On Expanding the Sine Function with Catalan Numbers: A Note on a Role for Hypergeometric Functions

Peter J. Larcombe and David R. French (Retired)

School of Mathematics and Computing
University of Derby, Kedleston Road, Derby DE22 1GB, U.K.
{P.J.Larcombe@derby.ac.uk}

Abstract

The theory of hypergeometric functions is brought to bear on a problem—namely, that of obtaining a certain power series expansion involving the sine function that is inclusive of the Catalan sequence and which serves as a prelude to the calculation of other related series of similar type. A general formulation provides the particular result of interest as a special case, into which Catalan numbers are introduced as desired.

Introduction

Based on a 1988 paper by Luo [1], Larcombe discussed the rather surprising role of the Catalan sequence $\{c_0, c_1, c_2, c_3, c_4, \ldots\} = \{1, 1, 2, 5, 14, \ldots\}$, defined by

$$c_n = \frac{1}{n+1}\begin{pmatrix} 2n \\ n \end{pmatrix}, \qquad n = 0, 1, 2, \ldots, \tag{1}$$

in the infinite series representation (in odd powers of $\sin(\alpha)$) of the trigonometric function $\sin(m\alpha)$ for integer m even [2]. It was observed that, beginning with an appropriate expansion of $\sin(2\alpha)$, those for subsequent functions $\sin(4\alpha), \sin(6\alpha), \sin(8\alpha)$, *etc.*, could in turn be generated via the recursion (valid for $m \geq 2$ and simple to establish)

$$\sin[(m+2)\alpha] = 2\left\{[1 - 2\sin^2(\alpha)]\sin(m\alpha) - (1/2)\sin[(m-2)\alpha]\right\}. \qquad (2)$$

From certain cases examined it was shown how the series for $\sin(m\alpha)$ starts explicitly with the first $\frac{1}{2}m$ odd powers of $\sin(\alpha)$, followed by higher order terms each possessing a coefficient dependent upon a specific linear function of $\frac{1}{2}m$ Catalan numbers. The sequential nature of building up expansions was explained in detail by example, with the result initiating the whole process being

$$\sin(2\alpha) = 2\left\{\sin(\alpha) - \sum_{n=1}^{\infty}\left[\frac{c_{n-1}}{2^{2n-1}}\right]\sin^{2n+1}(\alpha)\right\} \qquad (3)$$

(converging for $|\alpha| < \pi/2$) in which the appearance of the Catalan sequence is critical. This additional communication presents a means of arriving at (3) other than by the short proof detailed in [2].[1] It is found through a particular case of a pertinent general expansion of the sine function, whose formulation as a series in hypergeometric form has very much a first principles flavour to it and is felt to be both mathematically instructive and enlightening. In line with this, remarks have been made within the text where deemed appropriate and, in the context of the work, some further comments on the amenable nature of hypergeometric functions in developing other series forms comprise a final section before the Summary.

General Formulation

As indicated, we make use of hypergeometric function theory, writing

$$F\left(\begin{matrix}a_1, a_2\\ b_1\end{matrix}\,\middle|\, z\right) = \sum_{n=0}^{\infty}\frac{(a_1)_n(a_2)_n}{(b_1)_n}\cdot\frac{z^n}{n!} \qquad (4)$$

to denote the hypergeometric function of Gaussian type (with 2 upper parameters a_1, a_2, 1 lower parameter b_1 and argument z—all possibly complex variables),[2] where $(u)_n$ is the rising factorial function

$$(u)_n = u(u+1)(u+2)(u+3)\cdots(u+n-1) \qquad (5)$$

defined for integer $n \geq 0$ ($(u)_0 = 1$); we do not adopt the popular notation ${}_2F_1$, in favour of F, for the simple reason that the only hypergeometric

[1]Note that the shifted sequence $\{c_1, c_2, c_3, c_4, c_5, \ldots\} = \{1, 1, 2, 5, 14, \ldots\}$ was employed in [2] for consistency with [1].

[2]For typographical convenience we will also sometimes write $F(a_1, a_2; b_1|z)$.

function dealt with here is of '2-1' (Gaussian) type (excepting Remark 5).

