

On an intriguing distributional identity ¹

M.C. JONES^a, ÉRIC MARCHAND^b, WILLIAM E. STRAWDERMAN^c

a School of Mathematics and Statistics, The Open University, Walton Hall, Milton

Keynes MK7 6AA, UK (e-mail: m.c.jones@open.ac.uk)

b Université de Sherbrooke, Département de mathématiques, Sherbrooke Qc,

CANADA, J1K 2R1 (e-mail: eric.marchand@usherbrooke.ca)

c Rutgers University, Department of Statistics and Biostatistics, 561 Hill Center,

Busch Campus, Piscataway, N.J., USA, 08855 (e-mail: straw@stat.rutgers.edu)

SUMMARY

For a continuous random variable X with support equal to (a, b) , with c.d.f. F , and $g : \Omega_1 \rightarrow \Omega_2$ a continuous, strictly increasing function, such that $\Omega_1 \cap \Omega_2 \supseteq (a, b)$, but otherwise arbitrary, we establish that the random variables $F(X) - F(g(X))$ and $F(g^{-1}(X)) - F(X)$ have the same distribution. Further developments, accompanied by illustrations and observations, address as well the equidistribution identity $U - \psi(U) \stackrel{d}{=} \psi^{-1}(U) - U$ for $U \sim U(0, 1)$, where ψ is a continuous, strictly increasing and onto function, but otherwise arbitrary. Finally, we expand on applications with connections to variance reduction techniques, the discrepancy between distributions, and a risk identity in predictive density estimation.

AMS 2010 subject classifications: 60E05, 62E15.

Keywords and phrases: Equidistribution; Identity; Uniform distribution.

1 Introduction

Characterizations of probability distributions have a long history and include equidistributional results, some of which are more celebrated than others, but many of which

¹June 29, 2017

are very elegant. As an exemplar, for a Cauchy distributed random variable X with density $\{\pi(1+x^2)\}^{-1}$ on \mathbb{R} , the transformed random variable $(X - X^{-1})/2$ has the same distribution as X . Additionally, the equidistribution even characterizes the above Cauchy distribution among continuous random variables (Arnold, 1979).

Perhaps the most fundamental equidistributional result is the identity $U \stackrel{d}{=} 1 - U$ for $U \sim U(0, 1)$. From this, it follows that $F(X) \stackrel{d}{=} 1 - F(X)$ and therefore

$$X \stackrel{d}{=} F^{-1}(1 - F(X)), \quad (1)$$

where X is a continuous random variable with strictly increasing c.d.f. F on (a, b) (we note that this also extends to the case where F is not strictly increasing by defining $F^{-1}(u) = \inf\{x : F(x) \geq u\}$). What is less known, but still straightforward, is that the identity $X \stackrel{d}{=} g(X)$ with g continuous, strictly decreasing implies or characterizes $g \equiv F^{-1}(1 - F)$. Indeed, for $t \in (a, b)$, the condition $Y \stackrel{d}{=} g(X) \stackrel{d}{=} X \sim F$ tells us that $F(t) = \mathbb{P}(g^{-1}(Y) \leq t) = \mathbb{P}(Y \geq g(t)) = 1 - F(g(t))$, which implies that $g(t) = F^{-1}(1 - F(t))$. Similarly, it is also the case that $g \equiv g^{-1}$, i.e., g is its own inverse (e.g. Kucerovsky et al., 2005).

This note concerns the novel equidistributional identity

$$F(X) - F(g(X)) \stackrel{d}{=} F(g^{-1}(X)) - F(X), \quad (2)$$

for a continuous random variable X with support equal to (a, b) , with c.d.f. F , and $g : \Omega_1 \rightarrow \Omega_2$ a continuous, strictly increasing function, such that $\Omega_1 \cap \Omega_2 \supseteq (a, b)$, but otherwise arbitrary. Here, it is not required that g be an onto function (i.e., $\Omega_1 = \Omega_2$), and even less $\Omega_1 = \Omega_2 = (a, b)$, but we do require that g and g^{-1} be well defined on the support of X , whence the assumption $\Omega_1 \cap \Omega_2 \supseteq (a, b)$.

We put forth the particularly appealing nature of result (2), it seeming to be a ‘natural’ non-monotone complement to (1). The result will be established in Theorem 2.1 and, as far as we can tell, has not been presented previously in the literature. Obviously, identity (2) continues to hold if the c.d.f. F is replaced throughout by the survival function \bar{F} .

