Mathematics

AIMS Mathematics, 11(11): 36293–36355.
DOI: 10.3934/math.20261406
Received: 12 September 2026
Revised: 12 October 2026
Accepted: 15 October 2026
Published: 04 November 2026

https://www.aimspress.com/journal/Math

Research article

Cancellation-free distributional symmetry testing via reliability-inspired consecutive-system entropy scores

Ibrahim A. Nafisah[1], Tahani Alshathri[1] and Mashael A. Alshehri[2,*]

[1] Department of Statistics and Operations Research, College of Science, King Saud University, P. O. Box 2455, Riyadh 11451, Saudi Arabia
[2] Department of Quantitative Analysis, College of Business Administration, King Saud University, Riyadh 11362, Saudi Arabia

* **Correspondence:** Email: mealshehri@ksu.edu.sa; Tel: +966552022816.

Abstract: Testing distributional symmetry about an unknown center is a composite-null problem in which signed information contrasts can cancel under asymmetric alternatives. We developed a two-layer framework built on reflected probability kernels from paired consecutive reliability systems: Global fractional-entropy identities characterize symmetry, while a positive-weight complementary log-quantile-density discrepancy measures local departures without cancellation. The discrepancy identified symmetry for any strictly positive weight; the consecutive-system kernels supplied a structured, reliability-interpretable choice rather than a necessary or optimal one. We proved consistency of the moving-spacing estimator under explicit conditions and gave a compact-support sufficient class with uniform control of all complete windows, including boundary windows. For finite-sample inference, we introduced ACES-R, a distinct blockwise reflection-randomization core with finite-sample level control over the composite symmetry null, without estimating the center. Sign-paired ties generally made its complete-orbit levels conservative, and fixed-block consistency held on the stated compact-support class. A separate Monte Carlo member used a prespecified quadratic statistic. Blocking carried an information cost: ACES-R exhibited lower raw rejection rates than full-sample ACES in the reported asymmetric settings, and its finite-group guarantee did not validate the full-sample bootstrap calibration. Full-sample ACES combined the score orders 1, 3/2, and 2 in a symmetrized-bootstrap standardized maximum whose primitive bootstrap validity and joint null limit remained unproved. Paired-sample simulations quantified the blocking-power trade-off and examined Student-t, Cauchy, and skew-t settings without claiming equal-size or uniform power superiority; classical and characteristic-function benchmarks, ablations, rounding diagnostics, and two engineering

datasets mapped empirical strengths and limitations, with any heavy-tail power advantage remaining exploratory. Archived code, seeds, and replicate-level outputs support reproducibility of the paired-sample experiments.

Keywords: distributional symmetry; nonparametric symmetry testing; quantile density; consecutive-system weights; reliability systems; moving spacings; symmetrized bootstrap; randomization tests; finite-sample level control

Mathematics Subject Classification: 62B10, 62G09, 62G10

Abbreviations

ACES: aggregated consecutive-system entropy-score; CE: cumulative entropy; CLT: central limit theorem; CM: Cabilio–Masaro symmetry test; CRE: cumulative residual entropy; FGCE: fractional generalized cumulative entropy; FGCRE: fractional generalized cumulative residual entropy; FGE: fractional generalized entropy; GLD: generalized lambda distribution; HKM: Henze–Klar–Meintanis symmetry test; MGG: Miao–Gel–Gastwirth symmetry test; Mira-G: guarded implementation of Mira's symmetry test; Q–Q: quantile-quantile; ACES-R: distinct blockwise reflection-randomization inferential core

1. Introduction

Distributional symmetry about an unknown center is a basic structural assumption in location modeling, robust inference, reliability analysis, and many nonparametric procedures. In reliability and engineering applications, assessing symmetry can be important for evaluating the adequacy of location-based working models for repair-time, strength, and transformed lifetime data. Testing this assumption nevertheless remains a composite-null problem in which neither the center nor the symmetric parent distribution is specified. The difficulty is compounded by the local nature of asymmetry: Departures may be concentrated near the center, confined to one tail, or localized within a narrow band of probability levels. Procedures built from a single signed summary can therefore lose sensitivity through cancellation, because complementary contributions may integrate to zero even when substantial local asymmetry is present. In this paper, we developed a framework that deliberately separates two tasks that signed contrasts can conflate: characterizing symmetry through global entropy identities and detecting symmetry departures through a cancellation-free local discrepancy. The proposed ACES procedure is built on the second layer.

Three strands of literature motivate the development. First, information-theoretic approaches have yielded entropy-based characterizations and symmetry tests [1–3]. Related characterizations based on order statistics and information measures further develop this perspective [4–6]. Record-based characterizations and entropy tests extend the literature [7–9]. Extropy-based characterizations offer another related approach [10]. Second, classical nonparametric symmetry testing includes tests based on skewness and runs [11–13]. General nonparametric inference and sign-based procedures provide further foundations [14–16]. Hybrid and modified run tests constitute another strand [17–19]. Subsequent developments address hybrid and run procedures [20,21]. Location- and distribution-based tests include the Cabilio–Masaro and Mira approaches [22,23]. Other procedures and bootstrap-based

tests address symmetry about an unknown center [24–26]. Third, omnibus procedures based on empirical characteristic functions and related transforms provide an important complementary perspective; the test of Henze et al. [27], used later as an independent benchmark, relies on an inferential mechanism unrelated to consecutive systems or spacing entropy. These strands leave room for a construction that combines information-theoretic symmetry characterizations with a strictly nonnegative, locally interpretable testing functional and a structured family of positive probability-scale weights.

Consecutive-system lifetimes provide such a weighting family. A consecutive k-out-of-n:G system functions when at least k consecutive components operate, while the paired F system interchanges the roles of functioning and failure. Their lifetime distributions generate reflected probability-scale kernels with a natural reliability interpretation [28–30]. This places the construction naturally within statistical reliability modeling while retaining a genuinely nonparametric inferential target. As developed below, these paired kernels serve two distinct mathematical roles. They generate families of entropy equalities that characterize symmetry under explicit conditions, and they induce bounded positive weights for a local discrepancy. We emphasize from the outset that these weights are proposed as a structured and interpretable family, not as canonical or optimal choices. Their empirical value is therefore evaluated rather than assumed.

The starting point is the complementary quantile-density identity underlying symmetry. For an absolutely continuous distribution with quantile function Q and quantile density $q = Q'$, symmetry is equivalent, under the regularity conditions stated in Section 3, to

$$q(u) = q(1-u)$$

for almost every $u \in (0,1)$. A signed entropy contrast integrates a difference between complementary log-quantile densities, and such a difference can integrate to zero even when it is nonzero on a set of positive measure. Our construction therefore assigns characterization and detection to different functionals: global fractional-entropy equalities for the former, and a weighted magnitude of the complementary log-quantile-density imbalance for the latter.

The main contributions are as follows:

(i) **Probability-scale characterizations.** We derive representations of fractional generalized entropy (FGE), fractional generalized cumulative entropy (FGCE), and fractional generalized cumulative residual entropy (FGCRE) for paired consecutive-system lifetimes in terms of the parent density-quantile function. For the density-based family, we establish a converse characterization at $\alpha = 1$ for each fixed architecture parameter when equality holds across all admissible system sizes. For the cumulative family, we establish a converse for series-parallel pairs at each positive order, together with a fixed-system converse under an explicit sign restriction.

(ii) **A cancellation-free local discrepancy.** We introduce a weighted magnitude of the complementary log-quantile-density difference and show that, for a fixed architecture and any positive score order, its zero set coincides with the class of symmetric distributions under the stated regularity conditions. Because this identification property depends on positivity of the weight rather than on a unique reliability architecture, the consecutive-system family contributes structure and interpretability rather than necessity. A local perturbation expansion further separates directions that leave the discrepancy unchanged, including location, scale, and symmetry-preserving shape perturbations, from genuine asymmetric directions. The weighting scheme therefore becomes an explicit design choice that can be examined theoretically and empirically.

(iii) **The ACES test.** The ACES test estimates the local discrepancy using interior moving spacings, combines the prespecified score orders $\beta = 1, 3/2, 2$ through a standardized maximum, and

calibrates the resulting statistic using a Hodges–Lehmann-centered symmetrized bootstrap. We establish consistency of the spacing-based discrepancy estimator under explicit interior and boundary conditions and provide a verifiable compact-support sufficient class. We also show, through an occupancy argument, that the numerical zero-span floor is asymptotically inactive under continuous parent distributions when the spacing window grows at the stated rate. A broad reproducible numerical program, including the closest consecutive-system extropy predecessor [30], score, weight, and architecture ablations, larger-sample and rounding diagnostics, pooled-null calibration, and the independent characteristic-function benchmark [27], evaluates the individual design choices rather than presuming their superiority.

(iv) **Finite-sample inferential core.** In Section 5.5, we develop ACES-R, whose disjoint-block spacing vectors are jointly invariant under independent sign changes under the composite symmetry null. This supplies a test-level validity theorem distinct from the full-ACES bootstrap and a fixed-block consistency result. The complete-orbit standardized maximum and the prespecified quadratic Monte Carlo member have different calibration constructions; their guarantees are stated separately. ACES-R is a distinct finite-sample-valid inferential core based on the same spacing-asymmetry framework; it does not validate the bootstrap calibration of full ACES. Blocking loses information and gives lower raw rejection proportions on the asymmetric settings reported in Section 6.11.

The distinction from the closest consecutive-system testing precedent, the extropy construction in [30], is inferential rather than merely notational. That procedure is based on a single signed global extropy contrast, whereas ACES aggregates a nonnegative complementary log-quantile-density discrepancy. The zero-set property of the ACES discrepancy follows from positivity of the weight and the complementary symmetry identity; it does not require the consecutive-system architecture to be optimal. Accordingly, no optimality or uniform power claim is made for the proposed weighting family. The inferential scope is explicit. We derive no primitive joint null central limit theorem for the paired overlapping log-spacings, nor a method-specific theorem establishing that the implemented symmetrized bootstrap reproduces their joint null law. Existing limit theory for increasing-order spacings [31,32] provides relevant starting points. Further work on spacing-based tests and overlapping spacings [33,34] is also pertinent, but these results do not directly establish the reflected, weighted, max-combined ACES limit. Bootstrap transfer is therefore stated only as a conditional consequence of an explicit high-level joint approximation. For full ACES, finite-sample calibration and comparative power are assessed empirically rather than promoted to asymptotic theorems. This separation among proved, conditional, and empirical conclusions is deliberate. Consistent with that scope, the numerical evidence does not support uniform dominance: no weight, score order, or benchmark is uniformly best across the examined settings. Strong performance against MGG, CM, and Mira-G does not imply superiority over HKM, which has higher rejection rates for several smooth alternatives in the expanded benchmark comparisons.

The blockwise construction addresses the need for a test-level result without claiming to prove the bootstrap theory of full ACES. Reflection negates each block's signed complementary spacing vector. Under the continuous composite symmetry null, the vectors from disjoint independent blocks are jointly invariant in distribution under independent block reflections, and the unknown center cancels from the spacings. Complete enumeration therefore yields finite-sample level control that incorporates the conservative effect of sign-paired ties. The Monte Carlo construction in Theorem 5.2(iii) uses a prespecified statistic, independent uniform sign draws, and the plus-one correction; its finite-sample level control follows from conditional exchangeability under this null group invariance, and it does not enlarge the underlying sign orbit. Full ACES remains a separate, empirically calibrated procedure that uses the complete sample. The nominal level is generally not attained by complete

enumeration: Sign-paired orbit ranks impose a block-count-dependent upper rejection bound, and additional ties can make it more conservative.

Throughout, symmetry refers to the measurement scale on which the observations are analyzed. Applying the procedure after a logarithmic transformation addresses log-symmetry instead. In Section 2, we introduce the consecutive-system and fractional-information notation. In Section 3, we establish the global characterizations. In Section 4, we develop the cancellation-free local discrepancy and its perturbation geometry. In Section 5, we develop estimation and calibration for full ACES and introduce the finite-sample-valid ACES-R core. In Sections 6 and 7, we present the numerical study and data applications, respectively. In Section 8, we summarize the main conclusions and limitations. Auxiliary results and computational details are collected in the appendices and Appendix E.

2. Preliminaries and paired-system information measures

In this section, we fix the notation and objects used throughout the paper. In Section 2.1, we recall consecutive k-out-of-n systems and their paired survival and distribution representations, which generate the reflected probability-scale kernels that play two distinct roles later: As the integrand weights in the global entropy characterizations of Section 3, and as the positive weighting family for the local discrepancy of Section 4. In Section 2.2, we introduce the fractional information measures, fractional generalized entropy and its cumulative counterparts, together with the single notational convention adopted for all orders. The parameters that organize the paper are defined here and kept separate throughout: The architecture index (n, r) of the paired system and the fractional order α used for characterization, while the local score order β in Section 4 is reserved for the testing layer.

2.1. Consecutive systems and reflected probability kernels

A consecutive k-out-of-n: G system functions when at least k consecutive components operate; a consecutive k-out-of-n: F system fails when at least k consecutive components fail. Let the component lifetimes be independent and identically distributed, with distribution function F, survival function $\overline{F} = 1 - F$, and support S_X. For a linear system with $2k \geq n$, the system survival/distribution representations used here are [35]

$$\overline{F}_{k|n:G}(x) = (n-k+1)\overline{F}^{k}(x) - (n-k)\overline{F}^{k+1}(x), x \in S_X \tag{1}$$

and

$$F_{k|n:F}(x) = (n-k+1)F^{k}(x) - (n-k)F^{k+1}(x), x \in S_X. \tag{2}$$

The bar in (1) denotes the G-system survival function, whereas (2) is the F-system distribution function. Interchanging F and $\overline{F}$ converts one representation into the other, giving the paired reflection structure that underlies everything that follows. Below, $r = n - k$ is a nonnegative integer and $n \geq max\{1,2r\}$; n indexes the system architecture, while N, introduced in Section 5, denotes the statistical sample size.

2.2. Fractional information measures

Write $Q(u) = F^{-1}(u)$ and $q(u) = Q'(u) = 1/f\{Q(u)\}$ wherever defined; q is the quantile density. The Shannon differential entropy is $H(X) = -\int_{S_X} f(x)\, logf(x)\, dx$ [36]. Cumulative

residual entropy [37] and cumulative entropy [38] replace the density weighting by the survival and distribution functions, respectively. Using one convention throughout, set

$$\xi_\alpha(t) = t[-logt]^\alpha, \quad c(\alpha) = \{\Gamma(\alpha+1)\}^{-1}, \quad \alpha > 0, \tag{3}$$

with $\xi_\alpha(0) = \xi_\alpha(1) = 0$. The fractional generalized cumulative entropy of [39] and its residual counterpart are

$$CE_\alpha(X) = c(\alpha)\int_{S_X} F(x)\,[-logF(x)]^\alpha\,dx = c(\alpha)\int_0^1 \xi_\alpha\,(u)q(u)\,du, \tag{4}$$

$$E_\alpha(X) = c(\alpha)\int_{S_X} \overline{F}(x)\left[-log\overline{F}(x)\right]^\alpha\,dx = c(\alpha)\int_0^1 \xi_\alpha\,(1-u)q(u)\,du. \tag{5}$$

At $\alpha = 1$, these reduce to CE and CRE. Related fractional cumulative measures are studied in [40–42], and positive integer orders recover the generalized CRE construction in [43]. Integrals over S_X cover nonnegative lifetimes and other interval supports, whenever finite.
The density-based FGE of [1] is

$$FH_\alpha(X) = c(\alpha)\int_{S_X} f(x)\,[-logf(x)]^\alpha\,dx. \tag{6}$$

Every FGE used below is required to be finite and real valued. Integer orders avoid fractional powers of negative numbers; for noninteger orders, we require the relevant negative log density to be nonnegative almost everywhere. This condition applies to each system density when its FGE is evaluated, not merely to the parent density. Finally, the local score order β introduced in Section 4 is separate from α: α indexes the characterization layer, whereas β acts on a nonnegative magnitude in the testing layer.

3. Global entropy characterizations of symmetry

In this section, we establish the characterization layer of the framework: Symmetry is identified through global entropy equalities between paired consecutive-system lifetimes. The mechanism is uniform throughout. Under the probability integral transformation, the paired system densities decompose as a reflected probability-scale kernel times the parent density-quantile function; symmetry equates the two sides of this decomposition, and completeness of the kernel family turns a family of entropy equalities into a necessary and sufficient condition. In Section 3.1, we develop the density-based (FGE) results, where a converse is obtained at order $\alpha = 1$ for every fixed architecture parameter; in Section 3.2, we develop the cumulative (FGCE/FGCRE) results, where a converse holds for series-parallel pairs at every positive order; and in Section 3.3, we show how a fixed-system converse can still be obtained under an explicit sign restriction. A running caveat is stated once here and used throughout: These characterizations concern signed global contrasts, so they motivate, but do not by themselves provide, the cancellation-free testing functional in Section 4.

3.1. Density-quantile symmetry and FGE representations

Let X be an absolutely continuous random variable with the cumulative distribution function F, the probability density function f, the interval support S_X, and the quantile function F^{-1}. Assume that $f(x) > 0$ almost everywhere on the interior of S_X and that F^{-1} is absolutely continuous on every compact subinterval of $(0,1)$. Define the density-quantile transform by $f\left(F^{-1}(u)\right)$. The basic symmetry condition is

$$f(F^{-1}(u)) = f(F^{-1}(1-u)), \quad 0 < u < 1, \tag{7}$$

almost everywhere.

Lemma 3.1. Under the preceding assumptions, X is symmetric about some point μ if and only if (7) holds almost everywhere on $(0,1)$.

Proof. If X is symmetric about μ, then

$$F(\mu - x) + F(\mu + x) = 1$$

whenever both terms are defined. Equivalently,

$$F^{-1}(u) + F^{-1}(1-u) = 2\mu, \quad 0 < u < 1.$$

Differentiation, together with

$$\frac{d}{du}F^{-1}(u) = \frac{1}{f\left(F^{-1}(u)\right)},$$

gives (7) almost everywhere. Conversely, if (7) holds almost everywhere, then

$$\frac{d}{du}\{F^{-1}(u) + F^{-1}(1-u)\} = 0$$

almost everywhere. Absolute continuity on compact subintervals therefore implies that $F^{-1}(u) + F^{-1}(1-u) = 2\mu$ for some constant μ. This quantile identity is equivalent to symmetry about μ.