Consider the classic equation

$$\frac{d^2y}{dz^2} + n^2 y = 0 \tag{6}$$

in applied mathematics. It describes, in differential form, simple harmonic motion (of frequency n, assumed non-zero) with general solution

$$y(z) = A\sin(nz) + B\cos(nz) \tag{7}$$

for arbitrary constants A, B. Given any functional change in the independent variable z by a substitution $x = x(z)$, say, it is easy to show that

$$\frac{d^2}{dz^2} = \frac{d^2x}{dz^2}\frac{d}{dx} + \left(\frac{dx}{dz}\right)^2 \frac{d^2}{dx^2}, \tag{8}$$

and using the particular relation $x(z) = \sin^2(z)$ (6) is transformed to

$$x(1-x)\frac{d^2y}{dx^2} + \left(\frac{1}{2} - x\right)\frac{dy}{dx} + \left(\frac{n}{2}\right)^2 y = 0 \tag{9}$$

in $y(x)$. This is but a special case of the well known differential equation (termed the generalised hypergeometric equation)

$$x(1-x)\frac{d^2y}{dx^2} + [\gamma - (1+\alpha+\beta)x]\frac{dy}{dx} - (\alpha\beta)y = 0, \tag{10}$$

where $\alpha = \pm\frac{n}{2}$, $\beta = \mp\frac{n}{2}$ and $\gamma = \frac{1}{2}$ (invariance of (10) under the interchanging of α, β is obvious); we choose to take $\alpha = -\frac{n}{2}$, $\beta = \frac{n}{2}$.

It is known that the hypergeometric equation (10) has regular singular points at $x = 0, 1$ and ∞. Concentrating on the singularity at $x = 0$, then following the (standard) method of Fröbenius a trial solution of the form

$$y(x) = x^c \sum_{r=0}^{\infty} a_r x^r \tag{11}$$

yields roots $c = 0, 1-\gamma$ of the associated (quadratic) indicial equation. If γ is not a negative integer the exponent $c = 0$ gives rise to

$$F\left(\begin{matrix}\alpha, \beta \\ \gamma\end{matrix} \middle| \, x\right) \tag{12}$$

as a solution of (10), whilst for γ not a positive integer the exponent $c = 1-\gamma$ generates a further solution

$$x^{1-\gamma} F\left(\begin{matrix}\alpha+1-\gamma, \beta+1-\gamma \\ 2-\gamma\end{matrix} \middle| x\right). \tag{13}$$

Close to the system origin, therefore, the complete solution of the generalised hypergeometric equation is, subject to the stated restrictions on γ, the general linear combination of (12),(13)

$$y(x) = CF\left(\begin{matrix}\alpha, \beta \\ \gamma\end{matrix} \middle| x\right) + Dx^{1-\gamma} F\left(\begin{matrix}\alpha+1-\gamma, \beta+1-\gamma \\ 2-\gamma\end{matrix} \middle| x\right), \tag{14}$$

say, which converges for $|x| < 1$. For arbitrary C, D it thus follows that the general solution to (6) around $z = 0$ is, in hypergeometric form,

$$\begin{aligned} y(z) &= CF\left(\begin{matrix}-\frac{n}{2}, \frac{n}{2} \\ \frac{1}{2}\end{matrix} \middle| \sin^2(z)\right) \\ &\quad + D\sin(z) F\left(\begin{matrix}\frac{1}{2}-\frac{n}{2}, \frac{1}{2}+\frac{n}{2} \\ \frac{3}{2}\end{matrix} \middle| \sin^2(z)\right), \end{aligned} \tag{15}$$

which converges for $|z| < \pi/2$. If we now impose the condition $y(0) = 0$ then (7) simplifies trivially to

$$y(z) = A\sin(nz), \tag{16}$$

as does (15) to

$$y(z) = D\sin(z) F\left(\begin{matrix}\frac{1}{2}-\frac{n}{2}, \frac{1}{2}+\frac{n}{2} \\ \frac{3}{2}\end{matrix} \middle| \sin^2(z)\right). \tag{17}$$

Furthermore, matching these two equivalent solutions in dy/dz at $z = 0$, it is found that $D = An$ upon which (16),(17) immediately give

$$\sin(nz) = n\sin(z) F\left(\begin{matrix}\frac{1}{2}-\frac{n}{2}, \frac{1}{2}+\frac{n}{2} \\ \frac{3}{2}\end{matrix} \middle| \sin^2(z)\right), \qquad |z| < \pi/2. \tag{18}$$