Example 1.1. *Here is a simple, immediate, example, helpful to see how identity (2) can arise. Take $X \sim U(0, 1)$, and $g(x) = cx$ with $c \in (0, 1)$, for which (2) applies. Here is a direct proof, however. On the one hand, we have $F(X) - F(cX) = (1-c)X \sim U(0, 1-c)$. We also have $F\left(\frac{X}{c}\right) - F(X) = {}^d \left(\frac{1}{c} - 1\right) X \mathbb{I}_{(0,c]}(X) + (1-X) \mathbb{I}_{(c,1)}(X)$, so that*

$$\mathbb{P}\left(F\left(\frac{X}{c}\right) - F(X) \geq t\right) = \mathbb{P}\left(\frac{ct}{1-c} \leq X \leq 1-t\right) = 1 - \frac{t}{1-c},$$

which indeed implies a $U(0, 1-c)$ distribution as well for $F\left(\frac{X}{c}\right) - F(X)$.

For $U \sim U(0, 1)$, a second identity,

$$U - \psi(U) = {}^d \psi^{-1}(U) - U, \tag{3}$$

holds for continuous and strictly increasing functions ψ , such that $\psi(0) = 0$, $\psi(1) = 1$. This follows by setting $g \equiv F^{-1} \circ \psi \circ F$ in (2), since $F(X) \sim U(0, 1)$. In the latter case, g is an onto function such that $\Omega_1 = \Omega_2 = (a, b) = (0, 1)$, and identities (2) and (3) are equivalent for such g 's. (Indeed, setting $a = 0$, $b = 1$ and $g = \psi$, (3) also arises from (2) by choosing F to be the $U(0, 1)$ c.d.f.) However, (3) does not imply, or lead to (2), in situations where $\Omega_1 \neq \Omega_2$ or $\Omega_1 = \Omega_2 \neq (a, b)$. Identity (3) is appealing due to its simplicity, and as a ‘natural’ non-monotone complement to the simple monotone $U \sim 1 - U$ relationship. Also, observe that ψ can be taken to be a strictly increasing c.d.f. on $(0, 1)$ and ψ^{-1} the corresponding, strictly increasing, quantile function, or vice-versa.

The rest of the paper is organized as follows. Proofs, along with some related properties, are presented in Section 2. Developments relative to specific examples of identities (2) and (3) are provided in Section 3. In particular, these involve the common distribution in (3) when $\psi(t) = t^c$ for $c = 2, 3, 4$. Finally, Section 4 is devoted to applications.

2 Main results

Theorem 2.1. *Let X be a continuous random variable with c.d.f. F and support equal to or contained in (a, b) . Let $g : \Omega_1 \rightarrow \Omega_2$ be a continuous, strictly increasing function, such that $\Omega_1 \cap \Omega_2 \supseteq (a, b)$. Let $H_1(X) = F(X) - F(g(X))$ and $H_2(X) = F(g^{-1}(X)) - F(X)$. Then $H_1(X)$ and $H_2(X)$ have identical distributions.*

Proof. We show that $\mathbb{P}(H_1(X) > t) = \mathbb{P}(H_2(X) > t)$ for all $t \in \mathbb{R}$ (although $H_1(X)$ and $H_2(X)$ have support contained in $(-1, 1)$). Fix t and let

$$A = \{x : H_1(x) > t\} = \sum_{n=1}^M (a_n, b_n),$$

setting $M = \infty$ whenever the set of crossings of $H_1(x)$ and t is countably infinite. Here, the continuity of H_1 implies that A is an open set and thus decomposable as an at most countable number of intervals. Similarly, consider $B = \{x : H_2(x) > t\}$ and observe that, because $H_2(g(X)) = H_1(X)$, $x \in A \iff g(x) \in B$ implying that

$$B = \sum_{n=1}^M (g(a_n), g(b_n)).$$

It thus follows that

$$\mathbb{P}(H_1(X) > t) = \sum_{n=1}^M \{F(b_n) - F(a_n)\}, \quad \mathbb{P}(H_2(X) > t) = \sum_{n=1}^M \{F(g(b_n)) - F(g(a_n))\}. \quad (4)$$

Finally, since $H_1(a_n) = H_1(b_n)$ by definition of the a_n, b_n 's, we obtain that $F(b_n) - F(a_n) = F(g(b_n)) - F(g(a_n))$ for all $1 \leq n \leq M$, and the equality of $\mathbb{P}(H_1(X) > t)$ and $\mathbb{P}(H_2(X) > t)$ follows from (4), for all t . \square

As mentioned in the Introduction, identity (2) implies identity (3), but not vice-versa. An interesting consequence of (3) is, of course, equality of expectations of any absolutely continuous function, h say, of the random variables concerned, i.e.,

$$\int_0^1 h\{u - \psi(u)\} du = \int_0^1 h\{\psi^{-1}(z) - z\} dz. \quad (5)$$

On the other hand, if one were to establish (5) directly, then we would have a second, independent proof of (3); this is done in Theorem 2.2.