For each admissible system size, let $X_1, \dots, X_n$ be independent and identically distributed copies of X. Fix a nonnegative integer r and, for every integer $n \geq max\{1, 2r\}$, consider the linear consecutive $(n-r)$-out-of-n:G and $(n-r)$-out-of-n: F systems. Their lifetimes are denoted by $T_{n-r|n:G}$ and $T_{n-r|n:F}$, respectively. Differentiating the distributional representations in Eqs (1) and (2) gives

$$f_{n-r|n:G}(x) = f(x)\left\{a_{r,n}\overline{F}^{n-r-1}(x) - b_{r,n}\overline{F}^{n-r}(x)\right\}, \quad x \in S_X \tag{8}$$

and

$$f_{n-r|n:F}(x) = f(x)\left\{a_{r,n}F^{n-r-1}(x) - b_{r,n}F^{n-r}(x)\right\}, \quad x \in S_X, \tag{9}$$

where

$$a_{r,n} = (n-r)(r+1), \quad b_{r,n} = (n-r+1)r.$$

Set $U_j = F\left(X_j\right)$, $j = 1, \dots, n$, so that $U_1, \dots, U_n$ are independent uniform random variables on $(0,1)$.

Define

$$U_{n-r|n:G} = F(T_{n-r|n:G}), \quad U_{n-r|n:F} = F(T_{n-r|n:F}).$$

Their density functions are

$$\rho_{n-r|n:G}(u) = a_{r,n}(1-u)^{n-r-1} - b_{r,n}(1-u)^{n-r}, \quad 0 < u < 1 \tag{10}$$

and

$$\rho_{n-r|n:F}(u) = a_{r,n}u^{n-r-1} - b_{r,n}u^{n-r}, \quad 0 < u < 1. \tag{11}$$

The probability-scale kernels satisfy the following reflection identities.

Lemma 3.2. For every $0 < u < 1$,

$$\rho_{n-r|n:G}(u) = \rho_{n-r|n:F}(1-u).$$

Consequently,

$$B_{G,n-r}(u) = B_{F,n-r}(1-u),$$

where

$$B_{G,n-r}(u) = -log\rho_{n-r|n:G}(u), \qquad B_{F,n-r}(u) = -log\rho_{n-r|n:F}(u).$$

Proof. The first identity follows immediately from (10) and (11); the second follows by taking negative logarithms.

Define

$$C(u) = -logf\big(F^{-1}(u)\big), \qquad 0 < u < 1.$$

Lemma 3.3. The equality $C(u) = C(1-u)$ holds almost everywhere on $(0,1)$ if and only if (7) holds almost everywhere.

Proof. This follows from the one-to-one character of the logarithm on $(0,\infty)$.

For each fixed configuration, assume the displayed system entropies are finite and real valued. For a noninteger α, the expressions raised to that power are required to be nonnegative almost everywhere.

Theorem 3.1. For $\alpha > 0$, the fractional generalized entropy of the G-system lifetime is

$$FH_\alpha\big(T_{n-r|n:G}\big) = c(\alpha)\int_0^1 \rho_{n-r|n:G}\,(u)\big[-log\{\rho_{n-r|n:G}(u)f\big(F^{-1}(u)\big)\}\big]^\alpha\,du, \tag{12}$$

or, equivalently,

$$FH_\alpha\big(T_{n-r|n:G}\big) = c(\alpha)\int_0^1 \rho_{n-r|n:G}\,(u)\big[B_{G,n-r}(u) + C(u)\big]^\alpha\,du. \tag{13}$$

Similarly,

$$FH_\alpha\big(T_{n-r|n:F}\big) = c(\alpha)\int_0^1 \rho_{n-r|n:F}\,(u)\big[-log\{\rho_{n-r|n:F}(u)f\big(F^{-1}(u)\big)\}\big]^\alpha\,du, \tag{14}$$

or, equivalently,

$$FH_\alpha\big(T_{n-r|n:F}\big) = c(\alpha)\int_0^1 \rho_{n-r|n:F}\,(u)\big[B_{F,n-r}(u) + C(u)\big]^\alpha\,du. \tag{15}$$

Proof. From (8),

$$f_{n-r|n:G}\big(F^{-1}(u)\big) = \rho_{n-r|n:G}(u)f\big(F^{-1}(u)\big).$$

Substitution into the definition of fractional generalized entropy, followed by the change of variable $u = F(x)$, yields (12). Since

$$-log\{\rho_{n-r|n:G}(u)f\big(F^{-1}(u)\big)\} = B_{G,n-r}(u) + C(u),$$

Eq (13) follows. The proof of (14) and (15) is identical.

The following lemma supplies the completeness property required for the converse implication and applies to every fixed $r \geq 0$. Its proof uses a weighted-moment completeness argument related to classical moment-characterization methods [44,45].

Lemma 3.4. Fix an integer $r \geq 0$, and let $\Delta \in L^1(0,1)$. If

$$\int_0^1 \rho_{n-r|n:F}(u)\Delta(u)\,du = 0, \quad n \geq max\{1,2r\}, \tag{16}$$

then $\Delta(u) = 0$ almost everywhere on $(0,1)$.

Proof. For $r = 0$, (11) gives $\rho_{n|n:F}(u) = nu^{n-1}$. Hence, (16) implies

$$\int_0^1 u^j \Delta(u)\,du = 0, \quad j = 0,1,\ldots,$$

and the conclusion follows from the density of algebraic polynomials in $C[0,1]$.

Now let $r \geq 1$, put $m = n - r$, and define

$$M_j = \int_0^1 u^j \Delta(u)\,du.$$

Because $n \geq 2r$, one has $m \geq r$, and (11) transforms (16) into

$$m(r+1)M_{m-1} - r(m+1)M_m = 0, \quad m \geq r.$$

Thus,

$$M_j = M_{r-1}\frac{r}{j+1}\left(\frac{r+1}{r}\right)^{j-r+1}, \quad j \geq r.$$

By dominated convergence, $M_j \to 0$ as $j \to \infty$. The preceding expression forces $M_{r-1} = 0$, so $M_j = 0$ for all $j \geq r-1$. The function

$$h(u) = u^{r-1}\Delta(u)$$

belongs to $L^1(0,1)$ and defines a finite signed measure $h(u)\,du$. All its polynomial moments vanish.

Uniform density of polynomials in $C[0,1]$ therefore implies that this measure is zero. Since $u^{r-1} > 0$ on $(0,1)$, $\Delta = 0$ almost everywhere.

At $\alpha = 1$, FGE is Shannon differential entropy, and the preceding completeness lemma yields the following converse.

Theorem 3.2. Fix an integer $r \geq 0$. Under the assumptions of Lemma 3.1, suppose that

$$\int_0^1 |C(u)|\,du = \int_{S_X} |logf(x)|f(x)dx < \infty,$$

and that the required order-one system entropies are finite. Then F is symmetric if and only if

$$FH_1\left(T_{n-r|n:G}\right) = FH_1\left(T_{n-r|n:F}\right), \quad n \geq max\{1,2r\}, \tag{17}$$

for every integer $n \geq max\{1,2r\}$.

Proof. Suppose first that F is symmetric. By Lemmas 3.1 and 3.3, $C(u) = C(1-u)$ almost everywhere. Using (13), Lemma 3.2, and the change of variable $z = 1 - u$,

$$FH_1(T_{n-r|n:G}) = c(1)\int_0^1 \rho_{n-r|n:F}\,(z)\{B_{F,n-r}(z) + C(1-z)\}\,dz.$$

Replacing $C(1-z)$ by $C(z)$ and invoking (15) proves (17).

Conversely, assume (17). Subtracting the two representations after the same change of variable gives

$$\int_0^1 \rho_{n-r|n:F}\,(u)\{C(1-u) - C(u)\}\,du = 0, \quad n \geq max\{1,2r\}.$$

Set $\Delta(u) = C(1-u) - C(u)$. The assumed absolute integrability ensures that this function belongs to the space required in Lemma 3.4. Hence, Lemma 3.4 implies $\Delta = 0$ almost everywhere. Lemma 3.3 then gives (7), and Lemma 3.1 establishes symmetry of F.

For $r = 0$, the order-one minimum–maximum entropy criterion in Theorem 3.2 corresponds to the Shannon-order special case of the criterion stated in [1, Theorem 3.1(ii)]. This proof gives an explicit completeness argument for each fixed consecutive-system parameter r under the stated integrability assumptions. It does not infer a converse for general fractional orders from that linear completeness argument.

Remark 3.1. For other admissible fractional orders, symmetry implies equality of the paired FGE functionals. The converse involves the nonlinear difference

$$\left[B_{F,n-r}(u) + C(1-u)\right]^{\alpha} - \left[B_{F,n-r}(u) + C(u)\right]^{\alpha},$$

and does not follow from Lemma 3.4. No general converse for $\alpha \neq 1$ is claimed. Even at order one, equality for one fixed system size gives only a single signed integral constraint; the converse in Theorem 3.2 uses all admissible sizes. Appendix A treats the order-two signed statistic as an auxiliary diagnostic, not as a consequence of this converse.

Proposition 3.1. Let $Y = aX + b$, where $a > 0$ and $b \in R$, and let C_X and C_Y denote the corresponding negative log density-quantile transforms. Then

$$C_Y(u) = C_X(u) + loga, \quad 0 < u < 1. \tag{18}$$

Moreover, if the superscripts X and Y identify the parent scale used to construct the paired system lifetimes, then

$$FH_1\left(n-r|n: G^T_{(Y)}\right) - FH_1\left(n-r|n: F^T_{(Y)}\right) = FH_1\left(n-r|n: G^T_{(X)}\right) - FH_1\left(n-r|n: F^T_{(X)}\right). \tag{19}$$

Hence, the equality in Theorem 3.2 is invariant under positive affine transformations.

Proof. Since

$$Y^F_{-1}(u) = aX^F_{-1}(u) + b$$

and

$$f_Y(y) = a^{-1}f_X\left((y-b)/a\right), \qquad f_Y\left(Y^F_{-1}(u)\right) = a^{-1}f_X\left(X^F_{-1}(u)\right),$$

which proves (18). Since $\alpha = 1$ and $c(1) = 1$, substitution into (13) and (15) adds the same quantity $loga$ to both system entropies because each kernel integrates to one. Their difference is therefore unchanged, proving (19).

The order-one difference is thus invariant to positive affine changes of scale. A signed FGE contrast at $\alpha = 2$ generally lacks this raw invariance; the local discrepancy in Section 4 has its own, stronger cancellation of the scale factor.

3.2. Cumulative-entropy representations and converse

Retain the support and quantile regularity of Lemma 3.1 and assume all displayed cumulative entropies are finite. For $\alpha > 0$, use

$$\xi_\alpha(t) = t[-logt]^\alpha$$

as in (3). For the uniform-scale system variables defined above, the G-system survival function and the F-system distribution function are

$$\overline{Q}_{n-r|n:G}(u) = (r+1)(1-u)^{n-r} - r(1-u)^{n-r+1}, \quad 0 < u < 1 \tag{20}$$

and

$$Q_{n-r|n:F}(u) = (r+1)u^{n-r} - ru^{n-r+1}, \quad 0 < u < 1. \tag{21}$$

These identities follow from (1) and (2) after the probability integral transformation $u = F(x)$. They also satisfy the complementary relation

$$\overline{Q}_{n-r|n:G}(u) = Q_{n-r|n:F}(1-u).$$

Using $u = F(x)$ and $dx = du/f\left(F^{-1}(u)\right)$, the FGCRE of the G-system lifetime admits the representation

$$E_\alpha(T_{n-r|n:G}) = c(\alpha)\int_0^1 \frac{\xi_\alpha\left(\bar{Q}_{n-r|n:G}(u)\right)}{f\left(F^{-1}(u)\right)}\,du, \quad \alpha > 0. \tag{22}$$

Similarly, the FGCE of the paired F-system lifetime is

$$CE_\alpha(T_{n-r|n:F}) = c(\alpha)\int_0^1 \frac{\xi_\alpha\left(\bar{Q}_{n-r|n:F}(u)\right)}{f\left(F^{-1}(u)\right)}\,du, \quad \alpha > 0. \tag{23}$$

The series-parallel case gives a weighted-moment family with a direct completeness argument.

Lemma 3.5. Let Δ be measurable on $(0,1)$ and suppose that $z(-logz)^\alpha\Delta(z) \in L^1(0,1)$ for some fixed $\alpha > 0$. If

$$\int_0^1 z^n\,(-logz)^\alpha\Delta(z)\,dz = 0, \quad n = 1,2,\ldots, \tag{24}$$

then $\Delta(z) = 0$ for almost every $z \in (0,1)$.

Proof. Define

$$h(z) = z(-logz)^\alpha\Delta(z).$$

Equation (24) is equivalent to

$$\int_0^1 z^m\,h(z)\,dz = 0, \quad m = 0,1,2,\ldots.$$

Thus, h is orthogonal to every algebraic polynomial. Since polynomials are uniformly dense in $C[0,1]$ and $h \in L^1(0,1)$, the finite signed measure $h(z)\,dz$ annihilates every continuous function and must therefore be the zero measure. Hence, $h = 0$ almost everywhere. Because $z(-logz)^{\alpha} > 0$ on $(0,1)$, it follows that $\Delta = 0$ almost everywhere.

Theorem 3.3. Under the assumptions of Lemma 3.1, let the component lifetimes be independent copies of X. Fix $\alpha > 0$, set $r = 0$, and assume the functionals in (22) and (23) are finite for all $n \geq 1$, with

$$z(-logz)^{\alpha}\left[\frac{1}{f\left(F^{-1}(1-z)\right)} - \frac{1}{f\left(F^{-1}(z)\right)}\right] \in L^1(0,1).$$

Then F is symmetric if and only if

$$E_{\alpha}(T_{n|n:G}) = CE_{\alpha}(T_{n|n:F}), \quad n = 1,2, \ldots.$$

Proof. Suppose first that F is symmetric. By Lemma 3.1,

$$f\left(F^{-1}(u)\right) = f\left(F^{-1}(1-u)\right)$$

for almost every $u \in (0,1)$. When $r = 0$, Eqs (20) and (21) reduce to

$$\overline{Q}_{n|n:G}(u) = (1-u)^n, \quad Q_{n|n:F}(u) = u^n.$$

Substituting these identities into (22), using $z = 1-u$, and then applying density-quantile symmetry gives

$$E_{\alpha}(T_{n|n:G}) = c(\alpha)\int_0^1 \frac{\xi_{\alpha}(z^n)}{f\left(F^{-1}(z)\right)}\,dz = CE_{\alpha}(T_{n|n:F}),$$

which proves necessity.

Conversely, suppose that the displayed equality holds for every $n \geq 1$. From Eqs (22) and (23), and the substitution $z = 1-u$, one obtains

$$\int_0^1 \xi_{\alpha}(z^n)\left[\frac{1}{f\left(F^{-1}(1-z)\right)} - \frac{1}{f\left(F^{-1}(z)\right)}\right]dz = 0, \quad n = 1,2, \ldots. \tag{25}$$

Since

$$\xi_{\alpha}(z^n) = n^{\alpha}z^n(-logz)^{\alpha}$$

and $n^{\alpha} > 0$, (25) is equivalent to

$$\int_0^1 z^n\,(-logz)^{\alpha}\left[\frac{1}{f\left(F^{-1}(1-z)\right)} - \frac{1}{f\left(F^{-1}(z)\right)}\right]dz = 0, \quad n = 1,2, \ldots.$$

Lemma 3.5 therefore yields

$$\frac{1}{f\left(F^{-1}(1-z)\right)} = \frac{1}{f\left(F^{-1}(z)\right)}$$

almost everywhere. Positivity of f gives the density-quantile symmetry condition in Lemma 3.1, and hence F is symmetric.

Corollary 3.1. Fix $r \geq 0$ and $\alpha > 0$. If F is symmetric, then

$$E_\alpha(T_{n-r|n:G}) = CE_\alpha(T_{n-r|n:F})$$

for every $n \geq max\{1,2r\}$. This follows from the reflection identity and the substitution $z = 1-u$. For $r \geq 1$, the relevant kernels are

$$\xi_\alpha\big((r+1)u^{n-r} - ru^{n-r+1}\big).$$

These kernels are not the separable weighted-moment family used in Theorem 3.3. The corollary gives the symmetry-to-equality implication only; its general-architecture converse is not established.

3.3. Fixed-system identification under a sign restriction

Theorem 3.3 requires equality across all system sizes. A fixed-system characterization can nevertheless be obtained on a restricted class of density-quantile functions.

Let C denote the class of densities for which either

$$f\big(F^{-1}(u)\big) \geq f\big(F^{-1}(1-u)\big), \quad 0 < u < 1/2,$$

or the reverse inequality holds throughout $(0,1/2)$. For fixed $n \geq 1$ and $\alpha > 0$, define

$$H_{n,\alpha}(z) = z^n(-logz)^\alpha - (1-z)^n[-log(1-z)]^\alpha, \quad 0 < z < 1/2. \tag{26}$$

Condition 3.1. The function $H_{n,\alpha}$ has a fixed nonzero sign for almost every $z \in (0,1/2)$. A transparent sufficient choice is $\alpha = n$. Indeed,

$$-zlogz > -(1-z)log(1-z)$$

for $0 < z < 1/2$, and hence

$$H_{n,n}(z) = [-zlogz]^n - [-(1-z)log(1-z)]^n > 0, \quad 0 < z < 1/2. \tag{27}$$

Proposition 3.2. Suppose that $f \in C$ and that Condition 3.1 holds. If, for the fixed pair (n,α),

$$E_\alpha(T_{n|n:G}) = CE_\alpha(T_{n|n:F}),$$

then F is symmetric.

Proof. For $r = 0$, the assumed equality is equivalent to

$$\int_0^1 z^n\,(-logz)^\alpha \left[\frac{1}{f\big(F^{-1}(1-z)\big)} - \frac{1}{f\big(F^{-1}(z)\big)}\right] dz = 0.$$

Splitting the integral at $1/2$ and applying $z \mapsto 1-z$ to the second part gives

$$\int_0^{1/2} H_{n,\alpha}\,(z) \left[\frac{1}{f\big(F^{-1}(1-z)\big)} - \frac{1}{f\big(F^{-1}(z)\big)}\right] dz = 0. \tag{28}$$

By Condition 3.1, the first factor has a fixed nonzero sign almost everywhere. Since $f > 0$ and $f \in C$, the reciprocal difference in the second factor also has a fixed sign. The integrand therefore has one sign, and its integral can vanish only if the reciprocal difference is zero almost everywhere on $(0,1/2)$. Reflection then gives

$$f\big(F^{-1}(u)\big) = f\big(F^{-1}(1-u)\big)$$

almost everywhere on $(0,1)$, so Lemma 3.1 implies that F is symmetric.

