A characterization of the arcsine distribution

Karl Michael Schmidt* and Anatoly Zhigljavsky*

Abstract. The following characterization of the arcsine density is established: let ξ be a r.v. supported on $(-1,1)$, then ξ has the arcsine density $p(t) = 1/(\pi\sqrt{1-t^2})$, $-1 < t < 1$, if and only if $\mathbb{E}\log(\xi - x)^2$ has the same value for almost all $x \in [-1,1]$.

1 Introduction

The arcsine density on the interval $(-1,1)$ is

$$p(t) = \frac{1}{\pi\sqrt{1-t^2}}\,, \quad -1 < t < 1\,. \tag{1}$$

To define a r.v. ξ with the arcsine density (1) we can use the formula $\xi = \cos(\pi\alpha)$, where α is a r.v. with uniform distribution on (0,1). The arcsine density has several non-trivial appearances in probability theory and statistics. For example, for a general random walk $\{S_n\}$ satisfying the Lindeberg-Lévy condition, the limiting distribution of $\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{[S_i>0]}$ (as $n \to \infty$) has the arcsine density on (0,1), see §11 in Billingsley (1968), Erdős and Kac (1947), Lévy (1948). The arcsine density (1) is an invariant density for a number of maps of the interval $(-1,1)$ onto itself, see e.g. Rivlin (1990), Theorem 4.5. This density is the limiting density of the roots of the orthogonal polynomials which are defined on (-1,1) and orthogonal with respect to any weight function $w(\cdot)$ continuous on $(-1,1)$, see Ullman (1972), Erdős and Freud (1974), van Assche (1987).

In probability theory, the arcsine density has a number of characterizations, see Norton (1975, 1978), Arnold and Groenveld (1980), Kemperman and Skibinsky (1982). Below we consider a characterization of the arcsine density that is of a different nature than the ones considered in these papers.

Our main result is as follows.

Theorem. *Let ξ be a r.v. supported on $(-1,1)$. This r.v. has the arcsine density (1) if and only if $\mathbb{E}\log(\xi - x)^2$ has the same value for almost all $x \in [-1,1]$.*

As a motivation for the above theorem, assume that we have a sequence of points $x_1, x_2, \ldots$ in the interval $(-1,1)$ that has an asymptotic c.d.f. $F(\cdot)$ in the sense that

$$\lim_{k\to\infty} \frac{1}{k}\sum_{j=1}^{k} h(x_j) = \int_{-1}^{1} h(t)\, dF(t) \tag{2}$$

for any continuous function $h(\cdot)$ such that $\int |h(x)|\, dF(x) < \infty$. Consider an associated sequence of polynomials $H_k(x) = (x-x_1)^2\,(x-x_2)^2 \cdots (x-x_k)^2$. Then the result of the Theorem implies that the values of the normalized ratios $R_k(x,y) = [H_k(x)/H_k(y)]^{1/k}$ tend to 1 (as $k \to \infty$) for almost all $x, y \in (-1,1)$ if and only if the c.d.f. $F(\cdot)$ has the arcsine density (1). Indeed,

$$\log R_k(x,y) = \log[H_k(x)]^{1/k} - \log[H_k(y)]^{1/k} = \frac{1}{k}\sum_{j=1}^{k}\log(x-x_j)^2 - \frac{1}{k}\sum_{j=1}^{k}\log(y-x_j)^2\,.$$

*School of Mathematics, Cardiff University, Senghennydd Road, Cardiff, CF24 4YH, UK (SchmidtKM@cf.ac.uk, ZhigljavskyAA@cf.ac.uk)

Using (2), for almost all $x, y \in [-1,1]$ we obtain

$$\log\left[\lim_{k\to\infty} R_k(x,y)\right] = \lim_{k\to\infty} \log R_k(x,y) = \int_{-1}^{1} \log(x-t)^2\, dF(t) - \int_{-1}^{1} \log(y-t)^2\, dF(t)\,. \qquad (3)$$

The theorem implies that the r.h.s. of (3) is zero for almost all $x, y \in [-1,1]$ if and only if the c.d.f. $F(\cdot)$ has the arcsine density (1).