Theorem 2.2. *Equation (5) holds for absolutely continuous h . Consequently, for $U \sim U(0, 1)$, the random variables $U - \psi(U)$ and $\psi^{-1}(U) - U$ have identical distributions for any function $\psi : (0, 1) \rightarrow (0, 1)$ continuous, strictly increasing and onto.*

Proof. For the equidistributional part, it will more than suffice to establish (5) as this implies equality of the moment generating functions by the selection $h(t) = e^{st}$, $s \in \mathbb{R}$.

Write

$$\int_0^1 h\{u - \psi(u)\}du = \int_0^1 \{1 - \psi'(u)\}h\{u - \psi(u)\}du + \int_0^1 \psi'(u)h\{u - \psi(u)\}du. \quad (6)$$

The result follows by the substitution $w = u - \psi(u)$ in the first integral on the right-hand side of (6) yielding zero, and the substitution $z = \psi(u)$ for the second integral yielding the right-hand side of (5). \square

3 Examples and discussion

Identity (2) is universal with respect to F and g , while identity (3) is universal with respect to ψ , subject to the given conditions, of course. A long list of examples is not pertinent. Here is thus a selection of illustrative examples, complementary developments and remarks.

Example 3.1. *Consider $X \sim \text{Exp}(\lambda)$ with c.d.f. $F(t) = 1 - e^{-\lambda t}$ for $t > 0$. Theorem 2.1 tells us, by the selection $g(t) = ct$ with $c > 0$, that $e^{-c\lambda X} - e^{-\lambda X}$ and $e^{-\lambda X} - e^{-\lambda X/c}$ share the same distribution. Here g is onto and equivalent identity (3) yields the equidistribution of $U^c - U$ and $U - U^{1/c}$, for $U \sim U(0, 1)$, which we will investigate further in Example 3.2.*

The basic identity (2) will hold for exponential F and functions g satisfying the given conditions. For instance, with the location shift $g(t) = t + \alpha$, $\alpha > 0$, writing $F(t) = \max\{0, 1 - e^{-\lambda t}\}$, Theorem 2.1 implies that $e^{-\lambda X}(1 - e^{-\lambda \alpha})$ and $\min\{1 -$

$e^{-\lambda X}, e^{-\lambda X}(e^{\lambda\alpha} - 1)$ are equidistributed. This location shift is onto with $\Omega_1 = \Omega_2 = \mathbb{R}$, but it does not match the support $(0, \infty)$ here, and there is hence no immediate version of identity (3). By virtue of Theorem 2.1, the equality of $F(X) - F(X - \alpha)$ and $F(X + \alpha) - F(X)$ in distribution when $X \sim F$ is always true, and it also leads to the equality in distribution of $U - F(F^{-1}(U) - \alpha)$ and $F(F^{-1}(U) + \alpha) - U$, for $U \sim U(0, 1)$ and cases where the support of X is equal to \mathbb{R} , which is identity (3) for $\psi(u) = F(F^{-1}(u) - \alpha)$. Alternatively, whenever the density F' is symmetric about zero, the latter is also a consequence of the identity $U \stackrel{d}{=} 1 - U$ via the relationships $F(t) = 1 - F(-t)$, $t \in \mathbb{R}$, and $F^{-1}(u) = -F^{-1}(1 - u)$, $u \in (0, 1)$, but continues to hold in general for non-symmetric F' on \mathbb{R} also.

Remark 3.1. It is interesting that one can establish the equality of the expectations of both variables of identity (2) directly as follows. We have under the conditions of Theorem 2.1, for X_1, X_2 independent with pdf $f = F'$, denoting $\Omega = \Omega_1 \cup \Omega_2$,

$$\begin{aligned} \mathbb{E}[F(X) - F(g(X))] &= \int_{\Omega} \left\{ \int_{g(x_2)}^{x_2} f(x_1) dx_1 \right\} f(x_2) dx_2 \\ &= \int_{\Omega} \left\{ \int_{x_1}^{g^{-1}(x_1)} f(x_2) dx_2 \right\} f(x_1) dx_1 \\ &= \mathbb{E}[F(g^{-1}(X)) - F(X)]. \end{aligned}$$