Theorem 3.3 uses equality across all series–parallel system sizes. Proposition 3.2 instead uses a fixed size together with a shape restriction and a kernel sign condition. Neither result supplies a general fixed-system converse for a signed contrast without these qualifications. Appendix B gives illustrative parent distributions.

4. A cancellation-free local entropy-score discrepancy

The characterizations in Section 3 identify symmetry through families of signed equalities, which may lose sensitivity as test statistics because cancellation across complementary probability levels can mask a locally nonzero imbalance. In this section, we construct the object that the ACES test is built on, a strictly nonnegative discrepancy that aggregates the *magnitude* of the local imbalance rather than its signed value. The construction has two ingredients: a positive weight on the lower half of the probability scale, supplied here by the reflected consecutive-system kernels, and a score order that governs sensitivity to discrepancy magnitude. Two facts are established before any sampling is introduced: The discrepancy vanishes if and only if the distribution is symmetric (Theorem 4.1), and it is exactly invariant to positive affine transformations (Proposition 4.1). A perturbation analysis (Proposition 4.2) then shows precisely which directions the discrepancy is blind to, location, scale, and symmetry-preserving shape changes, and makes the weight a testable design object rather than a purely interpretive one.

The global identities in Section 3 identify symmetry through families of equalities or sign restrictions. For a fixed-architecture test, we instead aggregate the magnitude of the local imbalance. Write

$$n, r_F^{\rho} = \rho_{n-r|n:F} \text{ and } n, r_G^{\rho} = \rho_{n-r|n:G},$$

and define

$$w_{n,r}(u) = \frac{n,r_F^{\rho}(u)+n,r_G^{\rho}(u)}{\int_0^{1/2}[n,r_F^{\rho}(v)+n,r_G^{\rho}(v)]dv}, \quad 0 < u < \frac{1}{2}. \tag{29}$$

Reflection of the two kernels (Lemma 3.2) makes their sum positive on $(0, 1/2)$, and the denominator normalizes the weight to integrate to one on this interval. The score order $\beta > 0$ is independent of the FGE order α. The population discrepancy is

$$D_{\beta;n,r}(F) = \int_0^{1/2} w_{n,r}\,(u)\,|C(u) - C(1-u)|^{\beta}\,du. \tag{30}$$

Equivalently, since $C = logq$,

$$D_{\beta;n,r}(F) = \int_0^{1/2} w_{n,r}\,(u)\,\left|log\,\frac{q(u)}{q(1-u)}\right|^{\beta}\,du. \tag{31}$$

Theorem 4.1. *Under the interval-support and absolute-continuity assumptions of Lemma 3.1, fix an integer* $r \geq 0$*, an integer* $n \geq max\{1,2r\}$*, and* $\beta > 0$*. If* $D_{\beta;n,r}(F) < \infty$*, then*

$$D_{\beta;n,r}(F) = 0 \quad \Leftrightarrow \quad F\ is\ symmetric\ about\ some\ point. \tag{32}$$

Proof. If F is symmetric, Lemma 3.1 yields $C(u) = C(1-u)$ almost everywhere, and hence Eq (30) is zero. Conversely, the integrand in Eq (30) is nonnegative and $w_{n,r}(u) > 0$ almost everywhere on $(0,1/2)$. A zero integral therefore implies $C(u) = C(1-u)$ almost everywhere on $(0, 1/2)$, and hence on $(0,1)$. Lemma 3.1 completes the proof.

Remark 4.1. The result holds for a single architecture and any positive score order. Its proof uses only the positivity of the weight: Any strictly positive integrable weight, after normalization, has the same zero-set property on the stated distribution class. Consecutive systems are therefore retained as a closed-form, architecture-indexed weighting family, not as a necessary condition for identification and not as a guarantee of optimal power. The scientific role of consecutive systems in ACES is *weighting, not identification*: They generate a reflected polynomial family with a reliability interpretation, while uniform and other positive weights remain legitimate competitors. In Section 6.7, we evaluate empirically whether the system-induced choices add finite-sample value.

Proposition 4.1. *Let* $Y = a + bX$*, where* $a \in R$ *and* $b > 0$*. Then*

$$D_{\beta;n,r}(F_Y) = D_{\beta;n,r}(F_X). \tag{33}$$

Proof. Since $q_Y(u) = bq_X(u)$,

$$logq_Y(u) - logq_Y(1-u) = logq_X(u) - logq_X(1-u).$$

Substitution into (31) proves the result.

The discrepancy is thus a unit-free shape measure: For a positive affine transformation $Y = a + bX$, factor b cancels in the ratio of paired quantile densities. The same cancellation applies to paired sample spacings, so no sample-variance standardization is required for ACES, a practical contrast with the location-contrast benchmarks of Section 6.

A plot of $C(u) - C(1-u)$ identifies the probability regions contributing to the discrepancy: Small u compares complementary tail regions, while values near 1/2 compare the shoulders and center. These are descriptive local diagnostics, not simultaneous confidence statements. The global signed statistic of Appendix A is a different diagnostic and is not used to compute ACES.

The baseline architecture. The implementation fixes $(n,r) = (3,1)$, corresponding to paired consecutive 2-out-of-3 systems, for which

$$3{,}1_F^{\rho}(u) = 4u - 3u^2, \quad 3{,}1_G^{\rho}(u) = 4(1-u) - 3(1-u)^2, \tag{34}$$

and direct normalization gives

$$w_{3,1}(u) = 1 + 6u - 6u^2, \quad 0 < u < \tfrac{1}{2}, \quad \int_0^{1/2} w_{3,1}(u)\,du = 1. \tag{35}$$

This weight is positive and bounded and does not vanish at $u = 0$; the score order β controls sensitivity to discrepancy magnitude, while the architecture controls probability-scale emphasis. Figure 1 shows the paired kernels and a population score comparison. The $(3,1)$ choice is a prespecified baseline architecture, not a canonical one: Its empirical role is examined against uniform and alternative system-induced weights in Section 6.7. The three score orders used by ACES are specified in Section 5.3.

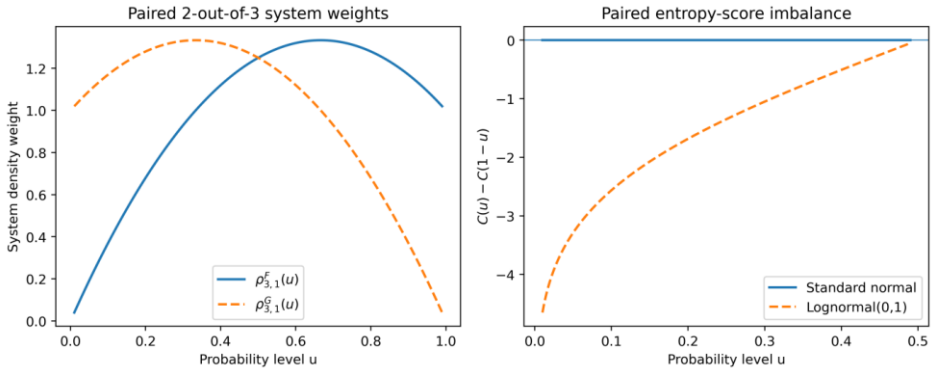


Figure 1. Paired consecutive 2-out-of-3 system densities and complementary entropy-score imbalance for normal and lognormal distributions. The local imbalance is zero under normal symmetry.

4.1. Local sensitivity and architecture geometry

The zero-set result in Theorem 4.1 does not by itself select a preferred positive weight. To make the role of the weight explicit, consider a smooth perturbation of a symmetric baseline quantile function. Let Q_0 have positive symmetric quantile density q_0, and let

$$Q_\varepsilon(u) = Q_0(u) + \varepsilon A(u),$$

where A is locally absolutely continuous and $q_\varepsilon = q_0 + \varepsilon A'$ is positive along the admissible perturbations specified in Proposition 4.2. These perturbations may be two-sided or one-sided; the proposition states the required positivity and domination conditions. Define the local asymmetry direction

$$\Delta_A(u) = \frac{A'(u)}{q_0(u)} - \frac{A'(1-u)}{q_0(1-u)}, \qquad 0 < u < \frac{1}{2}.$$

Proposition 4.2. (Local perturbation expansion) Let Q_0 be a symmetric quantile function with positive quantile density q_0, let A be locally absolutely continuous, and put $Q_\varepsilon = Q_0 + \varepsilon A$ and $\rho = A'/q_0$. Let w be a normalized strictly positive weight on $(0,1/2)$. Take a set E of nonzero admissible perturbations accumulating at zero, for which Q_ε is a quantile function. Suppose there exist a constant $c_0 > 0$ and a measurable function $G \geq 0$ such that, almost everywhere on $(0,1/2)$,

$$1 + \varepsilon\rho(u) \geq c_0, \qquad 1 + \varepsilon\rho(1-u) \geq c_0 \quad (\varepsilon \in E),$$

$$|\rho(u)| + |\rho(1-u)| \leq G(u), \qquad \int_0^{1/2} w\,(u) G(u)^\beta \, du < \infty.$$

Then, with $\Delta_A(u) = \rho(u) - \rho(1-u)$,

$$D_{\beta,w}(F_\varepsilon) = |\varepsilon|^\beta \int_0^{1/2} w\,(u) |\Delta_A(u)|^\beta \, du + o\big(|\varepsilon|^\beta\big), \quad \varepsilon \to 0,\ \varepsilon \in E.$$

Proof. Symmetry cancels the two baseline logarithms. Define

$$h_\varepsilon(u) = \{log\big(1 + \varepsilon\rho(u)\big) - log\big(1 + \varepsilon\rho(1-u)\big)\}/\varepsilon.$$

Pointwise, $h_\varepsilon \to \Delta_A$. The mean value theorem, applied between 1 and $1+\varepsilon\rho$, gives $|log(1+\varepsilon\rho)/\varepsilon| \le |\rho|/min(1,c_0)$. Thus, $|h_\varepsilon|^\beta \le G^\beta/min(1,c_0)^\beta$, an integrable dominating function. Dominated convergence proves the expansion for every $\beta > 0$, without a uniform Taylor remainder for an unbounded G.

A bounded ρ is a transparent two-sided sufficient condition: restrict $|\varepsilon| \parallel \rho \parallel_\infty \le 1/2$ and take $c_0 = 1/2$. The main statement also permits one-sided perturbations with an unbounded positive derivative ratio, provided the displayed lower bound and weighted integrability hold. In contrast, positivity for every sufficiently small positive and negative ε forces ρ to be essentially bounded; weighted integrability alone does not imply $|\varepsilon|G < 1/2$ uniformly.

Remark 4.2. (Scope) This is a population-functional expansion, not a stochastic local-power, detection-boundary, or Pitman-efficiency result.

The expansion separates nuisance and asymmetric directions. A pure location perturbation has $A' = 0$; a scale perturbation has A proportional to Q_0 minus its center, so that A'/q_0 is constant; and an odd perturbation about $u = 1/2$ preserves symmetry and satisfies $A'(u) = A'(1-u)$. Each gives $\Delta_A = 0$. Genuine asymmetric directions are precisely those for which Δ_A is nonzero on a set of positive measure.

Remark 4.3. (*Architecture is not one-to-one with weight shape*) For the normalized lower-half weights used here, direct algebra gives

$$w_{3,1}(u) = 1 + 6u - 6u^2, \qquad w_{4,1}(u) = 2, \qquad w_{4,2}(u) = 12u(1-u), \qquad w_{5,2}(u) = w_{3,1}(u).$$

Thus, different consecutive architectures can induce the same local weighting measure; the architecture label should not be interpreted as a unique statistical design parameter. In Section 6.7, we compare distinct induced weights and score orders directly.

5. Estimation and calibration

In this section, we separate estimation from calibration. In Sections 5.1–5.2, we define the moving-spacing estimator and its consistency conditions, including the compact-support sufficient class. In Sections 5.3 and 5.4, we retain the full-sample bootstrap construction and its explicitly conditional theoretical status. Then, in Section 5.5, we develop the separate blockwise ACES-R core with finite-group level control and fixed-block consistency. In Section 5.6, we summarize both algorithms. The new finite-group theorem does not validate the full-ACES bootstrap.

For independent complete observations, the symmetry hypothesis is

$$H_0: \exists \mu \in R \; such \; that \; F(\mu + x) = 1 - F(\mu - x) \; for \; all \; admissible \; x, \; H_1: no \; such \; \mu \; exists. \quad (36)$$

5.1. Interior spacings and recorded ties

Let $X_{(1)} \le \cdots \le X_{(N)}$ be the order statistics. Choose an integer window $m = m_N$ satisfying

$$m_N \to \infty, \quad \frac{m_N}{N} \to 0. \quad (37)$$

For full-sample ACES, the deterministic rule below is used on its admissible domain $N \ge 11$. The distinct small-block extension for ACES-R is stated as (51) in Section 5.5.

$$m_N = min\left\{\left\lfloor 2\sqrt{N} \right\rfloor, \left\lfloor \frac{(N-3)}{4} \right\rfloor\right\}, \quad m_N \ge 2. \quad (38)$$

Only interior paired positions are retained. For $i = m_N + 1, \dots, \lfloor N/2 \rfloor$, let $u_i = i/(N+1)$ and define

$$\hat{q}(u_i) = \frac{N}{2m_N}\{X_{(i+m_N)} - X_{(i-m_N)}\}. \tag{39}$$

The reflected estimate is $\hat{q}(1-u_i) = \hat{q}(u_{N+1-i})$. The retained index set $I_N = \{m_N + 1, \dots, \lfloor N/2 \rfloor\}$ uses complete windows and avoids endpoint replication. For the adopted rule, $m_{20} = 4$, $m_{50} = 11$, and $m_{100} = 20$, leaving 6, 14, and 30 lower-half evaluation positions, respectively. These are evaluation positions, not counts of observations retained: Each spacing uses observations at both window endpoints.

For rounded observations, let δ_N be the smallest strictly positive adjacent gap. If a local span is zero, its corresponding quantile-density estimate is replaced by $N\delta_N/(4m_N)$, equivalent to a span floor of $\delta_N/2$. The distinction between a span and its density-scaled estimate is important. A constant sample has no positive gap and does not support this calculation. For nonconstant samples, the estimator is

$$\widehat{D}_\beta = \frac{\sum_{i=m_N+1}^{\lfloor N/2 \rfloor} w_{3,1}(u_i)|log\hat{q}(u_i) - log\hat{q}(1-u_i)|^\beta}{\sum_{i=m_N+1}^{\lfloor N/2 \rfloor} w_{3,1}(u_i)}. \tag{40}$$

The floor is a deterministic numerical convention, not a model for rounded or discrete data. The implementation records its activation and rejects constant original samples; constant bootstrap samples are redrawn, with counts reported, and nonpositive bootstrap scales trigger an error. Under a continuous parent law, the original sample has no exact ties, but empirical resampling can still create ties, so original-sample continuity alone does not justify the bootstrap treatment. In (40), the retained-grid weights are normalized by their discrete sum, making the estimator a weighted mean over the retained index set.

Both positive spans and the resolution floor scale with the observations, so positive affine transformations leave the sample log-spacing discrepancy unchanged. The theoretical framework does not cover intrinsically discrete or heavily rounded distributions.

5.2. Consistency of the discrepancy estimator

Assumption 5.1. Retain the regularity of Lemma 3.1. For each selected score order $\beta > 0$, assume:

(i) f is positive and continuous in the interior of its support, and q is continuously differentiable on every compact subinterval of $(0,1)$.

(ii) $m_N/logN \to \infty$ and $m_N/N \to 0$.

(iii) The population discrepancy in (31) is finite.

(iv) For every $\eta > 0$, the weighted empirical log-spacing contributions in shrinking boundary bands satisfy

$$\lim_{\varepsilon \downarrow 0} \limsup_{N \to \infty} P\left\{\frac{\sum_{i:\, u_i < \varepsilon} w_{3,1}(u_i)|log\hat{q}(u_i) - log\hat{q}(1-u_i)|^\beta}{\sum_i w_{3,1}(u_i)} > \eta\right\} = 0. \tag{41}$$

Both sums in (41) are over I_N, with the numerator further restricted by $u_i < \varepsilon$. This is an explicit high-level boundary-negligibility condition, not a consequence of smoothness on fixed interior intervals alone. A suitable integrable envelope together with uniform integrability of the empirical boundary contributions is one sufficient route to it.

Theorem 5.1. Under Assumption 5.1,

$$\widehat{D}_\beta \xrightarrow{p} D_{\beta;3,1}(F), \quad \beta > 0. \tag{42}$$

Proof. Fix $0 < \varepsilon < 1/2$. On the compact interval $[\varepsilon, 1-\varepsilon]$, the probability-integral transform reduces the retained spans to moving 2m-spacings of uniform order statistics. Under conditions (i) and (ii), standard exponential/Beta representations, concentration, and a union bound over O(N) moving windows give uniform relative convergence of these spacings; positivity and uniform continuity of q on this compact interval then transfer the result to uniform consistency of $\hat{q}(u)$ for $q(u)$ [31,32]. Related density and quantile-density estimation references are [46,47]. Because q is positive there, the same holds for $log\hat{q}$. Hence, the portion of (40) over $[\varepsilon, 1/2]$ converges in probability to the corresponding weighted integral by a Riemann-sum argument. The population contribution over $(0, \varepsilon)$ vanishes as $\varepsilon \downarrow 0$ by condition (iii), and the empirical contribution vanishes in probability by condition (iv). The resolution adjustment is asymptotically inactive for an absolutely continuous distribution because the probability of tied original observations is zero. Letting $N \to \infty$ and then $\varepsilon \downarrow 0$ proves (42).

Consequently, under a fixed asymmetric alternative in the stated class, $\widehat{D}_\beta$ converges to a positive limit by Theorem 4.1. This is consistency of the discrepancy estimator; consistency of the bootstrap test additionally requires control of its critical values.

Remark 5.1. (Status of Theorem 5.1) Theorem 5.1 is a conditional consistency result: Interior consistency under (i) and (ii) is combined with the high-level boundary-negligibility condition (iv). It does not, by itself, establish global consistency from primitive conditions alone. An unconditional result on an explicit, checkable class is provided by Corollary 5.1 below, whose proof is completed in this revision using Lemma 5.1.

