) = 1\), then (up to change of location and scale) the resulting maximal
solutions may similarly be seen to be the family of gamma distributions, this
corresponds to Pearson's Type 3.

If \(\deg p(x) = 2\), then the situation is somewhat more complex.

1. If the discriminant \(\Delta = b_1^2 - 4b_0b_2\) of the polynomial \(p(x) = b_0 + b_1 x + b_2 x^2\) is

   negative, then \(p(x)\) has no real roots, and after an affine change of variable, the

   \(\Delta\) becomes non-negative, the solutions were periodic and the problem
   reduces to the classical case when \(\Delta = 0\).
density \( f(x) \) can be brought into the form

\[
f(x) = C(1 + x^2)^{-\alpha} \exp\{\beta \arctan x\},
\]

where \( C \) is the appropriate normalizing constant. If it is assumed that \( f(x) \) is defined on the maximal possible interval – here \((-\infty, \infty)\), – then \( \alpha > 1/2 \) and \(-\infty < \beta < \infty \) insure that \( f(x) \) is integrable. Except for special values of \( \alpha \) and \( \beta \), this corresponds to Pearson’s Type 4; in particular, the \( t \)-distributions are a (rescaled) subfamily of this class.

2. If the discriminant \( \Delta = b_1^2 - 4b_0b_2 \) is zero, then \( p(x) \) has a single repeated real root, and after an affine change of variable, the density \( f(x) \) can be brought into the form

\[
f(x) = Cx^{-\alpha} \exp\{-\frac{\beta}{x}\}.
\]

Here there are two maximal intervals, \((-\infty, 0)\) and \((0, \infty)\), but by the further change of variable \( y = -x \), every such “maximal” density can be thought of as defined on the positive reals. In this case \( \alpha > 1, \beta \geq 0 \) insure that \( f(x) \) is integrable. Except for special values of \( \alpha \) and \( \beta \), this corresponds to Pearson’s Type 5; in particular, the inverse Gaussian distributions are a (rescaled) subfamily of this class.

3. If the discriminant \( \Delta \) is positive, then \( p(x) \) has two distinct real roots, and after an affine change variable, the density \( f(x) \) can be brought into the form

\[
f(x) = C(1 - x)^\alpha (1 + x)^\beta.
\]

Here there are three maximal intervals, \((-\infty, -1), (-1, 1), \) and \((1, \infty)\), but after a further change of variable these can be taken to be either \((0, 1)\) or \((0, \infty)\). If the maximal interval is \((0, 1)\), then \( \alpha, \beta > -1 \) ensure that \( f(x) \) is integrable; these are the beta densities, and except for special values of \( \alpha \) and \( \beta \), correspond to Pearson’s Type 1 (the asymmetric beta) and Type 2 (the symmetric beta). If the maximal interval is \((0, \infty)\), then \( \alpha > -1, \alpha + \beta < -1 \) ensure that \( f(x) \) is integrable; and, in particular, the \( F \)-distributions are a subfamily of this class.
B) Closed Form Summation:

Let \( f(x) \) be a member of the Pearson family. Let \( \{ P_n \}_{n=0}^N \) be a sequence of orthogonal polynomials for \( f \). We admit here densities such as the \( t \), with only a finite number of moments, so that \( N < \infty \) is permitted. Thus \( \deg P_n = n \) and \( \int P_n P_m f = 0 \) for \( n \neq m \) provided the integral exists. The following regularity condition holds in all examples: The support of \( f \) is an interval \((a, b)\) and

\[
\frac{d^k}{dx^k} [f(x)p(x)^n] \text{ vanishes for } x = a, x = b, 0 \leq k < \alpha_n \leq N.
\]

\[
\frac{1}{f(x)} \frac{d^n}{dx^n} [f(x)p(x)^n] \text{ is a polynomial of degree } n, \quad 1 \leq n \leq N
\]

**Theorem 1.** Let \( \{ P_n \}_{n=0}^N \) be a sequence of orthogonal polynomials from the Pearson family (5.1) which satisfies (5.2). Then there are constants \( \lambda_n \neq 0 \) such that for all \( \alpha \) and \( n \geq 1 \)

\[
\int_{-\infty}^{\alpha} P_n(x)f(x)dx = \frac{1}{\lambda_n} f(\alpha)p(\alpha) \frac{dP_n}{dx}(\alpha).
\]

