PROVING WALLIS FORMULA FOR π IN A PROBABILITY CONTEXT WITH PROSPECTIVE PRIMARY SCHOOL TEACHERS

Michael KOURKOULOS[a], Constantinos TZANAKIS[b]
Department of Primary Education, University of Crete, Rethymnon, Greece
mkourk@uoc.gr[a], k.tzanakis@uoc.gr[b]

ABSTRACT

Motivated by the historical connections between the Wallis' product formula for π and the approximation of the binomial by the normal distribution in probability theory, we discussed with our students - prospective elementary school teachers - an elementary proof of this formula, which, though initially given in a geometric context, admits a probabilistic interpretation as we showed.

1. *Introduction*

In 1656, Wallis published his product formula for π (W.P.) and the relevant investigation in his *Arithmetica infinitorum* (Stedall, 2004, pp. xvii-xx, proposition 191). W.P. is an important mathematical result that historically had equally important applications, particularly in relation to the early development of probability theory in the 18th century. Specifically, in 1733 De Moivre used it as a key tool in his pioneering work on the normal approximation of the symmetric binomial distri-bution, which in turn was historically the first normal approximation (Hald, 2003, pp. 468-484; Khrushchev, 2006; Stigler, 1986, ch. 2). Moreover, in this context both Stirling and De Moivre found in 1730 (slightly different) approximations of n! (Hald, *ibid*). This use of the W.P. shows its crucial role in the development of prob-ability, so that even in the 18th century the connection between analysis and proba-bility was already coming to light. Therefore, given the importance of the W.P. in this context, we looked exhaustively at its historical development, for getting aided by history to find out how to teach its proof in our introductory course on probabil-ity and statistics addressed to prospective elementary school teachers (no publica-tion seems to exist on empirical didactical work concerning its proof). Our students have a very limited (and in many cases, weak) mathematical background, but never-theless, have to cope with elementary probability and statistics, including the bino-mial distribution, an understanding of the significance of the law of large numbers, the normal distribution and its significance in relation to the central limit theorem and in particular as an approximation to the binomial distribution. This led us to wonder whether there is an elementary probabilistic proof of this formula adequate for our purpose. Although in our historical investigation we identified 15 proofs of W.P. since Wallis' time (most after 1980), none was an elementary probabilistic one[48], and all but two, use advanced mathematics: Yaglom and Yaglom (1987, pp.24, 36-37) use complicated trigonometry; Wästlund (2007) uses elementary algebra and geometry (both proofs employ the squeezing theorem of limits). However, we realized that a feature implicit to the latter (not mentioned by Wästlund) is that basic quantities in the proof can be interpreted as probabilities, hence related algebraic relations as probabilistic properties.

So, with Wästlund's proof translated into an elementary probabilistic one and with a variant of his geometric argument, we taught the proof in the context of the normal approximation to the symmetric binomial. This allowed to comment on how W.P. was used in De Moivre's work even though we did not develop the connection in the way he did. Nevertheless, along these lines it became possible through history to provide the students with hints on the historical connections between probabilistic concepts and mathematical analysis.

2. *The teaching framework*

In the teaching activity 26 volunteers participated, 17 having followed no introductory course on calculus. All were taught in high school the algebra and geometry necessary for the proof, though in an initial test (including questions on polynomial expansion, solving a linear system of equations, properties of powers and fractions) 7 gave answers that indicated significant weaknesses in elementary algebra. Discussion on the proof and applications of W.P. started on the 5th course week for 12 hours.

Until then students had been taught basic elements of combinatorics, the additive and multiplicative rules of probability and its extension to more events (the chain rule of probability) and the binomial and hypergeometric distributions, together with examples and applications.

3. *The implemented teaching approach*

Originally, the teacher told the students that they were going to consider the limit

---

[48] Though several recent proofs employ advanced tools and results of probability theory and mathematical analysis (Chin, 2020; Kovalyov, 2011; Miller, 2008; Wei et al., 2017).

of an important product, $\frac{1\times3}{2^2}\times\frac{3\times5}{4^2}\times\ldots\times\frac{(2n-1)\times(2n+1)}{(2n)^2}$, and that this limit was initially found by Wallis in 1656 and provided information about Wallis, his work, and the importance of W.P. (Stedall, 2004, pp. xi-xxxiii; Khrushchev, 2006). He also said that they were going to study a variant of Wästlund's proof, because it needs only elementary mathematics.