Lemma 5.1. (Complete-window uniform relative convergence) Let $U_{(1)} \le \cdots \le U_{(N)}$ be the order statistics of N i.i.d. $Uniform\,(0,1)$ random variables and let $m = m_N$ satisfy $1 \le m \le \lfloor N/2 \rfloor$, $m \to \infty$, $m/N \to 0$, and $m/logN \to \infty$. Then the following uniform convergence holds:

$$\max_{m+1 \le i \le N-m} \left| \frac{N}{2m} \{U_{(i+m)} - U_{(i-m)}\} - 1 \right| \xrightarrow{p} 0. \tag{43}$$

The maximum ranges over all complete windows, including $i = O(m)$ (i.e., $u_i \to 0$).

Proof. Use the representation $U_{(k)} = S_k / S_{N+1}$, where $S_k = E_1 + \cdots + E_k$ and $E_1, \ldots, E_{N+1}$ are i.i.d. $Exp(1)$. For each admissible i, $G_i := S_{i+m} - S_{i-m} \sim Gamma(2m, 1)$ marginally, with the same law at every admissible index. The window sums need not be independent. Sub-exponential concentration for Gamma sums [48] gives

$$P(|G_i - 2m| > 2m\delta) \le 2exp(-m\delta^2/4)$$

for $0 < \delta < 1/2$, and a union bound over at most N windows yields

$$P(max_i |G_i/(2m) - 1| > \delta) \le 2Nexp(-m\delta^2/4) \to 0$$

because $m/logN \to \infty$. Since $S_{N+1}/N \to 1$ in probability and S_{N+1} is a single common factor, Slutsky's lemma gives (43). Overlapping windows do not invalidate the union bound because only marginal tails are used; the per-window fluctuation is index-free, so the boundary is no harder than the interior.

Corollary 5.1. (*Compact-support sufficient class*) Suppose the support is a compact interval, f is continuously differentiable there, and $0 < c \le f(x) \le C < \infty$. If $m = m_N$ satisfies $m/logN \to \infty$ and $m/N \to 0$, then the interior moving-spacing quantile-density estimates are uniformly consistent over the retained grid and the boundary-negligibility requirement in Assumption 5.1(iv) is automatic. Consequently, Theorem 5.1 holds without imposing (iv) as a separate high-level assumption on this class.

Proof. Write $X_{(k)} = Q\big(U_{(k)}\big)$. The assumptions imply that $q = 1/f(Q)$ extends continuously to $[0,1]$ and satisfies $1/C \le q \le 1/c$. By the mean value theorem,

$$\hat{q}(u_i) = q(\xi_i)\frac{N}{2m}\{U_{(i+m)} - U_{(i-m)}\}, \quad \xi_i \in \big[U_{(i-m)}, U_{(i+m)}\big].$$

To locate these intervals, put

$$\delta_N = max_{1\le k\le N}\big|U_{(k)} - k/(N+1)\big|.$$

Uniform convergence of the uniform empirical distribution gives $\delta_N \to_p 0$ [49]. In particular,

$$\max_i|\xi_i - u_i| \le \delta_N + \frac{m}{N+1} \xrightarrow{p} 0.$$

This step uses the locations of the order statistics as well as their spacing lengths; small window lengths alone do not establish it. Uniform continuity of q and Lemma 5.1 now yield

$$\max_{i\in I_N}|\hat{q}(u_i) - q(u_i)| \xrightarrow{p} 0, \qquad \max_{i\in I_N}|log\hat{q}(u_i)| \le M \quad with\ probability\ tending\ to\ one. \tag{44}$$

The same bounds hold for the reflected estimates. Here, M is a fixed sufficiently large constant. On this event, the boundary ratio in (41) is at most

$$\frac{(\varepsilon(N+1)+1)\, w_{max}(2M)^{\beta}}{(\lfloor N/2\rfloor - m)w_{min}},$$

where $1 \le w_{3,1} \le 5/2$. Its limit superior is bounded by a constant times ε. First, let $N \to \infty$, then $\varepsilon \downarrow 0$, to obtain (41). The normalized grid sums converge to the population integral by a Riemann-sum argument. Exact original-sample ties have probability zero.

5.3. Score aggregation and null resampling

The prespecified score set is

$$A = \{1, 3/2, 2\}, \quad \widehat{D} = \big(\widehat{D}_1, \widehat{D}_{3/2}, \widehat{D}_2\big)^{\top} \tag{45}$$

The order $\beta = 1$ aggregates absolute imbalance linearly, $\beta = 2$ places relatively greater weight on larger departures, and $\beta = 3/2$ provides an intermediate response. This finite set limits the multiplicity of the max-combined statistic; it is not claimed optimal. The architecture and order set are fixed before the observed sample is examined.

Let $\hat{\mu}_{HL}$ be the Hodges–Lehmann pseudomedian, defined as the median of the Walsh averages $\big(X_i + X_j\big)/2, i \le j$. Set $R_i = |X_i - \hat{\mu}_{HL}|$. A bootstrap sample is generated by independently drawing $1_*^R, \dots, N_*^R$ with replacement from $\{R_1, \dots, R_N\}$, independently generating Rademacher signs $i_*^S \in \{-1,1\}$, and setting

$$i_*^X = \hat{\mu}_{HL} + i_*^S i_*^R. \tag{46}$$

The center is not needed to evaluate the location-invariant discrepancy, but it is needed to construct the residual magnitudes from which the symmetric null law is estimated; adding the estimated center back leaves the discrepancy unchanged. Symmetrized empirical resampling of this type is used in symmetry testing [25,26], but validity for the spacing statistic requires the separate approximation stated in Section 5.4.

For bootstrap replicates $b = 1, \dots, B$, let $\beta_{*(b)}^{\widehat{D}}$ denote the component for $\beta \in A$, with bootstrap mean $\beta_*^{\overline{D}}$ and standard deviation β_*^S. Define

$$A_N = \max_{\beta \in A} \frac{\widehat{D}_\beta - \beta_*^{\overline{D}}}{\beta_*^S}, \tag{47}$$

and

$$N_{*(b)}^A = \max_{\beta \in A} \frac{\beta_{*(b)}^{\widehat{D}} - \beta_*^{\overline{D}}}{\beta_*^S}. \tag{48}$$

The Monte Carlo p-value is

$$\hat{p} = \frac{1 + \sum_{b=1}^B I\left\{N_{*(b)}^A \geq A_N\right\}}{B+1}. \tag{49}$$

The standard deviations in (47) and (48) must be positive. A degenerate bootstrap component requires investigation rather than division by zero. The $+1$ correction makes the smallest reported value $1/(B+1)$; it does not make the composite-null bootstrap an exact finite-sample randomization test.

Lemma 5.2. (Conditional occupancy and zero-span flooring) Conditional on the observations, suppose a bootstrap draw has at most $2N$ possible values and each atom has probability at most $1/N$. If N_*^M is the largest multiplicity in N independent bootstrap draws, then $N_*^M = O_{P^*}(logN/loglogN)$. Consequently, $m/logN \to \infty$ implies $P^*(N_*^M \geq 2m+1) \to 0$ and rules out any zero complete $2m$-span with conditional probability tending to one.

Proof. Each count is bounded in law by a binomial count with mean one. For $k \geq 1$, a binomial tail bound and a union bound give $P^*(N_*^M \geq k) \leq 2N(e/k)^k$. Taking $k = ClogN/loglogN$ with fixed $C > 1$ proves the rate, consistently with [50]. A zero $2m$-span requires at least $2m+1$ equal observations. For distinct original observations, ordinary empirical resampling has atom masses $1/N$. In the implemented symmetrized resampling, at most two observations share any nonzero absolute residual about a given center; after sign splitting, each nonzero atom has mass at most $1/N$, and a zero atom also has mass at most $1/N$. Thus, the lemma covers that resampling law as well. This controls numerical degeneracy only, not bootstrap distributional validity.

5.4. Conditional calibration logic under a high-level joint approximation

In this subsection, we record a conditional consequence only; it is not a bootstrap-validity theorem for ACES, and nothing in the ACES-R construction of Section 5.5 depends on it.

The following statement records what a valid joint bootstrap approximation would imply for the max-combined statistic. It does not prove that approximation for ACES from primitive conditions on F and the spacing window.

Condition 5.2. (High-level joint approximation) Under the symmetry null, let $a_{N,\beta} = E^*\left(\beta_*^{\hat{D}}\right)$ and $N, \beta_2^s = Var^*\left(\beta_*^{\hat{D}}\right)$ be the finite ideal conditional bootstrap moments, with $s_{N,\beta} > 0$. Define the standardized vectors

$$Z_N = \left(\frac{\hat{D}_\beta - a_{N,\beta}}{s_{N,\beta}}\right)_{\beta \in A}, \qquad N_*^Z = \left(\frac{\beta_*^{\hat{D}} - a_{N,\beta}}{s_{N,\beta}}\right)_{\beta \in A}. \tag{50}$$

Assume $Z_N \Rightarrow Z$ and, conditionally on the data, $L^*(N_*^Z) \Rightarrow L(Z)$ in probability. Assume also that $M = max_{\beta \in A} Z_\beta$ has a continuous distribution function. These are assumptions about the standardized joint laws. The raw scales $s_{N,\beta}$ may tend to zero and are not assumed bounded away from zero.

Conditional transfer consequence. Under Condition 5.2, the ideal conditional bootstrap distribution of $N_*^M = max_\beta N, \beta_*^Z$ consistently estimates the null distribution of $M_N = max_\beta Z_{N,\beta}$. The ideal-bootstrap rejection probability therefore converges to the nominal level at continuity levels of the limiting law.

Justification. The maximum is a continuous map on the finite-dimensional component space. The observed and conditional bootstrap convergences in Condition 5.2 therefore transfer to M_N and N_*^M. Continuity of the limiting distribution gives uniform convergence of the corresponding distribution functions and the resulting bootstrap quantile calibration; see the general framework in [49].
This consequence is not a primitive bootstrap-validity theorem for ACES: Condition 5.2 already assumes the difficult joint approximation, and the statement records only the finite-dimensional continuous-mapping consequence once that approximation is available.

Theorem 5.1 does not establish Condition 5.2. In particular, the component with $\beta = 1$ is nonsmooth at the null discrepancy, and a common root-N Gaussian expansion is not assumed. The kernel-based bootstrap argument in [51] concerns a different smoothing construction and cannot be transferred to intermediate spacings without proof. Finite-sample performance is therefore assessed empirically in Section 6, separately from this conditional transfer statement.

Corollary 5.1 and Lemma 5.2 strengthen the estimator and numerical-regularity side of the argument, but they do not identify the joint fluctuation law of the paired moving log-spacings. The relevant windows overlap strongly, the dependence range grows with m, and estimation of the symmetry center enters the resampling law. Hall [31,33] and more recent overlapping-spacing theory [34] provide closely related limit tools, but a source theorem covering the reflected log-ratio vector used by ACES has not been established here. Primitive validity of the implemented Hodges–Lehmann symmetrized bootstrap therefore remains open outside the conditional transfer statement.

The transfer statement concerns ideal conditional moments and distributions. For estimated standardization in (47) and (48), the centering error must be negligible relative to the ideal scale, estimated-to-ideal scale ratios must converge to one, and Monte Carlo approximation must be controlled as the number of bootstrap draws increases. A fixed finite bootstrap count is not covered by this ideal statement.

5.5. ACES-R: a finite-sample-valid reflection-randomization core

The bootstrap calibration of full ACES in Section 5.3 is empirically supported, but its primitive validity remains open (Section 5.4). We therefore introduce ACES-R, a blockwise reflection-randomization test that targets the same complementary log-quantile-density asymmetry functional (31) under the same cancellation-free weighting philosophy, but deliberately employs a different finite-

sample construction: disjoint blocks in place of full-sample overlapping windows, chosen so that finite-sample level control follows from group invariance (Theorem 5.2). This distinct inferential core does not validate the bootstrap calibration of full ACES. Its blocking incurs an information cost: Table 11 shows lower raw rejection rates than full ACES on every displayed asymmetric alternative, though these comparisons are between procedures whose effective finite-sample null rejection probabilities may differ.

Throughout the manuscript, K denotes the prespecified number of ACES-R blocks; B is reserved for full-ACES bootstrap and benchmark resampling draws; and L denotes the number of independent sign draws in Eq (55).

Construction and finite-block window. Prespecify a block count $K \geq 2$ and let $b = \lfloor N/K \rfloor$. Use the first Kb observations in their original order to form K disjoint blocks, and sort only within each block. The last $N - Kb < K$ observations are unused. Partitioning is not selected from the data or from resulting p-values. To distinguish the original full-sample rule (38) from its small-block extension, for $b \geq 8$ define

$$m_b = max\left\{2, min\left(\lfloor 2\sqrt{b} \rfloor, \left\lfloor \frac{b-3}{4} \right\rfloor\right)\right\}. \tag{51}$$

Rule (51) agrees with (38) for $b \geq 11$; its floor applies only to smaller blocks. It satisfies the same asymptotic rates. More generally, the theoretical construction permits any integer m with $1 \leq m < \lfloor b/2 \rfloor$, so that the retained grid is nonempty. Put $I_b = \{m+1, \ldots, \lfloor b/2 \rfloor\}$, $u_i = i/(b+1)$, and, for block r,

$$\hat{q}_r(u_i) = \frac{b}{2m}\{X_{r,(i+m)} - X_{r,(i-m)}\}, \quad Z_{r,i} = log\hat{q}_r(u_i) - log\hat{q}_r(1-u_i).$$

The reflected index is $i^* = b + 1 - i$, so $u_{i^*} = 1 - u_i$ exactly, and both windows are complete. Let

$$\pi_i = \frac{w_{3,1}(u_i)}{\sum_{i \in I_b} w_{3,1}(u_j)}$$

and $G = \{-1,1\}^K$. For $s \in G$, define

$$\overline{Z}_s(u_i) = K^{-1} \sum_{r=1}^{K} s_r Z_{r,i}, \qquad D_{\beta,R}(s) = K^{-1} \sum_{i \in I_b} \pi_i \left|\overline{Z}_s(u_i)\right|^{\beta}.$$

For $\beta \in A = \{1, 3/2, 2\}$, compute the complete-orbit mean $\mu_{\beta,R}$ and standard deviation $\sigma_{\beta,R}$ of $D_{\beta,R}(s)$. These are unchanged by multiplying the orbit by any fixed sign vector. The statistic at configuration s is

$$M_R(s) = \max_{\beta \in A:\, \sigma_{\beta,R} > 0} \frac{D_{\beta,R}(s) - \mu_{\beta,R}}{\sigma_{\beta,R}}. \tag{52}$$

If every orbit variance is zero, set $M_R(s) = 0$ for all s and return $p = 1$. The observed configuration is $s = 1$. Complete enumeration in Tables 11 and 12 uses this maximum. The raw quadratic member $T_Q(s) = D_{2,R}(s)$ is a separate, prespecified choice used for the Monte Carlo alternative below; it is not the statistic behind those tables.

Lemma 5.3. (Blockwise reflection invariance) Under the continuous composite symmetry null, the block vectors satisfy

$$(Z_1, \ldots, Z_K) \overset{d}{=} (s_1 Z_1, \ldots, s_K Z_K) \quad for\ every\ s \in G.$$

Proof. If F is symmetric about θ, replacing all observations of one block by $2\theta - X$ preserves their joint law, independently of the other blocks. Reflected order statistics are $X'_{r,(k)} = 2\theta - X_{r,(b+1-k)}$. Subtraction cancels θ and gives $\hat{q}'_r(u_i) = \hat{q}_r(u_{i^*})$; hence, $Z'_r = -Z_r$. The blocks are disjoint and independent, so composing these reflections proves the joint invariance. Continuity ensures positive spacings almost surely. The index-reversal identity also holds in the presence of ties; continuity is used here to make the logarithms well defined without an additional discrete-data convention.

Theorem 5.2. (Finite-sample validity and paired-orbit resolution) Define

$$p_F = 2^{-K} \sum_{S \in G} 1\{M_R(s) \geq M_R(1)\}. \tag{53}$$

(i) Under the composite null, $P_F(p_F \leq a) \leq a$ for every $a \in [0,1]$, uniformly over the unknown center and the continuous symmetric parent law.

(ii) Since $M_R(s) = M_R(-s)$, all orbit-tail counts are even. Thus, p_F is on the lattice with step $h_K = 2^{1-K}$ and $p_F \geq h_K$. A deterministic nominal level-a rule has rejection probability at most

$$a_K = h_K \left\lfloor \frac{a}{h_K} \right\rfloor. \tag{54}$$

Equality with a_K follows if, almost surely, the 2^{K-1} global-sign equivalence classes have distinct statistic values. Additional ties can make the rule more conservative; continuity alone is not asserted to rule out such ties.

(iii) For a statistic fixed before drawing transformations, in particular $T_Q = D_{2,R}$, let $S_1, \ldots, S_L$ be independent uniform draws with replacement from G, independent of the data, and put

$$p_{MC} = \frac{1 + \sum_{l=1}^{L} 1\{T_Q(S_l) \geq T_Q(1)\}}{L+1}. \tag{55}$$

Then $P_F(p_{MC} \leq a) \leq a$ for every $a \in [0,1]$. No off-grid additive error term is needed. The maximum in (52) is not standardized using those same Monte Carlo draws.

(iv) For complete-orbit, nonrandomized rejection at level a to be possible, $h_K \leq a$ is necessary. At $a = 0.05$, K=10 gives a_k=50/1024=0.048828125, whereas K=6 gives a_k=2/64=0.03125. These are upper attainable level bounds, not universal attained sizes. For design purposes, we require adequate complete-orbit support before choosing enumeration or Monte Carlo. A finite Monte Carlo p-value can fall below the complete-orbit floor through sampling fluctuation, so that floor is not a deterministic lower bound on (55). Such fluctuation is not an enlargement of the orbit or a reliable remedy for coarse support.

Proof. Conditional on the orbit, invariance makes the observed point uniform over its orbit. Group counts include stabilizer multiplicities. The proportion of points whose upper-tail rank is at most a is at most a, proving (i); the deterministic sign pairing gives (ii) and (54). For (iii), conditional on the orbit, the observed transformed point and the L independently uniformly transformed points are identically distributed and independent. Their fixed-statistic values are therefore exchangeable. The inclusive upper rank divided by $L+1$ is super-uniform. Equivalently, it is bounded above in rejection probability by $\lfloor a(L+1) \rfloor / (L+1) \leq a$ [52–54]. The leading 1 in (55) counts the identity observation, while identity draws among the S_l are retained. Finally, conditional on the data, the random

exceedance count for (55) is binomial with success probability equal to the full-orbit tail probability of T_Q, which explains the distinction in (iv).