Remark 1 Convergence of the hypergeometric here over the open interval $z \in (-\frac{\pi}{2}, \frac{\pi}{2})$ is established as a direct consequence of the methodology used, and we can examine its convergence properties at the endpoints to check (18) in two combined special cases. For $z = \pm\frac{\pi}{2}$ the argument of the hypergeometric $F(\frac{1}{2}-\frac{n}{2}, \frac{1}{2}+\frac{n}{2}; \frac{3}{2}|\sin^2(z))$ is 1, and under such circumstances any Gaussian function (4) converges if and only if $\mathrm{Re}\{b_1 - (a_1 + a_2)\} > 0$

(see, for example, [3-5]). Moreover, we can deduce that the hypergeometric in (18) converges to the same value at both interval extremes. Gauss' Theorem gives this as $\Gamma(\frac{3}{2})\Gamma(\frac{1}{2})/[\Gamma(1-\frac{n}{2})\Gamma(1+\frac{n}{2})]$, which simplifies to $\pi/[n\Gamma(\frac{n}{2})\Gamma(1-\frac{n}{2})]$ using the result $s\Gamma(s)=\Gamma(1+s)$ $(s\neq 0)$ together with $\Gamma(\frac{1}{2})=\sqrt{\pi}$, whereupon (18) reads $\sin(\pm\frac{n\pi}{2})=\pm\pi/[\Gamma(\frac{n}{2})\Gamma(1-\frac{n}{2})]$. This is a self-consistent equation, which confirms the parity of the sine function at $\frac{n\pi}{2}$, since it is known that $\Gamma(s)\Gamma(1-s)=\pi/\sin(s\pi)$ (for s non-zero and non-integer).

Remark 2 Whilst Bailey [3] does not include such a result as discussed here, Slater [4] (with no working shown) has an incorrect version of (18) merely listed as equation (1.5.6) on p.17. These are longstanding treatises dedicated to hypergeometric function theory which are still consulted today.

Remark 3 There is, of course, another result which emerges readily from the approach taken to the formulation of (18), being the natural complement to it. It is found that by first applying the condition $y'(0)=0$ to the two general solutions (7),(15), and then matching them at $y(0)$, the following relation is obtained (the reader is invited to check this as an exercise): $\cos(nz)=F(-\frac{n}{2},\frac{n}{2};\frac{1}{2}|\sin^2(z))$ $(|z|<\pi/2)$. For the purposes of this paper, however, it lies outside our remit of interest.

Special Case

It remains only to show that the above formulation of $\sin(nz)$ as a hypergeometric series leads directly to (3) (for completeness, its agreement with a formula attributed to Euler is demonstrated in the Appendix). Changing z to α, then for $n=2$ (18) becomes explicitly

$$\begin{aligned}\sin(2\alpha) &= 2\sin(\alpha)F\left(\begin{matrix}-\frac{1}{2},\frac{3}{2}\\ \frac{3}{2}\end{matrix}\middle|\sin^2(\alpha)\right)\\ &= 2\sum_{n=0}^{\infty}\frac{(-\frac{1}{2})_n}{n!}\cdot\sin^{2n+1}(\alpha)\\ &= 2\left\{\sin(\alpha)+\sum_{n=1}^{\infty}\frac{(-\frac{1}{2})_n}{n!}\cdot\sin^{2n+1}(\alpha)\right\}. \qquad (19)\end{aligned}$$

Consider, now, for $n>1$,

$$\begin{aligned}\left(-\frac{1}{2}\right)_n &= \left(-\frac{1}{2}\right)\left(-\frac{1}{2}+1\right)\left(-\frac{1}{2}+2\right)\cdots\\ &\quad\cdots\left(-\frac{1}{2}+n-2\right)\left(-\frac{1}{2}+n-1\right)\end{aligned}$$

$$= -\left(\frac{1}{2}\right)^n 1\cdot 3\cdot 5\cdot \;\cdots\; \cdot(2n-5)\cdot(2n-3). \qquad (20)$$

After some elementary algebraic manipulation, the product of odd numbers can be written as a function of the nth Catalan term c_{n-1} as

$$1\cdot 3\cdot 5\cdot \;\cdots\; \cdot(2n-5)\cdot(2n-3) = \frac{n!}{2^{n-1}}c_{n-1}, \qquad n > 1, \qquad (21)$$

so that

$$\frac{(-\frac{1}{2})_n}{n!} = -\frac{1}{2^{2n-1}}c_{n-1}, \qquad n > 1. \qquad (22)$$

Since this also holds for $n = 1$ (for which both sides are $-\frac{1}{2}$), (19) yields the required result (3).