**Proof:** By an elementary computation, the functions

\[
Q_n(x) = \frac{1}{f(x)} \frac{d^n}{dx^n} [f(x)(p(x))^n], \quad n = 0, 1, \cdots N
\]

are polynomials, where \( p(x) \) is the polynomial in the denominator of the Pearson differential equation (5.1). Chihara (1978, pg. 146-147) shows that any such polynomials are orthogonal with respect to \( f \). It follows that for some positive constants \( c_n \), the following Rodrigues formula holds:

\[
P_n(x) = \frac{1}{c_n f(x)} \frac{d^n}{dx^n} [f(x)(p(x))^n].
\]

From this, it follows as in Chihara (1978, p. 148), that \( y = P_n \) satisfies the differential equation

\[
\frac{d}{dx} [f(x)p(x) \frac{dy}{dx}] - \lambda_n f(x)y, \quad n = 0, 1, \cdots N - 1,
\]
where the $\lambda_n$ are appropriate constants, which are nonzero for $n \geq 1$. This yields the theorem.

The key to the above argument is the Rodrigues formula (5.3). The next argument shows that the only probability densities admitting such a representation are Pearson families. Fix a positive integer $N$. Let $f(x)$ be a $C^N$ probability density, defined and everywhere positive on an open interval $J$. A sequence of polynomials $\{p_n; 0 \leq n \leq N\}$ such that $\deg p_n = n$ is said to satisfy a Rodrigues formula with respect to $f(x)$ if there exists a polynomial $g(x)$ and a sequence of nonzero constants $C_n$ such that

\begin{equation}
 p_n(x) = \frac{1}{C_n} \frac{1}{f} \frac{d^n}{dx^n} [f(x)g(x)^n], \quad 0 \leq n \leq N.
\end{equation}

The following result, which is based on a theorem of Cryer (1970, p. 3), shows that having a Rodrigues formula for $N = 2$ implies that $f$ is a member of the Pearson family.

**Theorem 2.** Let $p_0, p_1, p_2$ be three polynomials such that $\deg p_j = j$ for each $j$, and such that (5.4) is satisfied for $n = 1, 2$. Then $f(x)$ is a member of the Pearson family of probability distributions (5.1).

**Proof:** By assumption there exists a polynomial $g(x)$ and nonzero constants $C_j$, $1 \leq j \leq 2$, such that $C_j p_j f = D^j \{fg^j\}$ for $j = 1, 2$. Thus

\[
 C_2 p_2 f = D^2 \{fg^2\} = D[\{(fg)Dg + gD(fg)\}]
\]

\[
 = fgD^2 g + D(fg)Dg + D[g(C_1 p_1 f)]
\]

\[
 = fgD^2 g + C_1 p_1 fDg + fgD(C_1 p_1) + C_1 p_1 D(fg)
\]

\[
 = fgD^2 g + f(C_1 p_1 Dg + gD(C_1 p_1)) + (C_1 p_1)^2 f
\]

\[
 = f(gD^2 g + D(C_1 p_1 g) + (C_1 p_1)^2);
\]

hence $C_2 p_2 = gD^2 g + D(C_1 p_1 g) + (C_1 p_1)^2$. But if $\deg g = m \geq 3$, then the degree of the right hand side would be $2m - 2 \geq 4$, which is impossible. Thus $\deg g \leq 2$. But $C_1 p_1 f = D(fg) = fDg + gDf$, hence

\[
 \frac{Df}{f} = \frac{C_1 p_1 - Dg}{g}.
\]
It follows that
\[
\frac{f'(x)}{f(x)} = \frac{a_1 x + a_0}{b_2 x^2 + b_1 x + b_0},
\]
for appropriate constants \(a_0, a_1, b_0, b_1,\) and \(b_2;\) and thus, by definition, \(f(x)\) is a member of the Pearson family.

REMARK: De Moivre’s MAD identity was given explicitly in Section 3.3. Stein (1986, Chapter 6) gives appropriate versions of his identity for general densities and explicitly specializes to Pearson curves \(f(x)\). Suppressing regularity conditions, a random variable \(X\) has density \(f(x)\) as at (4.7) if and only if for every smooth \(h\) of compact support
\[
E[Xh(X) - p(X)h'(X)] = 0, \quad \text{with } p(x) \text{ as in (5.1)}.
\]
Stein proves this by introducing an operator \(U\), just as in the normal case. We presume that the orthogonal polynomials give a singular value decomposition for this operator to the extent that this makes sense (e.g., existence of moments).