Theorem 5.3. (Consistency of complete enumeration at fixed K) Fix $K \geq 2$ with $2^{1-K} \leq a$. Let $b = \lfloor N/K \rfloor \to \infty$, $m_b/logb \to \infty$ and $m_b/b \to 0$. Suppose F belongs to the compact-support class of Corollary 5.1 and is asymmetric. Then the complete-enumeration test based on (52) and (53) has power tending to one. This is not a consistency claim for a fixed finite Monte Carlo budget.

Proof. Put

$$\Delta(u) = logq(u) - logq(1-u)$$

and

$$D_\beta = \int_0^{1/2} w_{3,1}(u)|\Delta(u)|^\beta du > 0.$$

Per-block uniform convergence and a finite union over K blocks yield, uniformly on the retained grid and the finite sign group,

$$\overline{Z}_s(u_i) = a_s\Delta(u_i) + o_p(1), \quad a_s = K^{-1}\sum_{r=1}^{K} s_r, \quad D_{\beta,R}(s) = |a_s|^\beta D_\beta + o_p(1).$$

Let c_β and $v_\beta > 0$ be the uniform-sign mean and variance of $|a_s|^\beta$. The limiting standardized component is $\left(|a_s|^\beta - c_\beta\right)/\sqrt{v_\beta}$. Every such component, and hence their maximum, is strictly increasing in $|a_s|$. Its maximum occurs only at $s = \pm 1$. The finite orbit therefore has a positive limiting gap between these two configurations and the others. Their equality is deterministic even before taking limits. With probability tending to one, the observed upper-tail count is exactly two; hence,

$$p_F = 2^{1-K} \leq a,$$

and the null hypothesis is rejected.

Remark 5.4. (Limits of the consistency claim) Theorem 5.3 covers a fixed, prespecified number of blocks with growing block size. A separate regime with $K \to \infty$ would require explicit triangular-array moment or uniform-integrability control and a specified statistic. For example, the conditional sign mean of the raw quadratic is $K^{-2}\sum_r \sum_i \pi_i r, i_2^Z$; this identity is not a vanishing-mean statement for the orbit-standardized maximum. No additional growing-K consistency theorem is claimed here.

Remark 5.5. (Grid and sample-use trade-off) For fixed K, the retained grid becomes dense in $(0, 1/2)$ under the rates of Theorem 5.3, so the theorem detects every asymmetric distribution in its stated compact-support class. At finite sample size, small blocks omit the most extreme probability levels; finite-sample level control does not remove this power limitation. Discarding fewer than K observations has vanishing fractional cost for fixed K, but its finite-sample count is reported with every design.

5.6. Algorithm and computational considerations

Symmetry tests can be found in Algorithms 1 and 2.

Algorithm 1. ACES symmetry test.	
1	Sort the nonconstant sample, compute m_N from (38), and construct the paired spacings over I_N.
2	Apply the recorded-tie convention when necessary and compute (40) for $\beta \in A$.
3	Compute the Hodges–Lehmann center and the absolute residuals.
4	Generate B samples from (46) and repeat the same discrepancy calculation, window rule, and tie convention in each sample.
5	Check the bootstrap scales, calculate (47)–(49), and reject at nominal level γ when $\hat{p} \le \gamma$.

Sorting costs O(N log N) per sample and the three discrepancies require linear additional work. The direct Hodges–Lehmann calculation uses O(N²) Walsh averages. The code enforces $N \ge 11$ for rule Eq (38), while the numerical performance evidence starts at $N = 20$; the computational threshold is not a calibration recommendation.

Algorithm 2. ACES-R symmetry test.	
1	Prespecify K and the target level a. Form blocks from the original observation order, using $b = \lfloor N/K \rfloor$ and the first Kb observations; record the unused remainder.
2	Set $m = m_b$ by (51), verify that the complementary grid is nonempty, and sort within each block to compute $Z_{r,i}$ and π_i.
3	Check $2^{1-K} \le a$ for the recommended design. If full enumeration is feasible, compute the complete-orbit moments and the maximum (52). Otherwise prespecify T_Q and use independent uniform sign draws with replacement in (55).
4	Compute upper-tail counts using the inclusive comparison $T \ge T_{obs}$. The implementation evaluates one representative per deterministic global-sign pair and restores its multiplicity exactly; observed and transformed statistics follow the same arithmetic path. No numerical tolerance is subtracted from the observed statistic. A zero-variance complete orbit returns $p = 1$.
5	Reject when the corresponding p-value is at most a. Do not reinterpret the Monte Carlo theorem as a proof for full-ACES symmetrized bootstrap calibration. For complete enumeration, a is nominal and (54) is its upper attainable rejection bound, not a universally attained size.

At $a = 0.05$, the reported defaults are $(N, K, b, m) = (50,6,8,2), (100,10,10,2), (200,10,20,4)$. The unused remainders are 2,0,0, and the upper attainable complete-orbit levels are $0.03125, 0.048828125,$ and 0.048828125 respectively. These settings are prespecified working choices, not optimal designs. The reference code exposes the block count corresponding to K explicitly; it does not search block counts to minimize the observed p-value. Appendix D maps the manuscript notation to the unchanged historical code arguments.

6. Simulation study

In this section, we evaluate ACES and its competitors across a deliberately extensive program: raw size and power at $N = 20, 50, 100$ (Sections 6.2 and 6.3); null-only sensitivity to the spacing-

window rule (Section 6.4); independent calibration and reference-specific recalibration checks (Section 6.5); a direct comparison with the closest consecutive-system extropy predecessor (Section 6.6); common-random-number ablations of the score orders, weights, and architectures (Section 6.7); larger-sample and rounding diagnostics (Section 6.8); a balanced portfolio calibration that places all methods on a common target (Section 6.9); an independent characteristic-function benchmark together with a Cauchy stress experiment (Section 6.10); a paired-sample comparison of ACES-R and full ACES (Section 6.11); and a targeted tail-path experiment (Section 6.12). Three interpretive rules are stated once and applied throughout: The tables evaluate the implemented finite-B procedures rather than any ideal-bootstrap limit; boldface marks the largest entry in a row and is descriptive, not a significance statement; and no experiment is read as evidence of uniform dominance, since the Monte Carlo uncertainty derived in Section 6.1 applies to every comparison. The ACES-R comparisons concern distinct procedures with potentially different effective finite-sample null rejection probabilities; they are raw rejection comparisons, not equal-size power comparisons.

6.1. Design and comparison scope

The reported experiments use $R = 1000$ Monte Carlo samples per configuration and $B = 399$ bootstrap draws per sample at nominal level 0.05. Sample sizes are $N = 20{,}50{,}100$. The tables evaluate the implemented finite-B procedures, not the ideal-bootstrap limit.

The original replay is supplemented by five independent diagnostic blocks using disjoint seed streams: a direct comparison with the consecutive-system extropy test [30]; score-order, weight, and architecture ablations; $N = 200$ and 500 checks plus rounding/tie stress tests; a balanced non-oracle portfolio calibration; and an independent comparison with the empirical-characteristic-function symmetry test of Henze, Klar, and Meintanis (HKM) [27], accompanied by a Cauchy stress null. These additional experiments are diagnostic and do not retrospectively redefine the original Tables 1–5.

The null panel contains six continuous symmetric distributions with substantially different shapes: standard normal, logistic, Laplace, Student t with three degrees of freedom, standardized uniform, and an equal-weight mixture of $N(-1.5, 1)$ and $N(1.5, 1)$. The panel therefore includes light and heavy tails, bounded support, a sharp center, and bimodality.

Ten asymmetric alternatives are considered: skew-normal distributions with shape parameters 3 and 6; lognormal distributions with log-scale standard deviations 0.5 and 1; gamma distributions with shapes 2 and 4 and unit scale; Beta (2,5); the contaminated normal mixture $0.90N(0,1) + 0.10N(3,1)$; Weibull with shape 1.5 and unit scale; and its left-skewed reflection. The complete generator definitions and original seed streams are included in the computational archive. These models also probe the boundary of the stated theory. The Laplace law is included as a deliberate finite-sample stress case because its quantile density is not continuously differentiable at the median and therefore does not satisfy all of Assumption 5.1(i). The asymmetric panel contains diffuse and tail-driven departures, while the reflected Weibull checks behavior under the opposite direction of skewness.

The original benchmarks are MGG [24], Cabilio–Masaro (CM) [22], and a numerically guarded version of Mira [23], denoted Mira-G, using the Python implementation of the candidate-size and Wasserstein-selection approach documented in [26,55]. Mira-G replaces its variance estimate by the stated positive floor before studentization; this safeguard is not in the inspected lawstat R formula, so the column is not labeled unmodified Mira. Appendix E, Section S2 documents this audit and no executed R-parity claim is made. Section 6.6 adds the direct consecutive-system extropy predecessor [30], and Section 6.10 adds HKM [27], an affine-invariant empirical-characteristic-function test calibrated here

by its conditional sign-permutation construction. The latter provides a methodologically independent modern benchmark rather than another variant of the location-contrast family.

The extropy and HKM implementations are source-audited additions. The extropy article has an ambiguous finite-sample window description, so both documented readings are retained; HKM uses $a = 3, B = 399$ sign permutations, and Gauss-Hermite evaluation of its weighted characteristic-function norm, with an internal regression check against the closed-form observed statistic. Neither comparator is represented as original-author-code parity. For a rejection probability p, the Monte Carlo standard error of an estimated rate is $\sqrt{p(1-p)/R}$: approximately 0.0069 at $p = 0.05$ and at most 0.0158 for $R = 1000$. Separately, the binomial approximation for a single bootstrap tail probability near 0.05 gives about 0.0109 for $B = 399$ and 0.0049 for $B = 1999$. The latter is a reference approximation, not an exact error formula for maxima using estimated shared standardization. Neither calculation supplies a universal significance threshold for differences between paired test procedures.

Tables 2–4 report raw rejection rates and must be read with Table 1. Size distortion under one symmetric model cannot provide a universal correction for every asymmetric alternative. A separate reference-specific recalibration experiment, using independent training and evaluation samples, is reported in Section 6.5 and Appendix E, Section S4. The accompanying archive contains the exact generators, seed streams, scripts, and replicate-level outputs for the original replay and the additional experiments.

6.2. Empirical size

Table 1. Empirical rejection rates under symmetric null distributions at nominal level 0.05.

Distribution	N	ACES	MGG	CM	Mira-G
Normal	20	0.022	0.015	0.010	0.024
Normal	50	0.037	0.034	0.041	0.033
Normal	100	0.040	0.042	0.046	0.052
Logistic	20	0.013	0.015	0.016	0.036
Logistic	50	0.047	0.042	0.048	0.046
Logistic	100	0.031	0.043	0.049	0.052
Laplace	20	0.017	0.024	0.017	0.030
Laplace	50	0.048	0.062	0.052	0.061
Laplace	100	0.039	0.051	0.058	0.064
Student-t(3)	20	0.019	0.040	0.026	0.029
Student-t(3)	50	0.038	0.068	0.048	0.057
Student-t(3)	100	0.030	0.063	0.062	0.055
Uniform	20	0.024	0.016	0.030	0.040
Uniform	50	0.036	0.048	0.061	0.038
Uniform	100	0.049	0.060	0.083	0.051
Symmetric mixture	20	0.051	0.061	0.055	0.060
Symmetric mixture	50	0.053	0.074	0.095	0.058
Symmetric mixture	100	0.059	0.123	0.159	0.080

Note: Entries are rejection proportions from 1000 Monte Carlo replications at nominal level 0.05. Mira-G denotes the explicitly guarded implementation in Section 6.1. Monte Carlo uncertainty and implementation scope apply to every column.

Table 1 shows conservative ACES rejection rates of 0.013–0.024 for the five unimodal nulls at $N = 20$. At $N = 50$ and 100 these range from 0.030 to 0.049. For the specified symmetric mixture, they are 0.051, 0.053, and 0.059, compared with MGG and Cabilio–Masaro rates of 0.123 and 0.159 at $N = 100$. This is favorable calibration evidence for the examined mixture, not a proof of general multimodal robustness or evidence that the competitors assume unimodality.

6.3. Empirical power

Tables 2–4 report rejection rates under the ten asymmetric alternatives. Boldface marks the largest entry in each row and is descriptive, not a significance assessment. This benchmark set comprises MGG, CM, and Mira-G; the expanded comparisons with HKM in Sections 6.9–6.10 favor HKM for several smooth alternatives. The early tables therefore do not establish broad power dominance.

Table 2. Empirical power at nominal level 0.05 for $\boldsymbol{N = 20}$.

Alternative	ACES	MGG	CM	Mira-G
Skew-normal (3)	0.047	0.055	0.047	0.060
Skew-normal (6)	0.086	0.082	0.084	0.097
Lognormal (0.5)	0.139	0.178	0.127	0.139
Lognormal (1.0)	0.462	0.569	0.292	0.267
Gamma (2)	0.145	0.159	0.122	0.164
Gamma (4)	0.078	0.086	0.062	0.086
Beta (2,5)	0.055	0.058	0.045	0.070
Contaminated normal	0.059	0.076	0.055	0.073
Weibull (1.5)	0.106	0.103	0.087	0.114
Reflected Weibull (1.5)	0.107	0.114	0.087	0.130

Note: All entries use 1000 Monte Carlo replications. Boldface marks the largest reported rate, not a significance result; ties are marked jointly. Mira-G is the guarded implementation defined in Section 6.1.

Table 3. Empirical power at nominal level 0.05 for $\boldsymbol{N = 50}$.

Alternative	ACES	MGG	CM	Mira-G
Skew-normal (3)	0.222	0.166	0.177	0.156
Skew-normal (6)	0.444	0.317	0.331	0.302
Lognormal (0.5)	0.652	0.542	0.521	0.490
Lognormal (1.0)	0.987	0.969	0.912	0.888
Gamma (2)	0.696	0.543	0.511	0.475
Gamma (4)	0.356	0.291	0.308	0.269
Beta (2,5)	0.258	0.175	0.172	0.162
Contaminated normal	0.200	0.185	0.178	0.179
Weibull (1.5)	0.538	0.395	0.392	0.360
Reflected Weibull (1.5)	0.526	0.391	0.377	0.391

Note: All entries use 1000 Monte Carlo replications. Boldface marks the largest reported rate, not a significance result; ties are marked jointly. Mira-G is the guarded implementation defined in Section 6.1.

Table 4. Empirical power at nominal level 0.05 for $\boldsymbol{N = 100}$.

Alternative	ACES	MGG	CM	Mira-G
Skew-normal (3)	0.490	0.385	0.389	0.365
Skew-normal (6)	0.828	0.638	0.643	0.618
Lognormal (0.5)	0.971	0.885	0.881	0.867
Lognormal (1.0)	1.000	1.000	0.997	0.997
Gamma (2)	0.973	0.845	0.843	0.833
Gamma (4)	0.740	0.559	0.578	0.551
Beta (2,5)	0.527	0.369	0.390	0.345
Contaminated normal	0.466	0.396	0.407	0.398
Weibull (1.5)	0.909	0.725	0.738	0.707
Reflected Weibull (1.5)	0.901	0.713	0.723	0.725

Note: All entries use 1000 Monte Carlo replications. Boldface marks the largest reported rate, not a significance result; ties are marked jointly. Mira-G is the guarded implementation defined in Section 6.1.

Several competitors have higher empirical power at $N = 20$. At $N = 50$ and 100, ACES has the largest reported rejection rate in each row, with a tie with MGG for Lognormal (1.0) at $N = 100$. The right- and left-skewed Weibull results are similar. These comparisons describe the experiments; they do not imply uniform dominance or equal-size efficiency, and small differences should be interpreted with the Monte Carlo uncertainty in Section 6.1. This statement concerns MGG, CM, and Mira-G only; the larger benchmark set including HKM leads to the more qualified comparison in Sections 6.9–6.10.

The original aggregate results alone do not identify which ingredient produces each gain. In Section 6.7, we address this directly: Single-order, alternative-weight, and alternative-architecture experiments are run with common random numbers. The resulting differences are treated as design diagnostics rather than retrospective tuning, and the baseline ACES specification remains unchanged in all primary tables.

6.4. Null-only window sensitivity

A separate null-only experiment compares

$$m_N = min\{\lfloor c\sqrt{N} \rfloor, \lfloor (N-3)/4 \rfloor\}$$

for $c = 1.00, 1.25, 1.50, 1.75, 2.00$. Table 5 summarizes the six null models at $N = 50$ and 100, with $R = 1000$ and $B = 399$. The original analysis selected $c = 2$ before alternative-model power evaluation. The nearby results for $c = 1.75$ and 2 do not establish a unique optimum; moreover, the interior cap can make different multipliers yield the same window at a particular sample size.

Table 5. Null-only sensitivity of ACES to the spacing-window multiplier c.

c	Mean size	Minimum	Maximum	Mean |size-0.05|	Max |size-0.05|
1.00	0.0305	0.021	0.047	0.0195	0.029
1.25	0.0335	0.022	0.047	0.0165	0.028
1.50	0.0372	0.025	0.054	0.0134	0.025
1.75	0.0398	0.029	0.057	0.0113	0.021
2.00	0.0413	0.025	0.061	0.0116	0.025

Note: The summary uses only symmetric null models and therefore does not tune c to alternative-model power. Differences of a few thousandths are within the Monte Carlo resolution of this experiment.

6.5. Independent calibration checks

An independent check uses

$$R = 1000,\ N = 100, \text{and } B = 1999$$

for each of the six null models. ACES rejection rates range from 0.035 to 0.044 for the five unimodal models and equal 0.066 for the symmetric mixture (Appendix E, Table S1). The latter indicates residual over-rejection; increasing the bootstrap count does not guarantee nominal size. These runs use new samples and therefore do not isolate a causal effect of changing B.

For a reference-specific size-matched diagnostic, each method is recalibrated using 4000 independent null samples at

$$N = 100 \text{ and } B = 39,$$

with randomized treatment of cutoff ties. Independent validation and power evaluation each use 1000 samples. Achieved validation sizes range from 0.046 to 0.054 under the normal reference and 0.043 to 0.059 under the mixture reference. ACES retains higher calibrated rejection rates in the four selected alternatives, including 0.553 versus 0.427 for MGG under skew-normal (3). All cutoffs, achieved sizes, powers, and uncertainty estimates are given in Appendix E, Tables S2 and S3. This diagnostic depends on the specified reference law and is not a new uniformly calibrated composite-null test.