The fact that $R_k(x,y) \to 1$ (as $k \to \infty$) for almost all $x, y \in (-1,1)$ means that the ratios $H_k(x)/H_k(y)$ are almost never very large (these ratios are smaller than δ^k with any $\delta > 1$ for sufficiently large k: $k > k^*(x,y)$) and very rarely are very close to 0 (they are larger than any δ^k with any $\delta < 1$ and $k > k_*(x,y)$). Note that if k is fixed and $x_j = \cos(\pi(2j-1)/(2k))$ for $j = 1, \ldots, k$, then $H_k(x) = c_k T_k^2(t)$ where c_k is some constant and $T_k(x) = \cos[k \arccos(x)]$ is the k-th Chebyshev polynomial. In this case, the fact that $R_k(x,y) \cong 1$ (as k is large) for typical $x, y \in (-1,1)$ follows from the properties of the Chebyshev polynomials.

2 Auxiliary statements and proofs

The proof of the theorem is based on three Lemmas. In Lemma 1 we observe that the expected value of $\log(\xi - x)^2$ is finite for almost all $x \in [0,1]$. In Lemma 2 we derive a specific characterization of the uniform measure on the interval $[0,\pi]$. To prove Lemma 2 we use a general characterization of the Lebesgue measure on the interval $[0,\pi]$ established in Lemma 3.

Lemma 3 uses Fourier series, which may seem surprising but is a natural reflection of the intrinsic relationship between the arcsine distribution and trigonometric powers, as apparent in the Chebyshev polynomials.

Lemma 1. *For any r.v. ξ supported on $(-1,1)$, the expectation $\mathbb{E}\log(\xi - x)^2$ is finite for almost all $x \in [-1,1]$.*

Proof of Lemma 1. Let $F(\cdot)$ be the c.d.f. of the r.v. ξ and $-1 < t < 1$. The integral

$$\int_{-1}^{1} \log(t-x)^2\, dx = (1+t)\log(1+t)^2 + (1-t)\log(1-t)^2 - 4$$

is bounded and continuous as a function of t, so the integral

$$\int_{-1}^{1}\int_{-1}^{1} \log(t-x)^2\, dx\, dF(t)$$

exists. By the Fubini-Tonelli theorem,

$$\mathbb{E}\log(\xi - x)^2 = \int_{-1}^{1} \log(t-x)^2\, dF(t) \in L^1([-1,1]),$$

so in particular it is finite for almost all $x \in [-1,1]$. ■

Lemma 2. *Let φ be a r.v. distributed according to a probability measure $\mu(\cdot)$ on $[0,\pi]$. Then the expectation*

$$E_x(\mu) = \mathbb{E}\log(\cos\varphi - x)^2$$

is constant for almost all $x \in [-1,1]$ if and only if the measure $\mu(\cdot)$ is uniform on $[0,\pi]$; in this case, the expectation $E_x(\mu)$ has the same value for all $x \in [-1,1]$.

Proof of Lemma 2. As $x \in [-1,1]$, we can set $\psi := \arccos x \in [0,\pi]$. Let us extend μ to $[0,2\pi]$ as an even measure (that is, we set $\mu(A) = \mu(2\pi - A)$ for all Borel sets $A \subset [\pi, 2\pi]$) and note that $\mu([0,2\pi]) = 2$. Using $\cos\varphi = \cos(2\pi - \varphi)$ for all $\varphi \in \mathbb{R}$, we calculate

$$E_x(\mu) = \frac{1}{2}\int_0^{2\pi} \log(\cos\varphi - x)^2\,\mu(d\varphi) = \frac{1}{2}\int_0^{2\pi} \log\left(2\sin\frac{\varphi-\psi}{2}\sin\frac{\varphi+\psi}{2}\right)^2 \mu(d\varphi)$$

$$= \frac{1}{2}\left[\int_0^{2\pi} 2\,\log 2\,\mu(d\varphi) + \int_0^{2\pi} \log\left(\sin\frac{\varphi-\psi}{2}\right)^2 \mu(d\varphi) + \int_0^{2\pi} \log\left(\sin\frac{\varphi+\psi}{2}\right)^2 \mu(d\varphi)\right]$$

$$= 2\,\log 2 + \int_0^{\pi} \log\left(\sin^2\left(\varphi - \psi/2\right)\right)\,\tilde{\mu}(d\varphi) + \int_0^{\pi} \log\left(\sin^2\left(\varphi + \psi/2\right)\right)\,\tilde{\mu}(d\varphi)\,, \tag{4}$$

where $\tilde{\mu}(A) = \frac{1}{2}\mu(2A)$ for all Borel sets $A \subset [0,\pi]$.