Remark 4 We mention here that a quicker route to (3)—which is perhaps more obvious—is offered by taking advantage of the known identity $(1-z)^{-\alpha} = F(\alpha,\beta;\beta|z)$ (β arbitrary, being either non-integer or an integer > 0), for then $(1-z)^{1/2} = F(-\frac{1}{2},\beta;\beta|z)$ which yields $1 - \sum_{n=1}^{\infty}[c_{n-1}/2^{2n-1}]\sin^{2n}(\alpha) = \cos(\alpha)$ (convergent for $|\alpha| < \pi/2$) on setting $z = \sin^2(\alpha)$ and using (22) (valid for $n \geq 1$ as just observed). This, however, amounts only to a reproduction of the result used to prove (3) in the previous paper of Larcombe [2].

Further Remarks

The usefulness of hypergeometric functions in a variety of mathematical settings is well known, and as a matter of interest we can illustrate this by simply deriving a different expanded form of $\sin(nz)$ which consists of odd powers of $\cos(\frac{1}{2}z)$, rather than $\sin(z)$, by first considering the identity

$$F\left(\begin{matrix}\alpha,\beta\\ \alpha+\beta+\frac{1}{2}\end{matrix}\,\middle|\,4\zeta(1-\zeta)\right) = F\left(\begin{matrix}2\alpha,2\beta\\ \alpha+\beta+\frac{1}{2}\end{matrix}\,\middle|\,\zeta\right) \qquad (23)$$

given in Bailey [3] as the quadratic transformation of Gauss. Putting aside the issue of convergence, for $\alpha = \frac{1}{2}-\frac{n}{2}$, $\beta = \frac{1}{2}+\frac{n}{2}$ this leads to

$$\begin{aligned} F\left(\begin{matrix}\frac{1}{2}-\frac{n}{2},\frac{1}{2}+\frac{n}{2}\\ \frac{3}{2}\end{matrix}\,\middle|\,\sin^2(z)\right) &= F\left(\begin{matrix}1-n,1+n\\ \frac{3}{2}\end{matrix}\,\middle|\,\frac{1}{2}[1\pm\cos(z)]\right) \\ &= F\left(\begin{matrix}1-n,1+n\\ \frac{3}{2}\end{matrix}\,\middle|\,\sin^2(z/2)\right), \qquad (24)\end{aligned}$$

choosing to take the negative sign. Thus, (18) now reads

$$\sin(nz) = n\sin(z)F\left(\begin{matrix}1-n, 1+n\\ \frac{3}{2}\end{matrix}\;\middle|\; \sin^2(z/2)\right), \tag{25}$$

and since Euler's identity [3] gives

$$\begin{aligned}[1-\sin^2(z/2)]^{\frac{1}{2}}F\left(\begin{matrix}1-n, 1+n\\ \frac{3}{2}\end{matrix}\;\middle|\; \sin^2(z/2)\right) \\ = \quad F\left(\begin{matrix}\frac{1}{2}+n, \frac{1}{2}-n\\ \frac{3}{2}\end{matrix}\;\middle|\; \sin^2(z/2)\right),\end{aligned} \tag{26}$$

then writing $\sin(z) = 2\sin(z/2)\cos(z/2) = 2\sin(z/2)[1-\sin^2(z/2)]^{\frac{1}{2}}$ equation (25) becomes

$$\sin(nz) = 2n\sin(z/2)F\left(\begin{matrix}\frac{1}{2}+n, \frac{1}{2}-n\\ \frac{3}{2}\end{matrix}\;\middle|\; \sin^2(z/2)\right). \tag{27}$$

The constant n is arbitrary so that the equivalence of this result to (18) is clear, validating the above procedure. We can, however, equally make the same line of argument—beginning by taking the positive sign in the first line of (24)—to yield

$$\sin(nz) = 2n\cos(z/2)F\left(\begin{matrix}\frac{1}{2}+n, \frac{1}{2}-n\\ \frac{3}{2}\end{matrix}\;\middle|\; \cos^2(z/2)\right); \tag{28}$$

as a partial check on this we note that for $n = 1$ (28) reduces to $\sin(z/2) = F(\frac{3}{2}, -\frac{1}{2}; \frac{3}{2}|\cos^2(z/2))$, which is just the standard hypergeometric representation $F(-\frac{1}{2}, \beta; \beta|z)$ of $(1-z)^{1/2}$ (see Remark 4 earlier) with $\beta = \frac{3}{2}$ and z replaced by $\cos^2(z/2)$.