6. Two Discrete Examples.

6.1 The Poisson distribution and Charlier polynomials.

The two examples of this section are included because the arguments do not require Rodrigues type representations. For \(\lambda > 0\), let \(q_\lambda(j) = e^{-\lambda} \lambda^j / j!\) denote the Poisson density on \(\{0, 1, 2, \cdots\}\). The orthogonal polynomials are called Charlier polynomials. Chihara (1978, pg. 170-72) gives background and details.

A monic form of the polynomials can be given explicitly as
\[
C_n^\lambda(x) = \sum_{k=0}^{n} \binom{n}{k} \frac{x^k}{k!} (-\lambda)^{n-k}.
\]
Then \(C_0 = 1, \ C_1 = x - \lambda, \ C_2 = x(x - 1) - 2\lambda x + \lambda^2\). The identity becomes

LEMMA 1. Let \(C_n\) be defined by 6.1. For \(n \geq 1\), and \(0 \leq a < \infty\) any integer,
\[
\sum_{k=0}^{a} C_n(k)q_\lambda(k) = -\lambda q_\lambda(a)C_{n-1}(a).
\]
PROOF: The polynomials satisfy the recurrence relation

\[ C_{n+1}^\lambda(x) = (x - n - \lambda)C_n^\lambda(x) - \lambda nC_{n-1}^\lambda(x). \]

The Christoffel-Darboux identity for polynomials given by \( P_n - (x - c_n)P_{n-1} - \lambda_n P_{n-2} \), is

\[ \sum_{k=0}^{\infty} \frac{P_k(x)P_k(y)}{\lambda_1 \cdots \lambda_{k+1}} = (\lambda_1 \cdots \lambda_{n+1})^{-1} \frac{P_{n+1}(x)P_n(y) - P_n(x)P_{n+1}(y)}{x - y}. \]

This holds for any \( n, x \) and \( y \). We will specialize this to the Charlier case, take \( y = 0 \), and use the duality relation

\[ C_n(k) = (-\lambda)^{n-k}C_k(n). \]

The left side of the Christoffel-Darboux identity becomes

\[ \sum_{k=0}^{\infty} \frac{C_k(x)C_k(0)}{\lambda_1 \cdots \lambda_{k+1}} = \sum_{k=0}^{n} \frac{(-\lambda)^{k-x}C_z(k)(-\lambda)^k}{\lambda^k k!} = (-\lambda)^{-x} \sum_{k=0}^{n} \frac{C_z(k)\lambda^k}{k!}. \]

The right side is easily seen to be

\[ \frac{-\lambda^{n+1}(-\lambda)^{-x}}{n!}C_{n-1}(n) \]

where the identities \( \Delta C_n(x) = nC_{n-1}(x) \) (Chihara 1978, pg. 171) were useful.

Equating the two sides then gives the stated result.

\[ \square \]

REMARK: There is a Rodrigues type formula involving finite differences which is available:

\[ C_n(x) = \lambda^{-x}\Gamma(x + 1)\Delta^n \left[ \frac{x^n}{\Gamma(x - n + 1)} \right] \quad \text{with} \quad \Delta f(x) = f(x + 1) - f(x). \]

Here \( x \) is treated as a variable and one easily sees this gives a polynomial. Direct use of the formula for integer \( x \) smaller than \( n \) requires care in its interpretation.

EXAMPLE 1: The analogue of De Moivre’s identity here becomes

\[ \sum_{k=0}^{\infty} |k - \lambda| \frac{e^{-\lambda} \lambda^k}{k!} = 2\lambda \frac{e^{-\lambda} \lambda^{|\lambda|}}{[\lambda]!}. \]
with $[\lambda]$ the largest integer less than or equal to $\lambda$.

Billingsley (1986, pp. 381-382) bases a proof of Stirling's formula on this identity. Similar proofs can be based on the other identities of this section.

**Example 2 (Stein's Identity):** The Poisson distribution is characterized by the identity

$$E[\lambda f(X + 1)] = E[Xf(X)]$$

for every bounded function $f$ on the integers. Solving for $f$ in terms of $g$ in

$$\lambda f(x + 1) = xf(x) = g(x)$$

leads to

$$f(x) = (Ug)(x) = \frac{1}{\lambda q(x - 1)} \sum_{j=0}^{x-1} g(j) q(\lambda j), \quad x \geq 1; f(0) = 0.$$  \hspace{1cm} (6.2)

As usual, $\sum_{x=0}^{\infty} g(x) q(\lambda x) = 0$ is assumed. Stein (1986, Chap. 9) gives background and motivation.