3.1 The first properties discussed with the students

(a) The probability at the center of the symmetric binomial with even number of trials is $B(n,2n,0{,}5) = \frac{(2n)!}{(n!)^2}\times\frac{1}{2^{2n}} = \frac{(2n-1)!!}{(2n)!!}$ with $b_n = B(n,2n,0{,}5)$ for $n \geq 1$, and $b_0 = 1$ by definition. (b) $2n\times b_n = b_{n-1}+\ldots+b_1+b_0$.

3.2 The partial Wallis product and its relation to b_n

The teacher explained that, for $\varphi_0 = 0$, $\varphi_n = b_n^2\times(2n+1)$, $n\in\mathbb{N}$, and rearranging the factors in the numerator of b_n^2, the inverse of the partial W.P. becomes

$$\varphi_n = \frac{1\times3}{2^2}\times\frac{3\times5}{4^2}\times\ldots\times\frac{(2n-1)\times(2n+1)}{(2n)^2}.$$

Moreover, setting $a_n = b_n^2\times 2n$, $a_0 = 0$, $a_1 = \frac{1}{2}$, and for $n \geq 2$ rearranging the factors in the denominator of b_n^2, gives $a_n = \frac{1}{2}\times\frac{3^2}{4\times6}\times\frac{5^2}{6\times8}\times\ldots\times\frac{(2n-1)^2}{(2n-2)\times2n}$.

Then he discussed that (i) $\varphi_n > \varphi_{n+1}$; (ii) $\alpha_{n+1} > \alpha_n$; (iii) $\frac{a_n}{\varphi_n} = \frac{2n}{2n+1}$

Thus $\varphi_n > \alpha_n$ and α_n approaches φ_n as n increases; (iv) by (i)-(iii), α_n and φ_n have the same limit C, still to be found, and $\varphi_n > C > \alpha_n$.

3.3 A simple Pólya-Eggenberger urn model

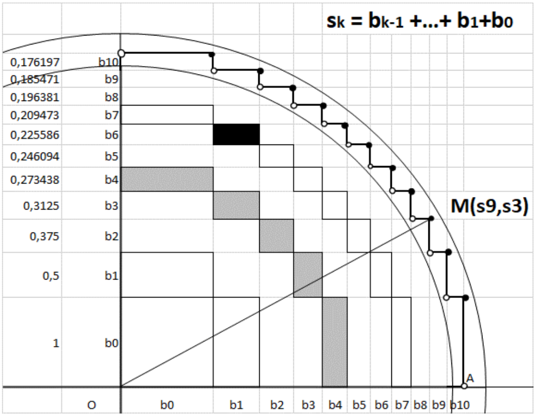
The following property is of central importance in Wästlund's proof

$$b_0b_n + b_1b_{n-1} + \cdot\ \ \cdot\ \ \cdot + b_nb_0 = 1 \qquad (1)$$

He proves it algebraically, but we proved it probabilistically. For this we conceived a simple Pólya urn model, running as follows: Initially, an urn contains a black and a red ball. At each trial, we draw randomly a ball from the urn and return it to the urn plus two new balls of the same color. The model was discussed with the students; in particular, finding the probability $\mathrm{Pl}(k,n)$ to get k black balls in n successive random trials, with their order of occurrence being immaterial. We proved that $\mathrm{Pl}(k,n) = b_k\times b_{n-k}$. Then we got $\mathrm{Pl}(0,n) + \mathrm{Pl}(1,n) +\ldots+ \mathrm{Pl}(n,n)= 1$, since it is the sum of probabilities of all possible events, all being mutually exclusive. Finally, substituting the probabilities in this sum and using $\mathrm{Pl}(k,n) = b_k\times b_{n-k}$ the desired property resulted.

The generalization of the urn model with three parameters was discussed next and information on Polya's work in mathematics, physics, and mathematics education was given (e.g., Alexanderson, 2000). Students' work on Polya's urn models, though limited, was a significant introduction to the subject.