Remark 5 It goes without saying that other established results involving hypergeometric forms can be applied to obtain series for $\sin(nz)$, although some avenues of thought prove not to be useful when pursued. Using the transformation $F^2(\alpha, \beta; \alpha+\beta+\frac{1}{2}|z) = G(2\alpha, 2\beta, \alpha+\beta; 2(\alpha+\beta), \alpha+\beta+\frac{1}{2}|z)$ first published by Clausen in 1828 (where G is a '3-2' hypergeometric function, see [3]), then squaring up (18), for instance, gives $\sin^2(nz) = n^2\sin^2(z)F^2(\frac{1}{2}-\frac{n}{2}, \frac{1}{2}+\frac{n}{2}; \frac{3}{2}|\sin^2(z)) = n^2\sin^2(z)G(1-n, 1+n, 1; 2, \frac{3}{2}|\sin^2(z))$, where for integer $n \geq 1$ the r.h.s. is a finite series. Writing now $\sin(nz) = n\sin(z)\sqrt{G}$, then for each n odd the appropriate known finite expansion, in odd powers of $\sin(z)$, is recovered, whilst for n even the only way to obtain an infinite series in such odd powers is by expanding

$\sqrt{G}$ binomially (the reader may care to check the first few simple cases $n = 1, 2, 3, 4, \ldots$, as the authors have); the point here is that no new results emerge in either case.

Remark 6 It is worth reminding readers that a transformation between hypergeometric forms does sometimes carry with it validity conditions. As an example, according to Bailey [3] (23) is valid inside that (left hand) loop of the lemniscate $|4\zeta(1-\zeta)| = 1$ which surrounds the origin of the complex plane (the general equation of a lemniscate is $|\zeta - k_1||\zeta - k_2| = k$ for real $k > 0$ and k_1, k_2 complex constants). Writing $\zeta = x + iy$, it is straightforward algebraically to show that the equation of the full (two loop) region is, in Cartesian co-ordinates, $(x^2 + y^2)[(x-1)^2 + y^2] = \frac{1}{16}$. This describes a lemniscate—in this case a 'horizontal' figure-of-eight (see Figure 1 below)—centred at $(\frac{1}{2}, 0)$ and extending across the x axis from $(\frac{1}{2}(1-\sqrt{2}), 0)$ to $(\frac{1}{2}(1+\sqrt{2}), 0)$ with a respective maximum and minimum value of $\pm\frac{1}{4}$ for $x = \frac{1}{4}(2 \pm \sqrt{3})$ (the four extremal points corresponding to the intersections of the said lemniscate and the circle $(x - \frac{1}{2})^2 + y^2 = \frac{1}{4}$ with co-incident centre and radius $\frac{1}{2}$). For ζ real the region of validity becomes the open interval $(\frac{1}{2}(1-\sqrt{2}), \frac{1}{2})$, so that in (24) $\sin^2(z/2) \in [0, \frac{1}{2})$. In other words, $\sin(z/2) \in (-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$, which means (24) only holds for $|z| < \pi/2$.

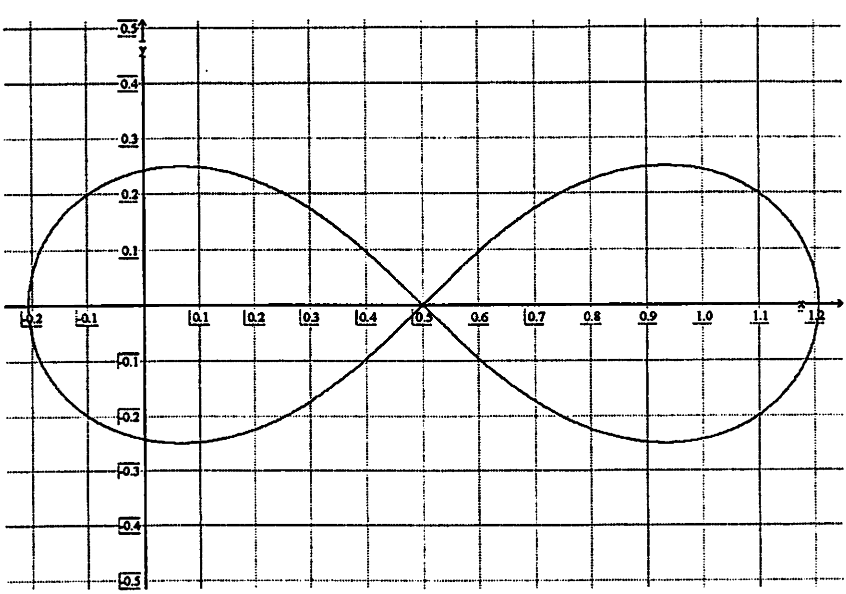

Figure 1: The Lemniscate $|4\zeta(1-\zeta)| = 1$.