Comparison with Lemma 1 shows that the Charlier polynomials give a singular value decomposition for $U$. To state this explicitly, we use $\sum_{\ell=0}^{\infty} C_{i}(\ell) C_{j}(\ell) q(\lambda \ell) = \lambda^{i} \delta_{ij}$ to form orthonormal polynomials $\tilde{C}_{n} = C_{n} \sqrt{\lambda^{n} n!}$. Let $\{\tilde{C}_{j}\}_{j=1}^{\infty}$ be an orthonormal basis for $L_{0}^{2}(q_{\lambda})$. Define $p_{\lambda}(j) = q_{\lambda}(j - 1)$. Let $L^{2}(p_{\lambda}) = \{f : \{1, 2, \ldots\} \rightarrow \mathbb{R} : \sum_{j=1}^{\infty} f^{2}(j) p_{\lambda}(j) < \infty\}$. Thus $\tilde{D}_{n}(j) = \tilde{C}_{n}(j - 1)$ form an orthonormal basis for $L^{2}(p_{\lambda})$, $0 \leq n \leq \infty$.

**Corollary 1.** The operator $U$ defined by (6.2) is a bounded linear operator from $L_{0}^{2}(q_{\lambda})$ to $L^{2}(p_{\lambda})$. If these spaces are given bases $\{\tilde{C}_{n}\}_{n=1}^{\infty}$ and $\{\tilde{D}_{n}\}_{n=0}^{\infty}$, then $U$ has singular values given by

$$U(\tilde{C}_{n}) = \frac{-1}{\sqrt{\lambda n}} \tilde{D}_{n-1}.$$  

6.2 Binomial distribution and Krawtchouk polynomials.
Let \( b(k; n, p) = \binom{n}{k} p^k (1-p)^{n-k} \) denote the binomial density on \( \{0, 1, \cdots, n\} \).

The Krawtchouck polynomials are orthogonal for \( b(k; n, p) \). They can be given by the explicit formula

\[
P_k(x) = \sum_{j=0}^{k} (-1)^j \gamma^{k-j} \binom{x}{j} \binom{n-x}{k-j}; \quad 0 \leq k \leq n; \quad \gamma = p/(1-p).
\]

Thus \( P_0 = 1, P_1 = \gamma(n-x)-x, P_2 = \gamma^2 \frac{(n-x)(n-x-1)}{2} - \gamma x(n-x) + \frac{x(x-1)}{2} \).

The orthogonality relation is

\[
\sum_{i=0}^{n} b(i; n, p) P_r(i) P_s(i) = \left( \frac{p}{1-p} \right)^r \binom{n}{r} \delta_{rs}.
\]

Basic properties of Krawtchouck polynomials with extensive references are given by Macwilliams and Sloane (1977, pg. 150-153).

The basic identity for the binomial density is

**Lemma 1.** For \( P^*_k \) defined by (6.3), with \( k \geq 1 \) and \( a \) an integer, \( 0 \leq a \leq n \),

\[
\sum_{i=0}^{a} P^*_k(i) b(i; n, p) = \frac{p}{1-p} \frac{n-a}{k} b(a, n, p) P^{*n-1}_{k-1}(a).
\]

**Proof:** Macwilliams and Sloane (1977, p. 152) give the identity

\[
P^*_0(x) + P^*_1(x) + \cdots + P^*_a(x) = P^{*n-1}_a(x-1),
\]

and the duality relation

\[
P_k(i) = \frac{\gamma^{k-i} \binom{n}{k}}{\binom{i}{i}} P_i(k).
\]

Substituting (6.5) into (6.4) and simplifying gives the desired result.

**Remarks:** The Krawtchouck polynomials have a Rodrigues type representation which can be used to give an alternative proof for Lemma 1. The Christoffel-Darboux formula and duality can also be used as in our treatment of the Poisson
distribution. Finally, as was our case originally, the correct formula can be guessed at from small cases and proved directly from (6.3).