6.6. Direct comparison with the consecutive-system extropy predecessor

Reference [30] is the closest structural predecessor, so we include a direct source-audited implementation. Its power section reports m=floor(N/2)−1 and, separately, m=sqrt(n) in its Monte Carlo description. We implement the latter using the integer convention m=floor(sqrt(N)) as the primary reading, and retain the half-minus-one rule as a sensitivity analysis. In both cases, the published symmetric GLD Case 1 calibration is reproduced with 50,000 reference samples. No original-author-code parity is claimed. Table 6 summarizes the corresponding empirical null rejection rates under both window conventions.

Table 6. Empirical size at *N*=100 for ACES and the published extropy calibration under two documented window readings.

Symmetric null	ACES	Extropy sqrt(N)	Extropy N/2-1
Normal	0.027	0.045	0.059
Logistic	0.039	0.074	0.134
Laplace	0.043	0.103	0.201
Student-t(3)	0.030	0.137	0.365
Uniform	0.043	0.232	0.022
Symmetric mixture	0.061	0.191	0.040

The extropy procedure has substantially higher raw rejection rates under several symmetric laws, including 0.232 under Uniform and 0.191 under the symmetric mixture for the sqrt(N) reading, while the half-minus-one reading shifts the distortion to other nulls. Its raw alternative rejection rates are correspondingly high and are not interpreted as equal-size power. This comparison therefore resolves the missing-predecessor issue but also shows why common-target calibration is needed before making power claims.

6.7. Score, weight, and architecture ablations

A separate common-random-number experiment at $N = 50$ and 100 compares the baseline max-combined score with each single score order, uniform and beta-shaped positive weights, a tail-emphasizing weight, and several admissible consecutive architectures. The baseline is not retuned after these results are observed. Table 7 reports selected ablation rejection rates for these prespecified comparisons.

Table 7. Selected $\boldsymbol{N = 100}$ ablation results; entries are 5% rejection proportions from 300 common-random-number replications.

Model	Baseline max	$\beta = 2$ only	Uniform weight	Architecture (5,1)
Mean over six nulls	0.039	0.037	0.041	0.042
Skew-normal (3)	0.467	0.470	0.480	0.513
Gamma (4)	0.740	0.763	0.760	0.803
Contaminated normal	0.423	0.433	0.457	0.487
Weibull (1.5)	0.923	0.933	0.937	0.950
Reflected Weibull (1.5)	0.907	0.917	0.923	0.943

The ablation does not identify a uniformly optimal design. At $N = 100$, the quadratic component is frequently the winning baseline score under strong asymmetry, but the max combination is comparable under weaker departures. Uniform and architecture-(5,1) weights improve several alternatives by a few percentage points while leaving average null rejection near the same range. Together with Remark 4.3, this supports interpreting consecutive systems as a structured weighting family rather than an optimality claim.

6.8. Larger-sample and rounding diagnostics

The larger-sample check extends ACES to $N = 200$ and 500 using the same implemented rule and $B = 399$. The symmetric mixture receives 1000 outer replications at each N; the other models use 500. Table 8 reports the six nulls.

Table 8. ACES empirical size at $\boldsymbol{N = 200}$ and 500.

Symmetric null	N=200	N=500
Normal	0.038	0.036
Logistic	0.042	0.032
Laplace	0.086	0.046
Student-t(3)	0.032	0.046
Uniform	0.046	0.028
Symmetric mixture	0.054	0.047

The $N = 100$ mixture over-rejection is not reproduced in these two independent larger-N runs: the mixture rate is 0.054 at $N = 200$ and 0.047 at $N = 500$. This is evidence of improving calibration in the examined sequence, not an asymptotic theorem. Laplace is the main exception at $N = 200$, with rate 0.086, before returning to 0.046 at $N = 500$; its median kink also lies outside the smooth compact

sufficient class of Corollary 5.1. A separate rounding experiment shows that 0.1-unit or two-significant-digit rounding is relatively benign in the examined settings, whereas integer rounding can activate the zero-span floor extensively and substantially alter size and power. ACES should therefore not be presented as a procedure for intrinsically discrete or coarsely rounded data.

6.9. Balanced portfolio equal-target diagnostic

To reduce raw-size confounding without assuming knowledge of the true symmetric parent, a balanced training portfolio pools 1000 independent samples from each of four prespecified symmetric laws: normal, student-t(3), uniform, and the symmetric mixture, all at $N = 100$. For each method and target level, one pooled cutoff is selected with randomized treatment of boundary ties and then evaluated on independent 1000-sample cells for all six nulls and ten alternatives. The same fixed portfolio is applied to four methods: ACES, MGG, the source-audited extropy comparator, and HKM. The diagnostic is non-oracle because the same fixed portfolio is used for every evaluation model; it is not a theorem of uniform composite-null calibration.

Table 9. Four-method portfolio-calibrated $N = 100$ rejection rates at target 0.05 for selected nulls and alternatives.

Model	Target	ACES	MGG	Extropy	HKM
Normal	Size	0.047	0.027	0.010	0.048
Logistic	Size	0.043	0.032	0.009	0.050
Laplace	Size	0.047	0.021	0.028	0.053
Student-t(3)	Size	0.036	0.043	0.052	0.055
Uniform	Size	0.055	0.048	0.085	0.045
Symmetric mixture	Size	0.061	0.085	0.044	0.037
Skew-normal (3)	Power	0.513	0.311	0.377	0.705
Lognormal (0.5)	Power	0.971	0.844	0.978	0.996
Gamma (4)	Power	0.758	0.479	0.747	0.931
Beta (2,5)	Power	0.563	0.309	0.650	0.812
Contaminated normal	Power	0.439	0.274	0.242	0.744
Weibull (1.5)	Power	0.917	0.647	0.973	0.986

Under the common portfolio target, HKM has the largest rejection rate for most smooth alternatives in Table 9, including skew-normal (3), Gamma (4), Beta (2,5), contaminated normal, and Weibull (1.5). ACES has higher calibrated rejection rates than MGG throughout the selected alternatives and has size estimates close to the common target for most nulls, while the extropy procedure occupies a different power-calibration tradeoff. The result therefore narrows the ACES claim: Its contribution is not universal power dominance over modern omnibus tests, but a cancellation-free spacing construction with interpretable local weighting, strong reproducibility, and a distinct robustness profile. Full four-method size-power curves at target levels 0.01, 0.025, 0.05, and 0.10 are reported in Appendix E and the retained computational archives.

6.10. Independent characteristic-function benchmark and Cauchy stress

HKM [27] supplies an omnibus comparator based on the empirical characteristic function rather than spacings or consecutive-system information measures. We implement its univariate affine-

invariant statistic with $a = 3$ and $B = 399$ conditional Rademacher sign permutations; a 32-node Gauss-Hermite rule evaluates the permutation norm, and a regression test verifies the all-positive-sign quadrature against the closed-form observed statistic. On the same $N = 100$ evaluation samples, its raw 5% rejection rates are 0.059, 0.062, 0.064, 0.066, 0.060, and 0.048 for Normal, Logistic, Laplace, student-t(3), uniform, and the symmetric mixture, respectively. Under the balanced portfolio calibration of Section 6.9, HKM is the strongest of the four procedures for most smooth alternatives. This is retained as unfavorable evidence against any claim of ACES power dominance.

A separate Cauchy experiment stresses the methods under an extremely heavy-tailed symmetric null. Because Cauchy data have no finite variance, the experiment is deliberately outside ordinary finite-moment asymptotic settings and is interpreted only as a robustness stress test. ACES rejection rates remain between 0.031 and 0.043 over $N = 50,100,200,500$, whereas HKM rates range from 0.357 down to 0.105 before rising to 0.138 at $N = 500$; MGG is also strongly oversized in the evaluated $N \leq 200$ cells. These results identify a setting in which the spacing-based procedure is empirically stable without converting a stress test into a general heavy-tail theorem. Table 10 presents the Cauchy-null rejection rates supporting this stress-test comparison.

Table 10. Cauchy-null stress rejection rates at nominal level 0.05.

N	ACES	HKM	MGG
50	0.035	0.357	0.368
100	0.032	0.280	0.423
200	0.031	0.105	0.414
500	0.043	0.138	—

Note: Each rate uses 1000 independent Cauchy samples and $B = 399$ resampling/sign-permutation draws. MGG was not extended to $N = 500$ in this stress protocol because of its substantially higher computational cost. The table is a finite-sample robustness diagnostic, not a comparison under a shared asymptotic theorem.

6.11. The ACES-R core: size, power, and the cost of blocking

The revised experiment uses 400 independently generated data samples per model and evaluates complete-enumeration ACES-R and full ACES on the same sample in each replication. ACES-R uses the max-over-β statistic (52), $K = 10$, $b = 10$, $m = 2$ from (51), and all 1024 sign configurations through their 512 global-sign pairs. At nominal level 0.05, (54) gives the upper attainable level 0.048828125. Full ACES uses $B = 399$ symmetrized-bootstrap draws. The cell seeds, all replicate-level p-values, and exact generator definitions are recorded in Appendix D and the revision code archive. No surrogate MGG or HKM implementation is added to these new tables; their comparisons remain in Sections 6.2–6.10.

The symmetric-mixture null in Table 11 is the equal-weight mixture of N(−2, 1) and N(2, 1); its component means differ from the −1.5 and 1.5 values used in the original null panel of Section 6.1.

Table 11 reports raw rejection proportions, not size-matched power. The ACES-R null estimates range from 0.0425 to 0.0550 with outer Monte Carlo uncertainty; finite outer simulations are not constrained to stay below the true rejection probability. On the six displayed alternatives, full ACES has higher raw rejection proportions, by 0.1275–0.1775. This is an observed trade-off, not an efficiency theorem or an equal-size comparison: full ACES has its own nonuniform null rejection rates. Appendix D reports cellwise Monte Carlo uncertainty and paired standard errors for the differences. The separate

$K = 6$ sensitivity run is retained in the archive with its own seeds and is not represented as a same-stream comparison.

Table 11. Paired-sample raw rejection proportions at $N = 100$, nominal level 0.05, with 400 replications per cell (ACES-R upper attainable complete-orbit level 0.048828125). ACES-R uses complete enumeration; full ACES uses bootstrap calibration. Parentheses contain estimated Monte Carlo standard errors.

Model	ACES-R (MCSE)	Full ACES (MCSE)
Normal	0.0550 (0.0114)	0.0250 (0.0078)
Logistic	0.0500 (0.0109)	0.0375 (0.0095)
Laplace	0.0425 (0.0101)	0.0225 (0.0074)
Student-t(3)	0.0500 (0.0109)	0.0375 (0.0095)
Uniform	0.0425 (0.0101)	0.0500 (0.0109)
Symmetric mixture	0.0450 (0.0104)	0.0625 (0.0121)
Skew-normal (3)	0.3150 (0.0232)	0.4775 (0.0250)
Lognormal (0.5)	0.8050 (0.0198)	0.9775 (0.0074)
Gamma (4)	0.5725 (0.0247)	0.7475 (0.0217)
Beta (2,5)	0.3950 (0.0244)	0.5725 (0.0247)
Contaminated normal	0.3275 (0.0235)	0.4550 (0.0249)
Weibull (1.5)	0.7400 (0.0219)	0.9050 (0.0147)

6.12. Tail-path experiment and an exploratory niche

The tail-path study is defined by the executable generators rather than by rounded table labels. Its symmetric laws are Student-t_3, Student-t_{10}, and Cauchy. Its asymmetric laws are generated by

$$X_\nu = \frac{Z_0 + (2/3)|Z_1|}{\sqrt{V/\nu}}, \qquad Z_0, Z_1 \sim N(0,1), \quad V \sim \nu_2^{\chi},$$

independently, with $\nu = 3$ or 10. Up to a positive scale factor, this is the skew-t construction of [56] with shape $2/3$. Scale does not affect either spacing discrepancy. The same-stream comparison again uses complete-enumeration ACES-R and full ACES; it does not substitute new implementations for the benchmark sections.

The $N = 50$ configuration uses $K = 6$, while $N = 100$ and 200 use $K = 10$. Their upper attainable complete-orbit level bounds are 0.03125 and 0.048828125, respectively; additional ties can lower the true rejection probability further. In particular, the $N = 50$ comparison includes stronger discreteness-induced conservativeness; an apparent power difference cannot be attributed to sample size alone. More Monte Carlo sign draws would not enlarge the underlying finite sign orbit. Although a sampled p-value can cross a threshold by chance, such a crossing is not a correction of the complete-orbit support limitation (see Table 12).

The new Student-t_3, Student-t_{10}, and Cauchy null estimates are compatible with the finite-group guarantee at the resolution of these simulations; values such as 0.0575 are estimates, not a violation of the bound on the true rejection probability. For skew-t_3, the ACES-R rejection proportions increase from 0.0675 to 0.1650 to 0.3100 as N increases from 50 to 100 to 200, with the accompanying change in block design noted above. The small skew-t_{10} difference at $N = 100$ does not establish a power advantage. The supported contribution is structural finite-sample level control for ACES-R on the

continuous symmetry null, alongside a transparent comparison with an empirically calibrated full-sample procedure. Neither heavy-tail power dominance nor equality of the competitors' sizes is claimed. Stable calibration under a heavy-tailed symmetric null is calibration-robustness evidence; it does not establish better power against heavy-tailed asymmetric alternatives. A practical preference would require further size-matched evidence, so the proposed niche remains exploratory.

Table 12. Student-t/Cauchy and skew-t path: 400 paired data replications per cell at nominal level 0.05 (ACES-R upper attainable levels 0.03125 for $N = 50$ and 0.048828125 for $N = 100,200$). ACES-R uses $(K, b, m) = (6,8,2)$ at $N = 50$, $(10,10,2)$ at $N = 100$, and $(10,20,4)$ at $N = 200$. Parentheses contain estimated Monte Carlo standard errors.

Model	N	ACES-R (MCSE)	Full ACES (MCSE)
Student-t(3) null	50	0.0225 (0.0074)	0.0400 (0.0098)
Student-t(3) null	100	0.0550 (0.0114)	0.0325 (0.0089)
Student-t(3) null	200	0.0575 (0.0116)	0.0225 (0.0074)
Student-t(10) null	100	0.0450 (0.0104)	0.0325 (0.0089)
Cauchy null	50	0.0325 (0.0089)	0.0425 (0.0101)
Cauchy null	100	0.0550 (0.0114)	0.0350 (0.0092)
Cauchy null	200	0.0425 (0.0101)	0.0475 (0.0106)
Skew-t(3)	50	0.0675 (0.0125)	0.1075 (0.0155)
Skew-t(3)	100	0.1650 (0.0186)	0.1975 (0.0199)
Skew-t(3)	200	0.3100 (0.0231)	0.3500 (0.0238)
Skew-t(10)	100	0.0650 (0.0123)	0.0550 (0.0114)

7. Data applications

The two examples in this section serve complementary purposes: They demonstrate the test in action on complete engineering observations, one rejection and one nonrejection, and they show how the local score curve localizes asymmetry descriptively, in probability-scale terms that can be related back to the underlying measurement scale. Three interpretive boundaries apply throughout. First, the system interpretation enters only through the probability weights; these are observed component-level measurements, not lifetimes of constructed consecutive systems. Second, the local score curve is a descriptive diagnostic without calibrated pointwise or simultaneous significance statements. Third, each inferential conclusion is stated at its proper strength: Rejection is interpreted at the stated nominal level of the implemented calibration, whereas nonrejection indicates compatibility with symmetry rather than proof of symmetry.

7.1. Airborne transceiver repair times

The first dataset contains 45 airborne transceiver repair times, measured in hours, reproduced through the source chain documented in Appendix E, Section S1 and attributed there to Qiu and Jia (see [57]). Direct primary-source verification of the original Qiu/Jia table is not claimed, and the separate 46-observation literature variant is not substituted for this input. The sample median is 1.5, hours and the sample skewness is 2.947; the values 22.0 and 24.5 hours extend the upper tail substantially. Although individual observations are tied, no retained local span was zero in the reported application runs.

Using $B = 1999$ and the recorded application seed stream, the replay gives ACES $p = 0.0005$; the minimum under (49), with zero exceedances. MGG and CM each return zero empirical two-sided tail frequency, and Mira-G gives $p = 0.0030$. These finite-resampling estimates reject at the 5% level; a zero frequency is not an exact zero tail probability. An independent $B = 9999$ ACES run gives $p = 0.0002$ (one exceedance), retaining the decision.

The standardized ACES components are approximately 5.49, 6.94, and 8.46 for $\beta = 1{,}3/2{,}2$, respectively, the largest being the quadratic component, which places the greatest relative weight on the largest departures. In Figure 2 the signed difference is negative, corresponding to greater quantile-density expansion at the upper complementary probability levels; correcting the historical plotting sign does not change the absolute discrepancy or the test. Read against the data scale, the curve localizes the asymmetry where the engineering concern is most visible: in the upper tail, where the two extreme repair times (22.0 and 24.5 hours, far above the median of 1.5 hours) inflate the quantile density at high complementary levels. In practical terms, a symmetric location working model should not be adopted without checking its tail fit; median-based turnaround summaries and upper-quantile summaries may describe different aspects of this sample. These are descriptive implications accompanying the reported rejection, not calibrated statements about specific quantiles.

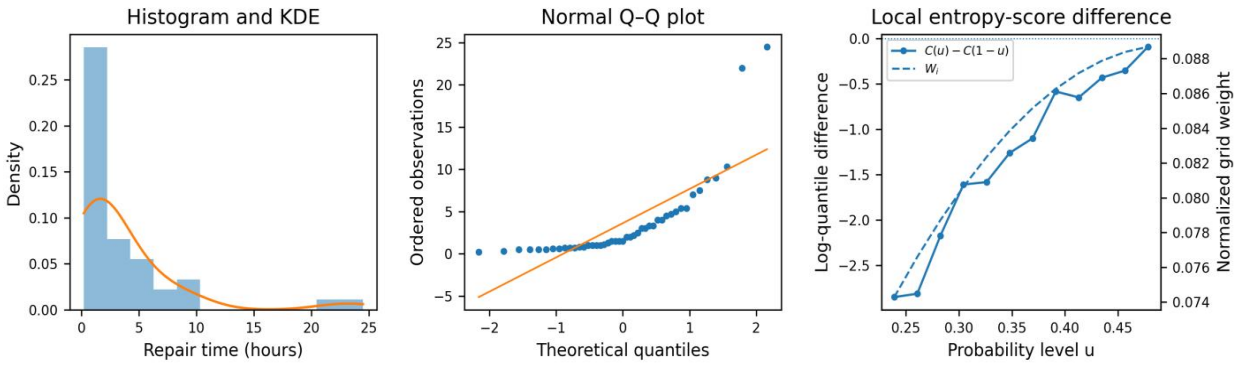


Figure 2. Repair-time histogram with kernel density estimate, normal Q–Q plot, and signed local score $\boldsymbol{C(u) - C(1-u) = log\, q(u) - log\, q(1-u)}$. The dashed curve shows the normalized discrete grid weights $\boldsymbol{W_i}$, not the continuously normalized weight function.