3.4 A grid for representing the probability Pl(k, n)=$b_k \times b_{n-k}$ and its properties



The first column of the grid has width b_0, the next has width b_1, etc. The rows are determined similarly. Columns and rows are enumerated starting from 0. So, column k has width b_k and row m has height b_m. Their intersection is a rectangle, $Rec(k, m)$, with dimensions b_k, b_m and area $AreaRec(k, m) = b_k \times b_m$, e.g. in black color the $Rec(1, 6)$. Let A_k be the sequence of rectangles $Rec(0,k)$, $Rec(1,k\text{-}1)$, …, $Rec(k,0)$ (e.g., A_4 is the sequence in grey). Since $AreaRec(i, j) = b_i \times b_j$, the union of the surfaces of the rectangles of A_k has area: $AreaA_k = b_0 \times b_k + b_1 \times b_{k-1} + \ldots + b_k \times b_0 = 1$, by (1). Let Π_n be the polygon whose surface is the union of the surfaces of the rectangles of the sequences A_0, A_1, A_2, ..., A_{n-1} (e.g., in the figure the last bold perimeter is the perimeter of polygon Π_{11}.). Since $AreaA_k = 1$, the area of Π_n is n. We set $s_k = b_{k-1} + ... + b_1 + b_0$ for $k \geq 1$ and $s_0 = 0$, $s_k = 2k \times b_k$ (recall §3.1). The outer corners of the polygon Π_n have coordinates (s_k, s_{n+1-k}), with k integer from 1 to n. The inner ones have coordinates (s_k, s_{n-k}), with k integer from 0 to n. These results were discussed with the students step by step and using many examples.

3.5 Circular quadrants containing and contained in Π_n

The teacher explained that the quadrant of a circle (O, R_n), with delimiting radii on Ox and Oy, contains Π_n if and only if $\sqrt{s_k^2 + s_{n+1-k}^2} \leq$ R$_n$, for $k \in \mathbb{N}$ $1 \leq k \leq n$. With $s_k = b_{k-1} + \ldots + b_1 + b_0 = 2k \times b_k$, $\alpha_k = b_k^2 \times 2k$ and α_k an increasing sequence, it was obtained that if R$_n = \sqrt{(2n+2) \times \alpha_n}$, then the above quadrant contains Π_n. Since the area of this quadrant is greater than that of Π_n, which is n, we have

$$n < \frac{1}{4}\pi \times \left(\sqrt{(2n+2) \times \alpha_n}\right)^2 = \frac{1}{4}\pi(2n+2) \times \alpha_n \Leftrightarrow \frac{2n}{\pi(n+1)} < \alpha_n \Rightarrow$$

$$\lim_{\nu \to +\infty} \frac{2n}{\pi(n+1)} \leq \lim_{\nu \to +\infty} \alpha_n = c \Rightarrow \frac{2}{\pi \times C} \leq 1 \quad (2)$$

Then, the teacher discussed that similarly there is a quadrant of the circle (O, r_n) with $r_n = \sqrt{(2n-2)\varphi_n}$, contained in Π_n, and from this it was obtained that:

$$1 \leq \frac{2}{\pi \times C} \qquad (3)$$

(3) and (2) imply that $C = \frac{2}{\pi}$ which is the limit sought $\lim_{\nu \to +\infty} \varphi_n = \lim_{\nu \to +\infty} \alpha_n$

3.6 The approximation of b_n

Using this limit and the properties in §3.2 the teacher discussed with the students the derivation of the approximation of b_n;

that for large n (i) $b_n \approx \frac{1}{\sqrt{\pi \times n}}$, (ii) $b_n \approx \sqrt{\frac{2}{\pi(2n+1)}}$ and $\frac{1}{\sqrt{\pi \times n}} > b_n > \sqrt{\frac{2}{\pi(2n+1)}}$.

He also discussed the historical importance of this approximation; that in 1729, after Stirling's suggestion, De Moivre used W.P. to obtain (i); that this was an important step in his work leading to the normal approximation of the symmetric binomial (the historically first normal approximation); and that in 1730 Stirling and De Moivre used W.P. for approximating $n!$ - still another important result (Hald, 2003, pp. 468-484).