Summary

The introduction of Catalan numbers into the series form of $\sin(2\alpha)$ in odd powers of $\sin(\alpha)$ is a necessary requisite for their retention in subsequent expansions (with the same structure) for $\sin(4\alpha), \sin(6\alpha), \sin(8\alpha)$, and so on, which have been discussed at length in [2] in proper historical context. In adding to the earlier publication, we have shown here how this can be achieved using a special case of a general expansion of the sine function which is expressed in hypergeometric form and which does not naturally contain elements of the Catalan sequence. Regarding the latter, the formulation employed makes an appeal to some fundamental hypergeometric function theory and is an illuminating one.

Appendix

Replacing n with m, and likewise z with α, (18) reads (for $|\alpha| < \pi/2$)

$$\begin{aligned}\sin(m\alpha) &= m\sin(\alpha) F\left(\left. \begin{matrix} \frac{1}{2} - \frac{m}{2}, \frac{1}{2} + \frac{m}{2} \\ \frac{3}{2} \end{matrix} \right| \sin^2(\alpha) \right) \\ &= m \sum_{n=0}^{\infty} \frac{(\frac{1}{2} - \frac{m}{2})_n (\frac{1}{2} + \frac{m}{2})_n}{(\frac{3}{2})_n} \cdot \frac{\sin^{2n+1}(\alpha)}{n!}. \qquad \text{(A1)}\end{aligned}$$

If $(0)_0$ is taken to be unity (as is standard) then this holds $\forall m$ integer (including $m = 1$), in which case, noting that $(\frac{1}{2} - \frac{m}{2})_0 = (\frac{1}{2} + \frac{m}{2})_0 = 1$, then for $n \geq 1$ we see that

$$\begin{aligned}\left(\frac{1}{2} \pm \frac{m}{2}\right)_n &= \left(\frac{1}{2} \pm \frac{m}{2}\right) \left(\frac{1}{2} \pm \frac{m}{2} + 1\right) \left(\frac{1}{2} \pm \frac{m}{2} + 2\right) \cdots \\ &\qquad \cdots \left(\frac{1}{2} \pm \frac{m}{2} + n - 2\right) \left(\frac{1}{2} \pm \frac{m}{2} + n - 1\right) \\ &= \frac{(1 \pm m)(3 \pm m) \cdots [(2n-3) \pm m][(2n-1) \pm m]}{2^n}. \qquad \text{(A2)}\end{aligned}$$

Thus, (A1) can be split as

$$\begin{aligned}&\frac{\sin(m\alpha)}{m} \\ &= \sin(\alpha) + \sum_{n=1}^{\infty} \frac{(\frac{1}{2} - \frac{m}{2})_n (\frac{1}{2} + \frac{m}{2})_n}{(\frac{3}{2})_n} \cdot \frac{\sin^{2n+1}(\alpha)}{n!} \\ &= \sin(\alpha) + \\ &\qquad \sum_{n=1}^{\infty} \frac{(1^2 - m^2)(3^2 - m^2) \cdots [(2n-1)^2 - m^2]}{(2n+1)!} \cdot \sin^{2n+1}(\alpha) \qquad \text{(A3)}\end{aligned}$$

since it is easy to show that

$$\left(\frac{3}{2}\right)_n = \frac{(2n+1)!}{4^n n!}, \qquad n \geq 0; \tag{A4}$$

equation (A3) is, according to Luo [1], due to Euler and gives both finite and infinite series for $\sin(m\alpha)$, as appropriate, for all integer m.

References

[1] Luo, J. (1988). Antu Ming, the first inventor of Catalan numbers in the world, *Neim. Daxue Xuebao*, **19**, pp.239-245.

[2] Larcombe, P.J. (2000). On Catalan numbers and expanding the sine function, *Bull. Inst. Comb. Appl.*, **28**, pp.39-47.

[3] Bailey, W.N. (1935). Generalized hypergeometric series (Cambridge tracts in mathematics and mathematical physics, No. 32), Cambridge University Press, London, U.K.

[4] Slater, L.J. (1966). Generalized hypergeometric functions, Cambridge University Press, London, U.K.

[5] Copson, E.T. (1935). An introduction to the theory of functions of a complex variable, Oxford University Press, London, U.K.