As $\tilde{\mu}$ and $\sin^2$ are π-periodic and even, we obtain by making the substitution $\tilde{\varphi} = \pi - \varphi$:

$$\int_0^{\pi} \log\left(\sin^2\left(\varphi + \psi/2\right)\right)\,\tilde{\mu}(d\varphi) = \int_0^{\pi} \log\left(\sin^2\left(\pi - \tilde{\varphi} + \psi/2\right)\right)\,\tilde{\mu}(d\tilde{\varphi}) = \int_0^{\pi} \log\left(\sin^2\left(\tilde{\varphi} - \psi/2\right)\right)\,\tilde{\mu}(d\tilde{\varphi})\,.$$

This implies that the two integrals in (4) are identical and therefore

$$E_x(\mu) = 2\log 2 + 2\int_0^{\pi} \log\left(\sin^2\left(\varphi - \psi/2\right)\right)\,\tilde{\mu}(d\varphi)\,.$$

Hence the expectation $E_x(\mu)$ is constant for almost all $x \in [-1,1]$ if and only if

$$\log\sin^2 \star\, \tilde{\mu}\,(y) = \int_0^{\pi} \log\left(\sin^2\left(\varphi - y\right)\right)\,\tilde{\mu}(d\varphi) \quad \text{is constant for almost all } y \in [0,\pi]\,. \tag{5}$$

The Fourier series for $\log\sin^2$ is not uniformly convergent, but it converges in the $L^2([0,\pi])$ sense, as $\log\sin^2 \in L^2([0,\pi])$ and $\{e^{2ikx} | k \in \mathbb{Z}\}$ is an orthonormal basis of this Hilbert space. Moreover, all (complex) Fourier coefficients of $\log\sin^2$ are real and non-zero. Indeed,

$$\int_0^{\pi} \log\left(\sin^2(\varphi)\right)\sin(2k\varphi)d\varphi = 0 \quad \forall k \in \mathbb{Z}$$

and

$$\int_0^{\pi} \log\left(\sin^2(\varphi)\right)\cos(2k\varphi)d\varphi = 2\pi\int_0^{1} \log\left(\sin(\pi t)\right)\cos(2\pi k t)dt = \begin{cases} -2\pi\log 2, & k = 0 \\ -\pi/k, & k \in \mathbb{Z}\setminus\{0\}\,, \end{cases}$$

see formula 4.384.3 in Gradshtein and Ryzhik (1965). The statement of Lemma 2 now follows from Lemma 3 below. ■

Lemma 3. *Let $\tilde{\mu}$ be a probability measure on $[0,\pi]$ and $f \in L^2([0,\pi])$ be such that*

$$f(x) = \text{l.i.m.}_{N\to\infty} \sum_{k=-N}^{N} \theta_k e^{2ikx} \quad (x \in [0,\pi])$$

where all Fourier coefficients are non-zero: $\theta_k \neq 0\ \forall k \in \mathbb{Z}$. Then, extending f to $\mathbb{R}$ as a π-periodic function, the convolution $f \star \tilde{\mu}\,(\cdot) := \int_0^{\pi} f(\cdot - t)\,\tilde{\mu}(dt)$ is constant almost everywhere if and only if $\tilde{\mu}$ is the uniform measure on $[0,\pi]$; in this case, $f \star \tilde{\mu}$ is constant everywhere.

Proof of Lemma 3. Assume

$$f \star \tilde{\mu}\,(x) = \int_0^\pi f(x-t)\,\tilde{\mu}(dt) = C = \text{const} \;\; (\text{for almost all } x \in [0,\pi])\,.$$

Then, for all $k \in \mathbb{Z} \setminus \{0\}$,

$$0 = \int_0^\pi e^{2ikx} C dx = \int_0^\pi e^{2ikx} \left[\int_0^\pi f(x-t)\,\tilde{\mu}(dt)\right] dx = \int_0^\pi \left[\int_0^\pi e^{2ikx} \;\text{l.i.m.}_{N\to\infty} \sum_{n=-N}^{N} \theta_n e^{2in(x-t)} dx\right] \tilde{\mu}(dt)$$

$$= \int_0^\pi \lim_{N\to\infty} \sum_{n=-N}^{N} \theta_n e^{-2int} \left[\int_0^\pi e^{2i(k+n)x} dx\right] \tilde{\mu}(dt) = \int_0^\pi \lim_{N\to\infty} \sum_{n=-N}^{N} \theta_n e^{-2int} \pi\, \delta_{n,-k}\, \tilde{\mu}(dt)$$

$$= \pi \theta_{-k} \int_0^\pi e^{2ikt} \tilde{\mu}(dt)\,.$$

As $\theta_{-k} \neq 0 \;\forall k \in \mathbb{Z} \setminus \{0\}$ we get $\int_0^\pi e^{2ikt}\tilde{\mu}(dt) = 0 \;\forall k \in \mathbb{Z} \setminus \{0\}$.