**EXAMPLE 1 (STEIN’S IDENTITY):** Binomial random variables are characterized by the identity

\[ E(p(n - X)f(X)) = E(qX f(X - 1)) \]

for every \( f : \{-1, 0, 1, \cdots n\} \) into \( \mathbb{R} \), where \( q = 1 - p \). The study of this identity involves solving for \( f \) given \( g \) in the following equation

\[ p(n - x)f(x) - qx f(x - 1) = g(x) \quad \text{where} \quad E_{n,p}(g(X)) = 0. \]

This can be solved explicitly as

\[ f(x) = (U g)(x) = \frac{1}{p(n - x) b(x, n, p)} \sum_{i=0}^{x} b(i, n, p) g(i), \quad 0 \leq x < n - 1. \]

The value of \( f(-1) \) and \( f(n) \) can be chosen arbitrarily.

Lemma 1 translates into a singular value decomposition for \( U \) after introducing orthonormal bases \( \hat{P}_r^n(x) = P_r^n / \sqrt{\binom{n}{r}(p/q)^r} \).

**COROLLARY.** For \( U \) defined by (6.6). Let \( L_0^n(b(k; n, p)) \) and \( L_2^n(b(k; n - 1, p)) \) have \( \{\hat{P}_i^n\}_{i=1}^n \) and \( \{\hat{P}_{i-1}^{n-1}\}_{i=0}^{n-1} \) as orthonormal bases. Then

\[ U(\hat{P}_i^n) = P_{i-1}^{n-1} / \sqrt{inp/q} \]

shows \( U \) is a 1-1, onto, linear map with singular values \( 1/\sqrt{inp/q} \), \( 1 \leq i \leq n \).

### 6.3 Other discrete densities.

Very similar results can be derived for other discrete densities. What is needed is either a Rodrigues type formula or a duality result along with the Christoffel-Darboux identity as outlined in Sections 6.1-6.2 above.

For example, the geometric and negative binomial distributions give rise to Mexner polynomials, the hypergeometric distribution to Hahn polynomials.
(There is, in fact, a discrete analog of Theorem 2 of Section 5 characterizing all discrete measures having Rodrigues type formulae.) Along these lines see Chihara (1978, Chapter 5, Section 3). Eagleson (1968) characterizes discrete orthogonal polynomials which admit a duality relation.

7. Coda.

This paper had its origin in the simple observation that buried in Problem 72 of De Moivre's *Doctrine of Chances* was the $L_1$ law of large numbers for Bernoulli trials. Somewhat to our surprise, however, what was initially regarded as a fairly straightforward (and short!) historical note soon began to acquire a life of its own: no sooner did we think that we had tracked down the earliest rediscovery of the result, then another cropped up; a routine intellectual credit check on Sir Alexander Cuming ended up leading us down the path of an 18th century con artist (see the Appendix below); and an attempt to understand Todhunter's proof of De Moivre's formula ultimately resulted in the discovery of a much more general phenomenon, valid for many of the classical distributions.

Most of us have probably had this experience at one time or another. But (for us, at least) it seems to happen with uncanny frequency when trying to read and understand the past masters of our subject, which is one reason why we enjoy doing it so much.
APPENDIX: SIR ALEXANDER CUMING

In the *Miscellanea analytica*, De Moivre states that Problem 72 in the *Doctrine of Chances* had been originally posed to him in 1721 by Alexander Cuming, whom he describes as an illustrious man (*vir clarissimus*) and a member of the Royal Society (*Cum aliquando labente Anno 1721, Vir clarissimus Alex. Cuming Eq. Au. Regiae Societatis Socius, quaestionem infra subjectam mihi proposuisse, solutionem problematis ei postero die tradideram*). At almost the same time, James Stirling acknowledged Cuming’s stimulus in the introduction to his *Methodus Differentialis* (Stirling, 1730). Thus Alexander Cuming appears to have played, for De Moivre and Stirling, a role similar to that of the Chevalier de Méré for Pascal and Fermat. Who was he?

At this remove of time, the question can only be partially answered, but the story that emerges is a strange and curious one, a wholly unexpected coda to an otherwise straightforward episode in the history of mathematics.

The British *Dictionary of National Biography* tells us that Cuming was a Scottish baronet, born about 1690, who briefly served in the Scottish bar (from 1714 to 1718) and then left it, under obscure but possibly disreputable circumstances. Shortly after, Cuming surfaced in London, where he was elected a fellow of the Royal Society of London on 30 June 1720, the year before that in which De Moivre says Cuming posed his problem. The *DNB* does not indicate the reason for Cuming’s election, and there is little if any indication of serious scientific output on his part. (No papers by him appear, for example, in the *Philosophical Transactions of the Royal Society of London*. This was not unusual, however, at the time; prior to a 19th century reform, members of the aristocracy could become members of the Royal Society simply by paying an annual fee.)