Since $\mathbb{E}[F(X)] = 1/2$, this also yields the relationship

$$\mathbb{E}[F(g(X)) + F(g^{-1}(X))] = 1.$$

Example 3.2. An interesting family of applications of identity (3) is given by the choices $\psi(u) = u^c$ with $c > 0$. We comment here on these equidistributional identities $U - U^c \stackrel{d}{=} U^{1/c} - U$ for $c = 2, 3, 4$, and the distributions that arise. For $c = 2$, we have for $t \in (0, 1/4)$:

$$\mathbb{P}(U - U^2 \geq t) = \mathbb{P}\left(\left|U - \frac{1}{2}\right| \leq \sqrt{\frac{1}{4} - t}\right) = \sqrt{1 - 4t}.$$

Multiplying by 4, we obtain a Beta distribution on its usual support $(0, 1)$, that is

$$4U(1 - U) \quad \text{and} \quad 4\sqrt{U}(1 - \sqrt{U}) \sim B(1, 1/2).$$

We point out that the part that says that $4U(1 - U) \sim B(1, 1/2)$ is the $\nu = 1$ special case of the following:

$$\text{if } V \sim B(\nu, \nu) \text{ then } 4V(1 - V) \sim B(\nu, 1/2).$$

When $c = 3$, multiply for convenience by $K_3 = 3\sqrt{3}/2$, and consider the distribution of $X_3 = K_3 U(1 - U^2)$ (taking values on $(0, 1)$). For $x \in (0, 1)$, the polynomial $u - u^3 - (x/K_3) = 0$ has two roots on $(0, 1)$ expressible as

$$u_i(x) = \frac{2}{\sqrt{3}} \cos \left\{ \frac{1}{3} \cos^{-1}(x) + \frac{(2i - 1)\pi}{3} \right\}, \quad i = 0, 1.$$

The common survival function of $X_3 = K_3 U(1 - U^2)$ and $K_3 U^{1/3}(1 - U^{2/3})$ is therefore given by

$$\mathbb{P}(X > x) = u_1(x) - u_2(x) = 2 \sin \left\{ \frac{1}{3} \cos^{-1}(x) \right\}, \quad x \in (0, 1),$$

with corresponding p.d.f.

$$g_3(x) = \frac{2}{3} \frac{\cos \left\{ \frac{1}{3} \cos^{-1}(x) \right\}}{\sqrt{1 - x^2}}, \quad x \in (0, 1).$$

Density g_3 is graphed in Figure 1 (dashed line) along with the very similar $B(1, 1/2)$ density $g_2(x) = 1/(2\sqrt{1 - x})$ (solid line).

When $c = 4$, multiply for convenience by $K_4 = 4^{4/3}/3$, and consider the distribution of $X_4 = K_4 U(1 - U^3)$ (taking values on $(0, 1)$). Solutions to $U - U^4 - (x/K_4) = 0$ are expressible as

$$\frac{1}{2} \left(\sqrt{2Y(x)} \pm \sqrt{\sqrt{\frac{2}{Y(x)} - 2Y(x)}} \right),$$

where

$$Y(x) = \frac{\sqrt{x}}{2^{1/3}} \cosh \left\{ \frac{1}{3} \cosh^{-1} \left(\frac{1}{x^{1/3}} \right) \right\} = \frac{1}{2^{4/3}} \left\{ (1 + \sqrt{1 - x^3})^{1/3} + (1 - \sqrt{1 - x^3})^{1/3} \right\}. \quad (7)$$

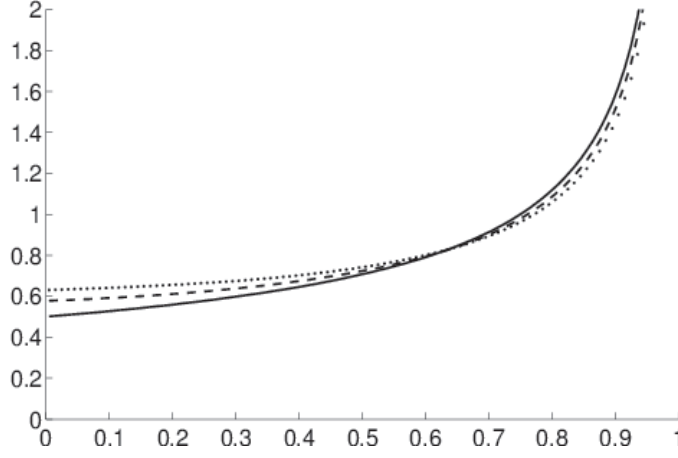


Figure 1: Densities g_2 (solid), g_3 (dashed) and g_4 (dotted)