For maintenance planning, the long-observed repairs suggest examining upper repair-time quantiles in addition to an average or median when setting turnaround targets. The symmetry test does not estimate those quantiles, spare-parts requirements, availability, or a scheduling policy. Any operational decision would also require a model of workload, failure arrivals, resources, and the costs of delays. The local-score plot identifies where the sample's discrepancy is concentrated; it is not a calibrated prediction of future delays.

7.2. Fiber tensile strength

The second example uses the 50 Nextel 610 fiber tensile strengths for lot A0168 in [58, Table 2], measured in MPa. All values are checked against the author-posted table transcription. The sample skewness is 0.409. Figure 3 gives the histogram and normal Q–Q diagnostics; Appendix E, Figure S1 adds the complementary local score curve.

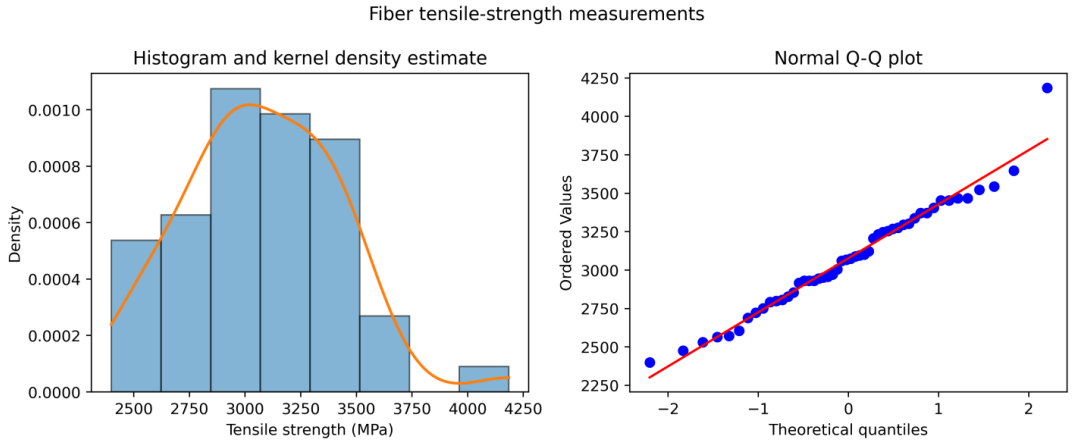


Figure 3. Histogram with kernel density estimate and normal Q–Q plot for the tensile-strength observations.

The application replay gives ACES $p = 0.6975$; MGG, CM, and Mira-G give 0.8914, 0.8884, and 0.8764, respectively. An independent $B = 9999$ ACES run gives $p = 0.7010$. None of the procedures rejects symmetry at the 5% level, and the four p-values are mutually consistent rather than in tension.

Operationally, nonrejection at this sample size supports treating a symmetric location model as a plausible working approximation for this lot, subject to the explicit caveat that nonrejection is compatibility rather than confirmation. A normal or other symmetric location-scale model may therefore be considered if scientifically appropriate. A lognormal model is not symmetric on the original strength scale; it corresponds instead to symmetry after logarithmic transformation and would require assessing the transformed data separately. Lot-specific confirmatory data would still be desirable before treating symmetry as established.

As a numerical robustness observation, multiplying the zero-span floor by 0.5 or 2 leaves both application results unchanged in the tested runs because the floor never activates; this is not evidence of general robustness to severe rounding. Finally, nonrejection is not proof of symmetry, and a normal Q–Q plot is not a distribution-free symmetry test; the formal test and graphical diagnostics simply agree in this example.

8. Discussion and concluding remarks

In this paper, we deliberately separate characterization from testing while making their common weighting mechanism explicit. The global FGE and cumulative-entropy results of Section 3 identify symmetry through complete families of reflected system equalities, whereas the ACES test uses the reflected kernels only to generate a structured family of positive probability-scale weights for a cancellation-free log-quantile-density discrepancy. Because any strictly positive integrable weight has the same zero set, consecutive systems are neither necessary for identification nor claimed to be canonical; their contribution is an interpretable reliability-based parameterization that can be compared on equal footing with simpler weights. We regard this separation of roles, characterization through signed global identities, detection through a nonnegative local functional, and weighting through a structured but non-essential family, as the paper's main conceptual message, and the ablations of Section 6.7 are its natural empirical counterpart.

The central inferential trade-off is finite-sample validity versus information loss from blocking. ACES-R provides a distinct finite-sample-valid inferential core based on the same spacing-asymmetry framework; it does not validate full-ACES bootstrap calibration. At $N = 100$ and $K = 10$, each block has only ten observations, and Table 11 shows lower raw rejection than full ACES for all six displayed asymmetric alternatives. These procedures may have different effective finite-sample null rejection probabilities, so the comparison is not an equal-size efficiency result. Theorem 5.2 controls the nominal level through an upper attainable orbit bound rather than an exact size equality, and Theorem 5.3 proves consistency for fixed K and growing block size on its stated compact-support class. We do not infer that every nonblocked symmetry statistic lacks a group-based calibration.

The enlarged evidence also sharpens the domain of contribution. ACES is conservative at *N*=20; the independent *N*=200 and 500 runs do not reproduce the earlier mixture over-rejection, although Laplace remains a finite-sample stress case at *N*=200; and coarse integer rounding is confirmed as a limitation. The direct comparison with the extropy predecessor reveals strong reference-calibration sensitivity, and the ablations show that neither the current system weight nor the three-score maximum is uniformly optimal. Most importantly, the independent HKM benchmark has higher rejection rates for many smooth alternatives in the balanced-portfolio diagnostic. ACES nevertheless remains competitive against the classical benchmarks and shows markedly more stable behavior in the separate Cauchy stress experiment. The defensible conclusion is therefore a reproducible, interpretable spacing methodology with a distinct calibration-robustness profile, not superiority over all symmetry tests. The Cauchy stability and favorable mixture calibration suggest an exploratory candidate niche in heavy-tailed or shape-heterogeneous settings. These are calibration-robustness observations, not evidence by themselves of a power advantage against asymmetric heavy-tailed alternatives; establishing a practical preference requires further size-matched study.

The main theoretical limitation is broader than bootstrap calibration: a primitive joint null limit for the paired overlapping log-spacing vector has not been proved here. The estimator is supported by Theorem 5.1 and the verifiable compact-support Corollary 5.1, while Lemma 5.2 controls the specific zero-span degeneracy of empirical resampling. Hall's increasing-order spacing results [31,33] and recent theory for higher-order overlapping spacings [34] indicate a plausible route to a primitive central limit theorem, but adapting those results to reflected log-ratios, nonconstant weights, and the three-component maximum requires a separate proof. Consequently, in Section 5.4, we provide only a conditional continuous-mapping consequence, and the implemented symmetrized bootstrap remains an empirically studied finite-sample calibration device rather than a proved asymptotic theorem for ACES.

Open problems. These limitations define a focused research agenda rather than an open-ended list. We state the problems in order of logical dependency, from the fluctuation law outward, and indicate what each would deliver if resolved.

P1. Primitive joint fluctuation law. Derive the joint null limit of the reflected, weighted log-spacing vector

$$\left(log\hat{q}(u_i) - log\hat{q}(1-u_i)\right)_{i\in I_N}$$

under

$$\frac{m_N}{logN} \to \infty, \qquad \frac{m_N}{N} \to 0.$$

This is the bottleneck step: The windows overlap strongly, and the dependence range grows with m_N, so standard spacing CLTs do not apply directly. A resolution would convert the max-combined statistic from an empirically calibrated object into one with a limit law and would quantify how the covariance of overlapping spacings propagates through the reflected log-ratio.

P2. Bootstrap validity for full ACES. Determine whether the Hodges–Lehmann-centered symmetrized bootstrap consistently reproduces the law in P1 (for ACES-R this question is moot: finite-sample level control is proved in Theorem 5.2). Arcones and Giné [59] prove symmetrized-bootstrap validity for other nonsmooth symmetry statistics, but their theorem does not cover the intermediate-spacing construction of ACES. The centering error must be shown negligible relative to s_N,β, and the estimated-to-ideal scale ratios must converge to one, before the conditional transfer statement of Section 5.4 becomes a primitive theorem.

P3. Local detection rate and efficiency. Once P1 is available, we do not assign a local detection rate, a universal covariance formula, or a Bahadur/Pitman efficiency comparison. Proposition 4.2 identifies how local asymmetry directions enter the discrepancy for a given weight; pairing that geometry with the fluctuation law would yield efficiency statements for cancellation-free spacing tests of symmetry and would clarify whether the quadratic score order's empirical advantage under strong asymmetry (Section 6.7) has a local-theoretic explanation.

P4. Weight and score-order selection theory. The ablations show that architecture $(5,1)$ and uniform weights improve several alternatives by a few percentage points, yet no selection rule is available that would choose a weight or score order from data without invalidating the calibration. A data-adaptive choice with controlled multiplicity, or a proved statement that prespecification is asymptotically costless, would close the gap between the ablation evidence and practice.

P5. Rounded and discrete data. The zero-span floor is a numerical convention, not a model, and integer rounding can activate it extensively (Section 6.8). An extension that treats rounded observations as interval-censored at the resolution scale, or a spacing estimator built for genuinely discrete distributions, would extend the framework beyond continuous parents; the occupancy analysis of Lemma 5.2 suggests the bootstrap-side difficulties would be nontrivial.

P6. Characterization converses beyond the proved cases. No general converse is claimed for fractional orders $\alpha \neq 1$ in the FGE family (Remark 3.1), nor for general architectures $r \geq 1$ in the cumulative family (Corollary 3.1). Whether the nonlinear moment equations at $\alpha \neq 1$ admit a completeness argument, or whether the sign-restriction route of Proposition 3.2 can be weakened, remains open; a positive answer would widen the class of symmetry characterizations supplied by consecutive systems, though it would not change the testing layer, whose zero-set property is already weight-agnostic.

This work provides a distinct finite-sample-valid ACES-R core, with the observed blocking cost and conservative orbit resolution stated explicitly. It does not validate full ACES, whose primitive joint null limit and bootstrap validity remain unresolved; full ACES is retained as an empirically calibrated procedure. Resolving P1 and P2 would give the full-sample procedure a primitive asymptotic foundation and enable subsequent efficiency analysis. No asymptotic optimality claim is made; that question belongs to P3. P3–P6 remain open, as stated.

Author contributions

Ibrahim A. Nafisah: conceptualization, methodology, formal analysis, and writing–review and editing; Tahani Alshathri: methodology, software, validation, and writing–review and editing; Mashael

A. Alshehri: conceptualization, methodology, supervision, writing–original draft, and writing–review and editing. All authors have read and approved the final version of the manuscript for publication

Use of Generative-AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgements and funding

This work was supported by the Ongoing Research Funding program (ORF-2026-1913), King Saud University, Riyadh, Saudi Arabia.

The authors thank the editors and anonymous reviewers for their careful reading and constructive comments, which helped improve the clarity and rigor of the manuscript.

Data and code availability

The computational documentation is consolidated in Appendices C–E. Appendix E incorporates the earlier computational supplement (Sections S1–S15, Tables S1–S16, and Figure S1); Appendix D documents ACES-R revision computations. The previously supplied original-computations and revision-computations archives contain the simulation code, application input datasets, random seeds, regression tests, and replicate-level outputs. These machine-readable files are retained as the authors' computational provenance, not embedded as executable code or data files in this Word article. No additional compressed file accompanies the present proof; the underlying files can be requested from the corresponding author. Claims about the executed replays, the limitations of bootstrap calibration, and the repair-data source qualifications remain unchanged.

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. M. S. Mohamed, M. A. Almuqrin, Properties of fractional generalized entropy in ordered variables and symmetry testing, *AIMS Math.*, **10** (2025), 1116–1141. https://doi.org/10.3934/math.2025053
2. M. Mahdizadeh, E. Zamanzade, Estimation of a symmetric distribution function in multistage ranked set sampling, *Stat. Pap.*, **61** (2020), 851–867. https://doi.org/10.1007/s00362-017-0965-x
3. M. Fashandi, J. Ahmadi, Characterizations of symmetric distributions based on Rényi entropy, *Stat. Probab. Lett.*, **82** (2012), 798–804. https://doi.org/10.1016/j.spl.2012.01.004
4. N. Balakrishnan, A. Selvitella, Symmetry of a distribution via symmetry of order statistics, *Stat. Probab. Lett.*, **129** (2017), 367–372. https://doi.org/10.1016/j.spl.2017.06.023
5. J. Ahmadi, M. Fashandi, Characterization of symmetric distributions based on some information measures properties of order statistics, *Phys. A*, **517** (2019), 141–152. https://doi.org/10.1016/j.physa.2018.11.009

6. J. Ahmadi, M. Fashandi, Characterization of symmetric distributions based on concomitants of ordered variables from FGMs family of bivariate distributions, *Filomat*, **33** (2019), 4239–4250. https://doi.org/10.2298/FIL1913239A
7. J. Ahmadi, Characterization of continuous symmetric distributions using information measures of records, *Stat. Pap.*, **62** (2021), 2603–2626. https://doi.org/10.1007/s00362-020-01206-z
8. J. Ahmadi, Characterization results for symmetric continuous distributions based on the properties of (k)-records and spacings, *Stat. Probab. Lett.*, **162** (2020), 108764. https://doi.org/10.1016/j.spl.2020.108764
9. H. A. Noughabi, Tests of symmetry based on the sample entropy of order statistics and power comparison, *Sankhya B*, **77** (2015), 240–255. https://doi.org/10.1007/s13571-015-0103-5
10. N. Gupta, S. K. Chaudhary, Some characterizations of continuous symmetric distributions based on extropy of record values, *Stat. Pap.*, **65** (2024), 291–308. https://doi.org/10.1007/s00362-022-01392-y
11. M. Amiri, B. Khaledi, A new test for symmetry against right skewness, *J. Stat. Comput. Simul.*, **86** (2016), 1479–1496. https://doi.org/10.1080/00949655.2015.1071374
12. T. P. McWilliams, A distribution-free test for symmetry based on a runs statistic, *J. Amer. Stat. Assoc.*, **85** (1990), 1130–1133. https://doi.org/10.1080/01621459.1990.10474985
13. A. Baklizi, A conditional distribution free runs test for symmetry, *J. Nonparametr. Stat.*, **15** (2003), 713–718. https://doi.org/10.1080/10485250310001634737
14. J. D. Gibbons, S. Chakraborti, *Nonparametric statistical inference*, Marcel Dekker, 1992.
15. I. Tajuddin, Distribution-free test for symmetry based on Wilcoxon two-sample test, *J. Appl. Stat.*, **21** (1994), 409–415. https://doi.org/10.1080/757584017
16. W. Cheng, N. Balakrishnan, A modified sign test for symmetry, *Commun. Stat. Simul. Comput.*, **33** (2004), 703–709. https://doi.org/10.1081/SAC-200033302
17. R. Modarres, J. Gastwirth, Hybrid test for the hypothesis of symmetry, *J. Appl. Stat.*, **25** (1998), 777–783. https://doi.org/10.1080/02664769822765
18. R. Modarres, J. Gastwirth, A modified runs test for symmetry, *Stat. Probab. Lett.*, **31** (1996), 107–112. https://doi.org/10.1016/S0167-7152(96)00020-X
19. A. Baklizi, Testing symmetry using a trimmed longest run statistic, *Aust. N. Z. J. Stat.*, **49** (2007), 339–347. https://doi.org/10.1111/j.1467-842X.2007.00485.x
20. A. Baklizi, Improving the power of the hybrid test, *Int. J. Contemp. Math. Sci.*, **3** (2008), 497–499.
21. J. Corzo, G. Babativa, A modified runs test for symmetry, *J. Stat. Comput. Simul.*, **83** (2013), 984–991. https://doi.org/10.1080/00949655.2011.647026
22. P. Cabilio, J. Masaro, A simple test of symmetry about an unknown median, *Can. J. Stat.*, **24** (1996), 349–361. https://doi.org/10.2307/3315744
23. A. Mira, Distribution-free test for symmetry based on Bonferroni's measure, *J. Appl. Stat.*, **26** (1999), 959–972. https://doi.org/10.1080/02664769921963
24. W. Miao, Y. R. Gel, J. L. Gastwirth, A new test of symmetry about an unknown median, In: Agnes C. Hsiung, Z. Ying, C. H. Zhang, *Random walk, sequential analysis and related topics*, World Scientific, 2006, 199–214. https://doi.org/10.1142/9789812772558_0013
25. T. Zheng, J. L. Gastwirth, On bootstrap tests of symmetry about an unknown median, *J. Data Sci.*, **8** (2010), 413–427. https://doi.org/10.6339/JDS.2010.08(3).614
26. V. Lyubchich, X. Wang, A. Heyes, Y. R. Gel, A distribution-free m-out-of-n bootstrap approach to testing symmetry about an unknown median, *Comput. Stat. Data Anal.*, **104** (2016), 1–9. https://doi.org/10.1016/j.csda.2016.05.004