4. *On the evaluation of the implemented teaching approach*

In the final test, 17 students answered correctly or with minor errors the questions on the proof of W.P., and 9 made important errors and/or gaps. Among the 7 students, with significant weaknesses in elementary algebra in the initial test (group B) only one answered correctly these questions, whereas, among the 19 students with no such weaknesses in the initial test (group A), 16 answered correctly or with minor errors. This difference between groups A and B is significant at the level of 0.01 (Fisher exact test $p = 0.00223$).

5. *Final remarks*

Teaching a version of Wästlund's proof about W.P. in a probabilistic context, was a journey based on the interplay among elementary algebra, probability, and geometry. In its context history played a double role:

(1) At the meta-cognitive level by contributing to the development of our didactical background as teachers/researchers (Tzanakis et al, 2000, section 7.2(c), p. 206) in the sense (i) of enriching our didactical repertoire; (ii) in getting aware of how "advanced" may be the subject to be taught in relation to the students to whom its teaching is addressed; and (iii) in getting involved into the creative process of "doing mathematics".

(2) Part of (iii) above, and a lot of information on the role of W.P. in De Moivre's and Stirling's related work were presented and discussed in the classroom, both in the form of "historical snippets" (Tzanakis & Arcavi, 2000, section 7.4.1) and strictly as a mathematical subject. For instance, we examined the application of W.P. for approximating b_n and discussed its importance in the context of De Moivre's work for the normal approximation to the symmetric binomial. Furthermore, we discussed some interesting properties of probabilities and their geometric representations, including an introduction to Pólya's urn models. This in turn, motivated further discussion on the origin and importance of these models and Polya's multifaceted contributions, including mathematics education. By devoting enough teaching time and a significant amount of work, the proof of W.P. presented here was accessible to students with no significant weaknesses in elementary algebra, though this was not possible for the weaker students.

In summary, our approach could be understood as an example of an "illumination approach" in the sense of Jankvist (2009, section 6.1, pp. 245-246).

REFERENCES

Alexanderson, G. L. (2000). *The random walks of George Pólya*. The Mathematical Association of America.

Chin, W. (2020). A probabilistic proof of a Wallis-type formula for the gamma function. *The American Mathematical Monthly*, *127*(1), 75–79.

Hald, A. (2003). *A history of probability and statistics before 1750*. Wiley.

Jankvist, U. T. (2009). A categorization of the "whys" and "hows" of using history in mathematics education. *Educational Studies in Mathematics*, *71*(3), 235–261.

Khrushchev, S. (2006). A recovery of Brouncker's proof for the quadrature continued fraction. *Publicacions Matemàtiques*, *50*(1), 3-42.

Kovalyov, M. (2011). Removing magic from the normal distribution and the Stirling and Wallis formulas. *The Mathematical Intelligencer*, *33*(4), 32–36.

Miller, S. J. (2008). A probabilistic proof of Wallis's formula for π. *The American Math-*

ematical Monthly, *115*(8), 740–745.

Stedall, J. A. (2004). *The Arithmetic of Infinitesimals, John Wallis 1656*. Springer.

Stigler, S.M. (1986). *The history of statistics: the measurement of uncertainty before 1900*. Harvard University Press.

Tzanakis, C., Arcavi, A., de Sá, C. C., Isoda, M., Lit, C-K., Niss, M., … & Siu, M.-K. (2000). Integrating history of mathematics in the classroom: an analytic survey. In J. G. Fauvel, & J. van Maanen (Eds.), *History in Mathematics Education: The ICMI Study*, New ICMI Study Series, vol. 6 (pp. 201–240). Kluwer.

Wästlund, J. (2007). An elementary proof of the Wallis Product Formula for pi. *The American Mathematical Monthly*, *114*(10), 914–917.

Wei, Z., Li, J., Zheng, X. (2017). A probabilistic approach to Wallis' formula. *Communications in Statistics-Theory and Methods*, *46*(13), 6491–6496.

Yaglom, A. M., & Yaglom, I. M. (1987). *Challenging mathematical problems with elementary solutions* (vol. II, problem 147). Dover.