The survival function of X_4 and of $K_4 U^{1/4}(1 - (U^{1/4})^3)$ will therefore be of the form

$$\sqrt{\sqrt{\frac{2}{Y(x)} - 2Y(x)}}.$$

The corresponding density is

$$g_4(x) = \frac{1}{2^{49/12} Y^{5/4}(x)} \frac{1 + (2Y)^{3/2}(x)}{\sqrt{1 - \sqrt{2Y}^{3/2}(x)}} \frac{1}{\sqrt{1 - x^3}} \left\{ (1 + \sqrt{1 - x^3})^{2/3} - (1 - \sqrt{1 - x^3})^{2/3} \right\}.$$

Density g_4 is also graphed in Figure 1 (dotted line), and it too is similar to the densities when $c = 2$ and $c = 3$.

Remark 3.2. Here are some further remarks concerning distributional identities of the type $U \stackrel{d}{=} \gamma(U)$ with γ non-monotone, $U \sim U(0,1)$. By composition with (3), such identities generate further equidistributional identities of the form

$$U - \psi(U) \sim \gamma(U) - \psi(\gamma(U)) \sim \psi^{-1}(\gamma(U)) - \gamma(U) \sim \psi^{-1}(U) - U. \quad (8)$$

As mentioned in the Introduction, requiring γ to be continuous and monotone leads to the unique pair of solutions $\gamma(u) = u$ and $\gamma(u) = 1 - u$. There are, however, many non-monotone functions of U that are also distributed as $U(0,1)$. One such is the symmetric triangular function

$$\gamma_1(U) = 2 \min(U, 1 - U),$$

another its complement

$$\gamma_2(U) = 1 - \gamma_1(U) = |2U - 1|.$$

We point out that $\gamma_1(U)$ is the $c = 1/2$ symmetric special case of asymmetric triangular functions of the form

$$\frac{U}{c}I(0 < U \leq c) + \frac{1-U}{1-c}I(c < U < 1),$$

which are also all $U(0,1)$ for any $0 < c < 1$. Another interesting, less obvious, non-monotone function of U that is also distributed as $U(0,1)$ is:

$$\gamma_3(U) = 2 \min(\sqrt{U}, 1 - \sqrt{U}),$$

along with its complement

$$\gamma_4(U) = |2\sqrt{U} - 1|.$$

As an illustration of (8), returning to the case $\psi(u) = u^2$, we obtain that $4(\gamma_i(U) - \gamma_i^2(U)) \stackrel{d}{=} 4(\sqrt{\gamma_i(U)} - \gamma_i(U)) \sim B(1, 1/2)$ for the above γ_i 's, yielding, for instance,

$$8 \left\{ \sqrt{U} - 2U + \max(0, 2\sqrt{U} - 1) \right\} \sim B(1, 1/2),$$

with the choice $4(\gamma_4(U) - \gamma_4^2(U))$.

The more standard way to generate a $B(1, 1/2)$ distribution is via the inverse c.d.f. method $F^{-1}(h(U))$, where $h(U) \stackrel{d}{=} U \sim U(0, 1)$ and $F^{-1}(t) = t(2-t)$ is the $B(1, 1/2)$ quantile function. For instance, the quantities

$$U(2-U); 1-U^2; |2U-1|(2-|2U-1|); |2\sqrt{U}-1|(2-|2\sqrt{U}-1|); |2\sqrt{1-U}-1|(2-|2\sqrt{1-U}-1|);$$

are all $B(1, 1/2)$ distributed and generated by $F^{-1}(h(U))$ with $h(u) = u; 1-u; \gamma_2(u); \gamma_4(u); \gamma_4(1-u)$, respectively. Finally, we point out that the inverse c.d.f. scheme $F^{-1}(h(U))$ matches the one represented by (8) for a given $\gamma(U) \sim U(0, 1)$ by the choices $h(u) = 2 \min(\gamma(u), 1 - \gamma(u)) = \gamma_1 \circ \gamma(u)$ and $h(u) = 2 \min(\sqrt{\gamma(u)}, 1 - \sqrt{\gamma(u)}) = \gamma_3 \circ \gamma(u)$, respectively for $4(\gamma(U) - \gamma^2(U))$ and $4(\sqrt{\gamma(U)} - \gamma(U))$.

4 Applications

We conclude this paper with a trio of applications which capitalize on the key identities (2) and (3) in different ways.