During the next decade Cuming seems to have taken on the role of intellectual go-between, and he brought the problem of the evaluation of binomial coefficients to Stirling’s attention when the latter returned to England in 1725 after nearly a decade abroad. Cuming’s chief claim to fame, however, lies in an entirely different
direction. In 1729 he undertook an expedition to the Cherokee Mountains in Georgia, several years prior to the time the first settlers went there, led by James Oglethorp, in 1734. Appointed a chief by the Cherokees, Cuming returned with seven of their number to England, presenting them to King George II in an audience at Windsor Castle on 18 June 1730. Before returning, an “Agreement of Peace and Friendship” was drawn up by Cuming and signed by the chiefs, which agreement, as the 19th century *DNB* so charmingly puts it, “was the means of keeping the Cherokees our firm allies in our subsequent wars with the French and American colonists”.

This was Sir Alexander’s status in 1730, when De Moivre refers to him as an illustrious man, and a member of the Royal Society; both conditions, unfortunately, were purely temporary. For the surprising *denouement* to Sir Alexander’s career, we quote the narrative of the *DNB*:

By this time some reports seriously affecting Cuming’s character had reached England. In a letter from South Carolina, bearing date 12 June 1730, … he is directly accused of having defrauded the settlers of large sums of money and other property by means of fictitious promissory notes. He does not seem to have made any answer to these charges, which, if true, would explain his subsequent ill-success and poverty. The government turned a deaf ear to all his proposals, which included schemes for paying off eighty millions of the national debt by settling three million Jewish families in the Cherokee mountains to cultivate the land, and for relieving our American colonies from taxation by establishing numerous banks and a local currency. Being now deeply in debt, he turned to alchemy, and attempted experiments on the transmutation of metals.

Fantastic as Cuming’s alleged schemes might seem, they were of a type not new to the governments of his day. A decade earlier, thousands had lost fortunes in England and France with the bursting of the South Sea and Mississippi “bubbles”. In England, the South Sea Company, an initially legitimate enterprise, like
the better known East India Company, had been formed in 1711, with a monopoly of trade in the South Seas and South America. At first a sound enterprise, in 1719 it undertook to assume the English national debt in return for stock in the Company. The price of stock soared dramatically, but in 1720, in the first great stock-market panic, prices plummeted and thousands were ruined. At almost the same time in France, the Scotsman John Law was responsible for the so-called "Mississippi bubble". (Unlike the South Sea bubble and its manipulators, there is no evidence of blatant fraud on the part of Law himself, merely monumental bad judgment, but the consequences were no less disastrous for the thousands entangled in its web.)

For Cuming it was all downhill from here. A few years later, in 1737, the law finally caught up with him, and he was confined to Fleet prison, remaining there perhaps continuously until 1766, when he was moved to the Charterhouse (a hospital for the poor), where he remained until his death on 23 August 1775. He had been expelled from the Royal Society on 9 June 1757 for nonpayment of the annual fee, and when his son, also named Alexander, died some time prior to 1796, the Cuming baronetcy became extinct. By 1738, when the second edition of De Moivre's *Doctrine of Chances* appeared, association with the Cuming name had clearly become an embarrassment, and unlike the corresponding passage in the *Miscellanea Analytica*, no mention of Cuming appears when De Moivre discusses the problem Cuming had posed to him.

Thus Cuming's life in outline. Nevertheless, there remain tantalizing and unanswered questions. The account in the *Dictionary of National Biography* appears largely based on an article by H. Barr Tomkins (1878). Tomkins's article several times quotes a manuscript written by Cuming while in prison (see also Drake, 1872), and this manuscript is presumably the ultimate source for the curious schemes mentioned by the *DNB*. But although they are there presented as serious proposals, at the time that Cuming wrote the manuscript his mind had been substantially deranged for several years, and is evidentiary value is questionable.
Acknowledgement. We thank Richard Askey, Ian Johnstone, Charles Stein, Steve Stigler, and Gérard Letac for their comments as our work progressed.
REFERENCES


Alexander Cuming to the Country of the Cherokees, in the Year 1730. David Clapp and Son, Boston.