Now set $\mu' = \tilde{\mu} - \tilde{\mu}([0,\pi])\lambda/\pi$, where λ is the Lebesgue measure on $[0,\pi]$. Then

$$\int_0^\pi \mu'(dt) = 0 \;\text{ and }\; \int_0^\pi e^{2ikt}\mu'(dt) = \int_0^\pi e^{2ikt}\tilde{\mu}(dt) - \frac{\tilde{\mu}([0,\pi])}{\pi} \int_0^\pi e^{2ikt} dt = 0\,, \;\; \forall k \in \mathbb{Z} \setminus \{0\}$$

as shown above, so $\int_0^\pi e^{2ikt}\mu'(dt) = 0 \;\; \forall k \in \mathbb{Z}$. As every continuous function on $[0,\pi]$ can be uniformly approximated by a linear combination of $\{e^{2ikt} | k \in \mathbb{Z}\}$ and μ' is finite, this implies

$$\int_0^\pi f(t)\mu'(dt) = 0 \;\; \forall f \in C([0,\pi])$$

and hence $\mu' = 0$. This completes the proof of the 'only if' part of Lemma 3. The converse is obvious, bearing in mind that f is π-periodic and $f(x-\cdot) \in L^1([0,\pi])$ for all $x \in \mathbb{R}$. ■

Proof of the Theorem. Consider the expectation

$$I_x = \mathbb{E}\log(\xi - x)^2 = \int_{-1}^{1} \frac{\log(t-x)^2}{\pi\sqrt{1-t^2}}\, dt\,, \tag{6}$$

where r.v. ξ has the arcsine density (1). By changing the variable $t = \cos\varphi$ we obtain

$$I_x = \int_0^\pi \frac{\log(\cos\varphi - x)^2}{\pi \sin\varphi} \sin\varphi\, d\varphi = \frac{1}{\pi}\int_0^\pi \log(\cos\varphi - x)^2\, d\varphi \; = E_x(\mu_0), \tag{7}$$

where μ_0 is the uniform measure on $[0,\pi]$. Hence, by applying Lemma 2 with $\mu = \mu_0$, we conclude that I_x has the same value for all $x \in [-1,1]$. This proves the 'only if' statement in the Theorem.

To complete the proof of the Theorem, we now show the converse, i.e. that if, for a random variable ξ supported on $(-1,1)$, $\mathbb{E}\log(\xi - x)^2$ has the same value for almost all $x \in [-1,1]$, then ξ has the arcsine density. In view of Lemma 1 the constant value of $\mathbb{E}\log(\xi-x)^2$ must be finite. Denote by $F(\cdot)$ the c.d.f. of ξ. Then $F(-1) = 0$, $F(1) = 1$ and

$$\mathbb{E}\log(\xi - x)^2 = \int_{-1}^{1} \log(t-x)^2\, dF(t) = \int_0^\pi \log(\cos\varphi - x)^2\, d\tilde{F}(\varphi),$$

where $t = \cos\varphi$ and $\tilde{F}(\varphi) = 1 - F(\cos\varphi)$. By Lemma 2, the probability measure generated by $\tilde{F}$ is uniform on $[0,\pi]$; that is, $\tilde{F}(\varphi) = \varphi/\pi \;\; \forall\varphi \in [0,\pi]$. This implies

$$F(x) = 1 - (\arccos x)/\pi \;\; \forall x \in (-1,1),$$

so the density of ξ is $F'(x) = 1/(\pi\sqrt{1-x^2})$. ■

3 Explicit formulae for the integrals and a generalization

3.1 Explicit formulae for the expectations

The value of the expectation (6) can be easily computed based on our result that it is independent of x in the interval $[-1,1]$.

Corollary 1. *Let the r.v.* ξ *have density* (1). *Then*

$$I_x = \mathbb{E}\log(\xi - x)^2 = \begin{cases} -2\log 2 & \text{if } |x| \le 1 \\ 2\log\left(|x| + \sqrt{x^2-1}\right) - 2\log 2 & \text{if } |x| \ge 1\,. \end{cases} \tag{8}$$

Proof. For $|x| \le 1$ we use $I_x = I_0 = -2\log 2$ by evaluating the integral I_0:

$$I_x = I_0 = \frac{1}{\pi}\int_0^{\pi} \log\left(\sin^2(\varphi)\right)\, d\varphi = \frac{2}{\pi}\int_0^{\pi} \log\left(\sin\varphi\right)\, d\varphi = -2\log 2\,. \tag{9}$$