4.1 A reduction of variance application

Suppose that one wishes to approximately calculate

$$\mathcal{I} \equiv E\{U - \psi(U)\},$$

or equivalently $\mathbb{E}(\psi(U) = 1/2 - \mathcal{I})$, where $U \sim U(0, 1)$, $\psi : (0, 1) \rightarrow (0, 1)$ satisfies the conditions of Theorem 2.1, and $u - \psi(u)$ is non-degenerate. We know from the findings above that, inter alia, it is also the case that

$$\mathcal{I} = E\{\psi^{-1}(U) - U\}$$

and, therefore, also, that

$$\mathcal{I} = \frac{1}{2} [E\{U - \psi(U)\} + E\{\psi^{-1}(U) - U\}] = \frac{1}{2} E\{\psi^{-1}(U) - \psi(U)\}.$$

Now, define $T_1(U) = U - \psi(U)$, $T_2(U) = \psi^{-1}(U) - U$ and, in the spirit of antithetic variables, $T_M(U) = \frac{1}{2}\{T_1(U) + T_2(U)\}$ so that

$$T_M(U) = \frac{1}{2}\{\psi^{-1}(U) - \psi(U)\}.$$

It is straightforward to see that

$$V\{T_M(U)\} < V\{T_1(U)\} = V\{T_2(U)\}.$$

Thus, we would need fewer simulations to estimate \mathcal{I} by averaging simulated values of $T_M(U)$ than by directly averaging simulated values of $T_1(U)$ or $T_2(U)$. More precisely, one shows that the relative reduction in variance is equal to

$$\frac{V\{T_1(U)\} - V\{T_M(U)\}}{V\{T_1(U)\}} = \frac{1}{2} - \frac{1}{2} \rho\{T_1(U), T_2(U)\}, \quad (9)$$

with ρ the Pearson correlation coefficient.

Example 4.1. *As an illustration, and for ease of calculations, consider the toy problem case $\psi(u) = u^2$, for which $\mathcal{I} = 1/6$, with $T_1(U) = U - U^2$, $T_2(U) = \sqrt{U} - U$, $T_M(U) = \frac{1}{2}(\sqrt{U} - U^2)$. It is easy to calculate and show that*

$$V\{T_1(U)\} = V\{T_2(U)\} = 1/180 \simeq 0.00556, \text{ while } V\{T_M(U)\} = 11/252 \simeq 0.00437,$$

a reduction in variance of 3/14. Alternatively, this can be calculated from (9) with $\rho\{U - U^2, \sqrt{U} - U\} = 2/7$.

Potential gains are also available with the use of antithetic variables per se. These would combine $T_1(U)$ with $T_1(1-U)$, or $T_2(U)$ with $T_2(1-U)$ to result in consideration of

$$T_{A_1}(U) = \frac{1}{2}\{1 - \psi(U) - \psi(1-U)\}, \quad T_{A_2}(U) = \frac{1}{2}\{\psi^{-1}(U) + \psi^{-1}(1-U) - 1\}.$$

We will always have $V\{T_{A_i}(U)\} \leq V\{T_i(U)\}$, $i = 1, 2$, with equality iff T_{A_i} coincides with T_i . In fact, expression (9) for the reduction in variance applies with the correlation taken instead between $T_i(U)$ and $T_i(1-U)$.

Remark 4.1. *One may further consider $T_{MA}(U) = \frac{1}{2}\{T_M(U) + T_M(1-U)\}$ (also equal to $\frac{1}{2}\{T_{A_1}(U) + T_{A_2}(U)\}$). This will improve on $T_M(U)$, but not necessarily on T_{A_1} or T_{A_2} , since the latter do not have the same variances in general and the use of antithetic variables as above is not guaranteed to be more efficient in such cases.*

Example 4.2. *(Example 4.1 continued) Consider again $\psi(U) = U^2$ in which case $T_{A_1}(U) = T_1(U)$ and nothing is gained; which is a rare example where antithetic variables per se doesn't work directly. However, $T_{A_2}(U) = \frac{1}{2}(\sqrt{U} + \sqrt{1-U} - 1)$ and a calculation tells us that*

$$V\{T_{A_2}(U)\} = \frac{\pi}{16} - \frac{7}{36} \simeq 0.00191,$$

a more considerable reduction of over 65% on $T_1(U)$ (and over 56% on $T_M(U)$). Note that $T_{A_2}(U)$ is itself only available because we know we can alternatively work with

$T_2(U)$. Finally, we have for $\psi(U) = U^2$: $T_{MA}(U) \equiv \frac{1}{2}\{T_M(U) + T_M(1-U)\} = \frac{1}{4}(\sqrt{U} + \sqrt{1-U} - 2U^2 + 2U - 1)$, and a calculation yields $V\{T_{MA}(U)\} = \frac{\pi}{64} - \frac{23}{504} \simeq 0.00345$; a good improvement on $T_1(U)$ and $T_M(U)$, but not on $T_{A2}(U)$ which is illustrative of Remark 4.1.

4.2 An application to the discrepancy of distributions

On the one hand, the (asymmetric, unweighted) Cramér-von Mises discrepancy between distributions with c.d.f. G and F is

$$\int \{G(x) - F(x)\}^2 f(x) dx.$$

On the other hand, if $X \sim F(x)$, then defining $Y = g(X) \sim F(g^{-1}(x))$ where g is a monotone transformation involving one or two parameters is a leading way of generating more flexible distributions from simpler starting points (e.g. introducing skewness and tailweight flexibility into normal F by, say, Tukey's g - and h transformation). For appropriate g , it is also interesting to define $Z = g^{-1}(X)$, with c.d.f. $F \circ g$, to get a different family of distributions with, presumably, 'opposite' properties. For $X \in \mathbb{R}$, an interesting family of g 's are the sinh-arcsinh transformations:

$$g_{a,b}(x) = \sinh(a + b \sinh^{-1}(x)), \quad g_{a,b}^{-1}(x) = \sinh\left(-\frac{a}{b} + \frac{1}{b} \sinh^{-1}(x)\right) = g_{-a/b, 1/b}(x),$$

$a \in \mathbb{R}, b > 0$ (Jones & Pewsey, 2009)), with $g_{0,1}$ the identity transformation.

Now, observe that identity (2) (and only the second-moment equivalence) tells us that

$$\int \{F(g^{-1}(x)) - F(x)\}^2 f(x) dx = \int \{F(g(x)) - F(x)\}^2 f(x) dx.$$

That is, the Cramer-von Mises discrepancy between a 'base' distribution F and the transformed distribution using transformation g is the same as the Cramer-von Mises discrepancy between the base distribution F and the transformed distribution using transformation g^{-1} . In the sinh-arcsinh special case, the Cramer-von Mises discrepancy between base distribution F and the transformed distribution using $g_{a,b}$ is the

same as the Cramer-von Mises discrepancy between the base distribution F and the transformed distribution using $g_{-a/b,1/b}$. Our identity therefore allows us to state in what concrete way the distributions resulting from inverse transformations are ‘the same distance from’ (but ‘in different directions to’) the base distribution. To conclude, we point out that the above argument does not depend on what F is, and would hold too for any even function of $G(x) - F(x)$ in the Cramér-von Mises discrepancy such as the absolute value version which underlies a test statistic of Green & Hagezy (1975).

4.3 An application for a Gamma model predictive density estimation problem

We describe here how Theorem 2.1 implies a frequentist risk property, appearing below in (10), in a predictive density estimation framework. For the model $X|\beta \sim Ga(\alpha_1, \beta)$, $Y|\beta \sim Ga(\alpha_2, \beta)$ independently distributed, L’Moudden et al. (2017) study, for restricted parameter spaces, either $\beta \in C = (a, b)$ or $\beta \in C = [a, b]$, the Kullback-Leibler risk performance of several predictive density estimators, including Bayesian predictive density estimators, associated with a prior density π for β , given by

$$\hat{q}_\pi(y; x) = \int_{\mathbb{R}_+} q(y|\beta) \pi(\beta|x) d\beta,$$

where $q(\cdot|\beta)$ is the model density for Y and $\pi(\cdot|x)$ is the posterior density. Two such priors, along with the corresponding predictive densities, are given by $\pi_0(\beta) = \frac{1}{\beta} \mathbb{I}_{(0,\infty)}(\beta)$, $\pi_{0,C}(\beta) = \pi_0(\beta) \mathbb{I}_C(\beta)$,

$$\hat{q}_{\pi_0}(y; x) = \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1) \Gamma(\alpha_2)} \frac{1}{x} \left(\frac{y}{x}\right)^{\alpha_2-1} \left(1 + \frac{y}{x}\right)^{-(\alpha_1+\alpha_2)} \mathbb{I}_{(0,\infty)}(y),$$

and

$$\hat{q}_{\pi_0, C}(y; x) = \hat{q}_{\pi_0}(y; x) \frac{F_{\alpha_1 + \alpha_2}\left(\frac{x+y}{a}\right) - F_{\alpha_1 + \alpha_2}\left(\frac{x+y}{b}\right)}{F_{\alpha_1}\left(\frac{x}{a}\right) - F_{\alpha_1}\left(\frac{x}{b}\right)},$$

with F_α representing, hereafter, the c.d.f. of a $Ga(\alpha, 1)$ distribution. When evaluating the performance of these predictive densities under Kullback-Leibler risk