Let now $x \ge 1$. From (9) we have $I_1 = -2\log 2$. Differentiating I_x we get

$$I_x' = \left(\int_{-1}^{1} \frac{\log(x-t)^2}{\pi\sqrt{1-t^2}}dt\right)' = \int_{-1}^{1} \frac{2}{\pi(x-t)\sqrt{1-t^2}}\, dt = 2\int_0^1 \frac{ds}{\pi(x+1-2s)\sqrt{s(1-s)}} = \frac{2}{\sqrt{x^2-1}}\,;$$

(see Gradshtein and Ryzhik (1965) 3.121.2 — note that interchanging the differentiation and integration is justified as the derivative of the integrand is bounded by an integrable function of t locally uniformly in x, $|x| > 1$). Therefore, for $x > 1$,

$$I_{-x} = I_x = I_1 + \int_1^x I_z' dz = -2\log 2 + \int_1^x \frac{2}{\sqrt{z^2-1}}dz = 2\log\left(\frac{x+\sqrt{x^2-1}}{2}\right)\,. \tag{10}$$

Combining (9) and (10) we obtain (8). ■

3.2 Arcsine density on an arbitrary interval

The arcsine density on an interval (a,b) is

$$p(t) = \frac{1}{\pi\sqrt{(t-a)(b-t)}}\,, \quad a < t < b\,. \tag{11}$$

If $a = -1$ and $b = 1$ then (11) is reduced to (1). A simple change of variables generalizes Theorem 1 to the following statement.

Corollary 2. *Let* $-\infty < a < b < \infty$ *and let* ζ *be a r.v. supported on the interval* (a,b)*. The r.v.* ζ *has the arcsine density* (11) *if and only if* $\mathbb{E}\log(\zeta - z)^2$ *has the same value for almost all* $z \in [a,b]$.

Corollary 1 is generalized as follows.

Corollary 3. *Let* $-\infty < a < b < \infty$ *and let r.v.* ζ *have density* (11). *Then*

$$\mathbb{E}\log(\zeta - z)^2 = \begin{cases} 2\log(b-a) - 4\log 2 & \text{if } a \le z \le b \\ 2\log(b-a) + 2\log\left(|x_z| + \sqrt{x_z^2 - 1}\right) - 4\log 2 & \text{if } z < a \text{ or } z > b\,, \end{cases} \tag{12}$$

where $x_z = -1 + 2(z-a)/(b-a)$.

Proof. By changing variables $t = -1 + 2(u-a)/(b-a)$ and $x = -1 + 2(z-a)/(b-a)$ in the integral

$$\int_a^b \frac{\log(u-z)^2}{\pi\sqrt{(u-a)(b-u)}}\, du = \mathbb{E} \log(\zeta - z)^2 \,,$$

we obtain $\mathbb{E} \log(\zeta - z)^2 = 2\log(b-a) - 2\log 2 + I_x$, where I_x is defined in (8). This immediately implies (12). ■

References

[1] B.C. Arnold and R.A. Groeneveld. Some properties of the arcsine distribution. *J. Amer. Statist. Assoc.*, 75(369):173–175, 1980.

[2] P. Billingsley. *Convergence of probability measures.* John Wiley & Sons Inc., New York, 1968.

[3] P. Erdős and G. Freud. On orthogonal polynomials with regularly distributed zeros. *Proc. London Math. Soc. (3)*, 29:521–537, 1974.

[4] I.S. Gradshteyn and I.M. Ryzhik. *Table of integrals, series, and products.* Academic Press, New York, 1965.

[5] J.H.B. Kemperman and M. Skibinsky. On the characterization of an interesting property of the arcsin distribution. *Pacific J. Math.*, 103(2):457–465, 1982.

[6] P. Lévy. *Processus Stochastiques et Mouvement Brownien.* Gauthier-Villars, Paris, 1948.

[7] R.M. Norton. On properties of the arc sine law. *Sankhyā Ser. A*, 37(2):306–308, 1975.

[8] R.M. Norton. Moment properties and the arc sine law. *Sankhyā Ser. A*, 40(2):192–198, 1978.

[9] T.J. Rivlin. *Chebyshev polynomials.* John Wiley & Sons Inc., New York, second edition, 1990.

[10] J. L. Ullman. On the regular behaviour of orthogonal polynomials. *Proc. London Math. Soc. (3)*, 24:119–148, 1972.

[11] W. Van Assche. *Asymptotics for orthogonal polynomials.* Springer-Verlag, Berlin, 1987.