$$R_{KL}(\beta, \hat{q}) = \mathbb{E}_\beta^X \left\{ \int_{\mathbb{R}_+} q(y|\beta) \log \left(\frac{q(y|\beta)}{\hat{q}(y; X)} \right) dy \right\},$$

they show that $\hat{q}_{\pi_0, C}$ dominates \hat{q}_{π_0} for $\beta \in C$. In the process of doing so, they obtain the frequentist risk representations

$$R_{KL}(\beta, \hat{q}_{\pi_0}) = (\alpha_1 + \alpha_2)\Psi(\alpha_1 + \alpha_2) + \log \Gamma(\alpha_1) - \log \Gamma(\alpha_1 + \alpha_2) - \alpha_2 - \alpha_1\Psi(\alpha_1),$$

and

$$R_{KL}(\beta, \hat{q}_{\pi_0, C}) = R_{KL}(\beta, \hat{q}_{\pi_0}) + \phi(\alpha_1, \beta) - \phi(\alpha_1 + \alpha_2, \beta),$$

where Γ, Ψ are the gamma and the digamma functions given by $\Gamma(t) = \int_0^\infty y^{t-1} e^{-y} dy$ and $\Psi(t) = \frac{d}{dt} \log \Gamma(t)$ respectively, and with $\phi(\alpha, \beta) = \mathbb{E} [\log (F_\alpha(\beta T/a) - F_\alpha(\beta T/b))]$, the expectation being taken with respect to $T \sim Ga(\alpha, 1)$ with c.d.f. F_α . Now, consider the difference in risks $(\phi(\alpha_1, a) - \phi(\alpha_1, b)) - (\phi(\alpha_1 + \alpha_2, a) - \phi(\alpha_1 + \alpha_2, b))$ at the endpoints a, b of the parameter space $[a, b]$ with $a > 0, b < \infty$. Since

$$\phi(\alpha, a) = \mathbb{E} [\log (F_\alpha(T) - F_\alpha(aT/b))] , \phi(\alpha, b) = \mathbb{E} [\log (F_\alpha(bT/a) - F_\alpha(T))] ,$$

and since Theorem 2.1 tells us that $F_\alpha(T) - F_\alpha(aT/b) =^d F_\alpha(bT/a) - F_\alpha(T)$, for $T \sim Ga(\alpha, 1)$, by taking $g(t) = at/b, t > 0$, we obtain $\phi(\alpha, a) = \phi(\alpha, b)$ for all $\alpha > 0$ and hence

$$R_{KL}(a, \hat{q}_\pi) = R_{KL}(b, \hat{q}_\pi), \tag{10}$$

i.e., the frequentist risks at the endpoints of the parameter space $\beta = a$ and $\beta = b$ coincide. Such is a consequence of the ‘‘intriguing’’ distributional identity. We refer to L’Moudden et al. (2017) for illustrations (e.g., Figure 4) and further details.

Concluding Remarks

We have introduced, established and illustrated novel and intriguing distributional identities (2) and (3), which are, to the best of our knowledge, previously unknown. Other than an intrinsic motivation for presenting these identities, we have expanded upon applications with connections to variance reduction techniques, the discrepancy between distributions, and a risk identity in predictive density estimation. Finally, it would also be of interest to explore extensions to multivariate cases.

Acknowledgements

We are indebted to a referee and an associate editor for useful suggestions and constructive comments. We are grateful to Aziz L'Moudden for numerical evaluations leading to the Gamma application. Otherwise, we are thankful to Tomasz Kaczynski for a useful discussion of some technical aspects of Theorem 2.1. Eric Marchand's and William Strawderman's research are partially supported by the Natural Sciences and Engineering Research Council of Canada and grants from the Simons Foundation (#209035 and #418098) respectively.

References

- Arnold, B.C. (1979). Some characterizations of the Cauchy distribution. *Australian Journal of Statistics*, **21**, 166-169.
- Hegazy, Y. A. S. & Green, J. R. (1976). Powerful modified-EDF goodness-of-fit tests. *Journal of the American Statistical Association*, **71**, 204-209.
- Jones, M.C. & Pewsey, A. (2009). Sinh-arcsinh distributions. *Biometrika*, **96**, 761-780.
- Kucеровsky, D., Marchand, É., Small, R.D. (2005). On the equality in distribution of the random variables X and $g(X)$. *International Journal of Pure and Applied*

Mathematics, **23**, 93-114.

L'Moudden, A., Marchand, É., Kortbi, O. & Strawderman, W.E. (2017). On predictive density estimation for Gamma models with parametric constraints. *Journal of Statistical Planning and Inference*, **185**, 56-68.