27. N. Henze, B. Klar, S. G. Meintanis, Invariant tests for symmetry about an unspecified point based on the empirical characteristic function, *J. Multivar. Anal.*, **87** (2003), 275–297. https://doi.org/10.1016/S0047-259X(03)00044-7
28. M. Kayid, M. Shrahili, Information properties of consecutive systems using fractional generalized cumulative residual entropy, *Fractal Fract.*, **8** (2024), 568. https://doi.org/10.3390/fractalfract8100568
29. M. Kayid, M. A. Alshehri, Shannon differential entropy properties of consecutive *k*-out-of-*n*: G systems, *Oper. Res. Lett.*, **57** (2024), 107190. https://doi.org/10.1016/j.orl.2024.107190
30. T. Alshathri, M. Kayid, Extropy-based characterization and nonparametric testing of symmetry in continuous distributions via consecutive systems, *AIMS Math.*, **11** (2026), 10638–10667. https://doi.org/10.3934/math.2026438
31. P. Hall, Limit theorems for sums of general functions of m-spacings, *Math. Proc. Cambridge Philos. Soc.*, **96** (1984), 517–532. https://doi.org/10.1017/S0305004100062459
32. N. Cressie, On the logarithms of high-order spacings, *Biometrika*, **63** (1976), 343–355. https://doi.org/10.1093/biomet/63.2.343
33. P. Hall, On powerful distributional tests based on sample spacings, *J. Multivar. Anal.*, **19** (1986), 201–224. https://doi.org/10.1016/0047-259X(86)90027-8
34. S. M. Mirakhmedov, Asymptotic properties of goodness of fit tests based on higher order overlapping spacings, *Stat. Probab. Lett.*, **225** (2025), 110461. https://doi.org/10.1016/j.spl.2025.110461
35. S. Eryılmaz, Reliability properties of consecutive *k*-out-of-*n* systems of arbitrarily dependent components, *Reliab. Eng. Syst. Saf.*, **94** (2009), 350–356. https://doi.org/10.1016/j.ress.2008.03.027
36. C. E. Shannon, A mathematical theory of communication, *Bell Syst. Tech. J.*, **27** (1948), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x
37. M. Rao, Y. Chen, B. C. Vemuri, F. Wang, Cumulative residual entropy: a new measure of information, *IEEE Trans. Inf. Theory*, **50** (2004), 1220–1228. https://doi.org/10.1109/TIT.2004.828057
38. A. Di Crescenzo, M. Longobardi, On cumulative entropies, *J. Stat. Plan. Inference*, **139** (2009), 4072–4087. https://doi.org/10.1016/j.jspi.2009.05.038
39. A. D. Crescenzo, S. Kayal, A. Meoli, Fractional generalized cumulative entropy and its dynamic version, *Commun. Nonlinear Sci. Numer. Simul.*, **102** (2021), 105899. https://doi.org/10.1016/j.cnsns.2021.105899
40. H. Xiong, P. Shang, Y. Zhang, Fractional cumulative residual entropy, *Commun. Nonlinear Sci. Numer. Simul.*, **78** (2019), 104879. https://doi.org/10.1016/j.cnsns.2019.104879
41. G. Alomani, M. Kayid, Fractional survival functional entropy of engineering systems, *Entropy*, **24** (2022), 1275. https://doi.org/10.3390/e24091275
42. M. Kayid, M. Shrahili, Some further results on the fractional cumulative entropy, *Entropy*, **24** (2022), 1037. https://doi.org/10.3390/e24081037
43. G. Psarrakos, P. Economou, On the generalized cumulative residual entropy weighted distributions, *Commun. Stat. Theory Methods*, **46** (2017), 10914–10925. https://doi.org/10.1080/03610926.2016.1252402
44. U. Kamps, Characterizations of distributions by recurrence relations and identities for moments of order statistics, *Handbook Stat.*, **16** (1998), 291–311. https://doi.org/10.1016/S0169-7161(98)16012-1

45. J. S. Hwang, G. D. Lin, On a generalized moment problem. II, *Proc. Amer. Math. Soc.*, **91** (1984), 577–580. https://doi.org/10.1090/S0002-9939-1984-0746093-4
46. O. Vasicek, A test for normality based on sample entropy, *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **38** (1976), 54–59. https://doi.org/10.1111/j.2517-6161.1976.tb01566.x
47. M. C. Jones, Estimating densities, quantiles, quantile densities and density quantiles, *Ann. Inst. Stat. Math.*, **44** (1992), 721–727. https://doi.org/10.1007/BF00053400
48. S. Boucheron, G. Lugosi, P. Massart, *Concentration inequalities: a nonasymptotic theory of independence*, Oxford University Press, 2013.
49. A. W. van der Vaart, *Asymptotic statistics*, Cambridge University Press, 1998. https://doi.org/10.1017/CBO9780511802256
50. M. Raab, A. Steger, *Balls into bins—a simple and tight analysis*, Springer, 1998, 159–170.
51. D. J. Henderson, C. F. Parmeter, A consistent bootstrap procedure for nonparametric symmetry tests, *Econ. Lett.*, **131** (2015), 78–82. https://doi.org/10.1016/j.econlet.2015.03.038
52. M. D. Ernst, Permutation methods: a basis for exact inference, *Stat. Sci.*, **19** (2004), 676–685. https://doi.org/10.1214/088342304000000396
53. J. Hemerik, J. J. Goeman, Exact testing with random permutations, *TEST*, **27** (2018), 811–825. https://doi.org/10.1007/s11749-017-0571-1
54. N. W. Koning, J. Hemerik, More efficient exact group invariance testing: using a representative subgroup, *Biometrika*, **111** (2024), 441–458. https://doi.org/10.1093/biomet/asad050
55. W. Hui, Y. R. Gel, J. L. Gastwirth, lawstat: an *R* package for law, public policy and biostatistics, *J. Stat. Softw.*, **28** (2008), 1–26. https://doi.org/10.18637/jss.v028.i03
56. A. Azzalini, A. Capitanio, Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution, *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **65** (2003), 367–389. https://doi.org/10.1111/1467-9868.00391
57. G. Qiu, K. Jia, Extropy estimators with applications in testing uniformity, *J. Nonparametr. Stat.*, **30** (2018), 182–196. https://doi.org/10.1080/10485252.2017.1404063
58. D. M. Wilson, Statistical tensile strength of Nextel™ 610 and Nextel™ 720 fibres, *J. Mater. Sci.*, **32** (1997), 2535–2542. https://doi.org/10.1023/A:1018538030985
59. M. A. Arcones, E. Giné, Some bootstrap tests of symmetry for univariate continuous distributions, *Ann. Stat.*, **19** (1991), 1496–1511. https://doi.org/10.1214/aos/1176348258
60. S. K. Chaudhary, N. Gupta, Extropy and Varextropy estimators with applications, *ArXiv*, 2024. https://doi.org/10.48550/arXiv.2401.13065
61. CRAN lawstat source, symmetry.test.R. Available from: https://github.com/cran/lawstat/blob/master/R/symmetry.test.R

Appendices

Appendix A. Auxiliary signed FGE diagnostic

This appendix retains the signed FGE contrast of Section 3 as an auxiliary diagnostic, complete with a plug-in estimator and a consistency statement. Its status is deliberately limited: It is not used to compute ACES or any of the main simulation, ablation, portfolio-calibration, or data-analysis results, and, as Remark 3.1 emphasizes, a zero value of the contrast need not imply symmetry at a fixed system size or at order two. The material is collected here for completeness of the characterization layer and for possible diagnostic use.

Define the signed contrast

$$F_{r,n,\alpha} = FH_{\alpha}(T_{n-r|n:G}) - FH_{\alpha}(T_{n-r|n:F}). \tag{A1}$$

For compact notation define

$$G_{r,n,\alpha}(u,y) = \rho_{n-r|n:G}(u)\{-log[\rho_{n-r|n:G}(u)y]\}^{\alpha} - \rho_{n-r|n:F}(u)\{-log[\rho_{n-r|n:F}(u)y]\}^{\alpha},$$

so that

$$F_{r,n,\alpha} = c(\alpha)\int_0^1 G_{r,n,\alpha}\,(u, f\{Q(u)\})\,du.$$

For an independent sample, write $u_k = k/(N+1)$ and use the endpoint-corrected density-quantile estimate

$$\widehat{f_k} = \frac{2m}{N(X_{k+m:N} - X_{k-m:N})} \tag{A2}$$

with $X_{k-m:N} = X_{1:N}$ for $k - m < 1$ and

$$X_{k+m:N} = X_{N:N}$$

for $k + m > N$. The plug-in diagnostic is

$$\hat{F}_{r,n,\alpha} = \frac{c(\alpha)}{N}\sum_{k=1}^{N} G_{r,n,\alpha}\left(u_k, \hat{f}_k\right). \tag{A3}$$

Assumption A.1. (High-level plug-in approximation) Assume the parent regularity of Lemma 3.1, $m \to \infty$, $m/N \to 0$, and finite real-valued integrands. Assume also that the population Riemann sums converge to the corresponding integral and that

$$\frac{1}{N}\sum_{k=1}^{N}\left|G_{r,n,\alpha}\left(u_k, \hat{f}_k\right) - G_{r,n,\alpha}(u_k, f\{Q(u_k)\})\right| \xrightarrow{p} 0. \tag{A4}$$

Proposition A.1. (Conditional plug-in consistency) Under Assumption A.1, $\hat{F}_{r,n,\alpha} \xrightarrow{p} F_{r,n,\alpha}$ for each fixed admissible (r,n,α).

Proof. Subtract the population Riemann sum from (A3). Its absolute difference is bounded by $c(\alpha)$ times the quantity in (A4), which tends to zero in probability. The assumed Riemann-sum convergence gives the stated limit.

This is a transfer under an explicit transformed-spacing assumption, not a primitive theorem for arbitrary supports. A small proportion of boundary indices alone would not justify ignoring unbounded logarithmic contributions.

Appendix B. Illustrative parent distributions

The three examples below exercise the density-quantile symmetry condition (7) in opposite directions. Example B.1 gives a one-parameter family that is symmetric only at a single parameter value and whose local imbalance has an explicit closed form; Examples B.2 and B.3 verify the symmetry condition for two non-trivial symmetric laws on (0,1), confirming that Theorem 3.3 and Lemma 3.1 apply to distributions beyond the usual location-scale families.

Example B.1. Let X have density $f(x) = bx^{b-1}$, $0 < x < 1$, with $b > 0$. Then $F(x) = x^b$, $F^{-1}(u) = u^{1/b}$, and

$$f\left(F^{-1}(u)\right) = bu^{(b-1)/b}, \quad C(u) = -logb - \frac{b-1}{b}logu.$$

Consequently,

$$C(u) - C(1-u) = -\frac{b-1}{b} log\left(\frac{u}{1-u}\right).$$

For $b = 1$, the distribution is uniform on $(0,1)$ and the difference vanishes. For $b \neq 1$, it is nonzero except at $u = 1/2$, so the density-quantile symmetry condition fails. The example illustrates the local imbalance targeted by the paired-system construction; it does not assert a universal sign for the integrated contrast at a general order α.

Example B.2. For the arcsine distribution on $(0,1)$,

$$f(x) = \frac{1}{\pi\sqrt{x(1-x)}}, \quad F(x) = \frac{2}{\pi} arcsin\left(\sqrt{x}\right),$$

and

$$F^{-1}(u) = sin^2\left(\frac{\pi u}{2}\right).$$

Consequently,

$$f\left(F^{-1}(u)\right) = \frac{1}{\pi sin(\pi u/2) cos(\pi u/2)} = f\left(F^{-1}(1-u)\right),$$

so the density-quantile symmetry condition holds and Theorem 3.3 applies.

Example B.3. Let

$$f(x) = \frac{\phi(logit x)}{x(1-x)}, \quad 0 < x < 1,$$

where ϕ is the standard normal density. Since

$$F^{-1}(u) = \frac{1}{1 + exp\{-\Phi^{-1}(u)\}},$$

normal symmetry implies

$$F^{-1}(1-u) = 1 - F^{-1}(u) \text{ and } \phi\left(\Phi^{-1}(1-u)\right) = \phi\left(\Phi^{-1}(u)\right).$$

Therefore,

$$f\left(F^{-1}(u)\right) = f\left(F^{-1}(1-u)\right)$$

and the distribution is symmetric about $1/2$.

Appendix C. Implementation details and numerical provenance

In this appendix, we record the computational conventions and the seed architecture that make the study reproducible, and states which outputs were verified by replay.

Computation. For ACES, compute the window in (38), the index set I_N, and the paired quantile-density estimates once for each sorted sample. Normalize the weights on the retained grid and evaluate all three powers of the same absolute log-spacing difference. The window depends on N, not on the data values.

The null resampling law uses absolute residuals about the Hodges–Lehmann pseudomedian and independent random signs. Every replicate is processed with the same spacing and numerical conventions as the original sample. The bootstrap means and standard deviations are shared by the observed and replicate maxima, retaining the dependence among score orders.

Seeds. The main study uses NumPy SeedSequence(2026061907), spawning 48 independent cell seeds in the null-then-alternative order of Section 6.1, with sample sizes 20, 50, and 100 within each model. The application stream starts at seed 20260619 and processes repair then tensile data, with ACES, MGG, CM, and Mira-G, in this order. The archived null-only window script draws signs before radii, unlike the main ACES routine; its original order is preserved for exact replay. The environment and all cell seeds are recorded in the archive.

Revision replay and provenance. The new experiments use base cell seeds $1100 + j$ for Table 11 ($j = 0, \ldots, 11$), $3100 + j$ for Table 12 ($j = 0, \ldots, 10$), and $2100 + j$ for the separate $K = 6$ sensitivity block. Replication $r = 0, \ldots, 399$ uses data seed $100003 \times cell\ seed + r$ and bootstrap seed $10^6 \times cell\ seed + r + 3$. The primary replay contains 23 cells and 9200 paired samples, including the 12 sensitivity cells gives 14000 samples and 23200 saved test p-values. Every new printed value is generated from the saved counts, not transcribed from an external audit. The script was executed twice in separate processes; result tables and replicate-level output agreed byte for byte. The original 192000-p-value replay and legacy supplemental diagnostics described in the submitted archive are not repeated in this revision replay.

Appendix D. Reproducibility details for ACES-R and the revision experiments

This appendix consolidates the original Supplement R narrative for the ACES-R experiments reported in Tables 11 and 12. It records the procedures, models, seeds, numerical-precision conventions, and reproducibility checks. The earlier computational supplement is now reproduced in Appendix E (Sections S1–S15, Tables S1–S16, and Figure S1).

D.1. Scope and version.

This appendix documents Tables 11 and 12 and the separate sensitivity panel. The computational supplement supporting Tables 1–10 and the engineering applications is incorporated as Appendix E, Sections S1–S15. The reference implementation and results retained in the previously supplied computation archives remain provenance material, but no additional ZIP file accompanies this integrated Word article. A fresh complete replay of all original simulations and supplemental diagnostics is not claimed. The two engineering applications were replayed separately during revision.

The reference implementation is code/aces_final_v6.py. It requires NumPy and SciPy. It was executed with Python 3.13.5, NumPy 2.3.5 and SciPy 1.17.0 on Linux using float64 arithmetic. The precise environment and source hash are in results/execution_environment.json.

D.2. Reproduction commands.

The commands below document how the experiments were reproduced from the authors' archived code; the executable Python files are not embedded in this Word manuscript: archive, use a new output directory for each fresh replay:

python -m pip install -r code/requirements.txt;

python -m unittest discover -s code -p 'test_*.py' -v;

python code/aces_final_v6.py --replay --out replay_results --reps 400;

replay_results can be compared with the shipped results directory. The deterministic CSV files, including replicate-level output, matched byte for byte in two separate process runs in the recorded environment. The JSON environment files additionally contain elapsed time and therefore are not expected to be byte-identical. No promise of bitwise agreement across arbitrary numerical-library versions or platforms is made.

D.3. Procedures and numerical definitions.

Full ACES uses the original full-sample window on $N \geq 11$, the retained complementary grid, score orders 1, 3/2 and 2, Hodges–Lehmann-centered residual magnitudes, 399 independent bootstrap samples and the +1 upper-tail proportion. Its calibration remains empirical. The code raises an error on nonpositive bootstrap scale, records zero-span flooring and redraws a constant bootstrap sample; neither exceptional event occurred in the final replay.

ACES-R uses blocks formed before within-block sorting and the distinct finite-block rule (51), with b≥8. Every primary cell uses the complete-orbit standardized maximum, not the quadratic-only statistic. Opposite-sign pairs are enumerated through one representative and restored algebraically. The final upper-tail comparison is >= without tolerance. The quadratic Monte Carlo fallback is supplied and tested, but is not the source of Tables 11 and 12. In the revised manuscript K is the block count, B the full-ACES or benchmark resampling count, and L the sign-draw budget in (55). To preserve reproducibility, the reference code and frozen CSV files are unchanged: their historical argument or column B corresponds to manuscript K, and Supplement_B6.csv is the $K = 6$ sensitivity panel.

Numerical precision. Deterministic global-sign pairs are represented algebraically and retain their prescribed multiplicity. Additional mathematical ties can nevertheless be separated by floating-point rounding in the orbit-standardized maximum. A diagnostic with $K = 2$ exhibited such a separation; it is outside the $K = 6$ and $K = 10$ designs of Tables 11 and 12 and is not a counterexample to the mathematical group-rank argument. No ad hoc tolerance is introduced into the inclusive upper-tail comparison. The finite-group theorems concern the mathematically defined statistic; numerical implementations remain subject to floating-point precision.

Table D.1. Prespecified ACES-R block and orbit settings.

N	Blocks K	Block size b	Window m	Grid size	Used / unused	Upper level bound
50	6	8	2	2	48/2	0.031250000
100	10	10	2	3	100/0	0.048828125
200	10	20	4	6	200/0	0.048828125
100 (sensitivity)	6	16	3	5	96/4	0.031250000

These are finite-group level bounds; extra ties may lower the actual size. A realized simulation estimate can exceed the true size because it is a binomial proportion, not a theoretical probability. At nominal $\alpha = 0.05$, the upper attainable complete-orbit levels are 0.03125 for $K = 6$ and 0.048828125 for $K = 10$; these bounds are not universally attained sizes. The exact calculation is enumeration of the finite sign group, while the inferential guarantee is finite-sample level control.

D.4. Models.

Table 11 uses standard normal, standard logistic, standard Laplace, Student-t with 3 degrees of freedom, Uniform($-\sqrt{3},\sqrt{3}$), and an equal-weight mixture N(−2,1)/N(2,1). Alternatives are skew-

normal with shape 3, lognormal with log mean zero and log standard deviation 0.5, Gamma(shape 4, scale 1), Beta(2,5), the mixture 0.9N(0,1)+0.1N(3,1), and Weibull(shape 1.5, scale 1).

The symmetric-mixture null here uses component means −2 and 2, rather than the −1.5 and 1.5 means in the original null panel of manuscript Section 6.1. These are different experimental configurations and are not treated as a same-model comparison. Table 12 uses Student-t3 at $N = 50,100,200$, Student-t10 at $N = 100$, Cauchy at $N = 50,100,200$, skew-t3 at $N = 50,100,200$, and skew-t10 at $N = 100$. The asymmetric generator is

$$X_\nu = \{Z_0 + (2/3)|Z_1|\}/\sqrt{V/\nu},$$

with independent standard normal Z0,Z1 and chi-square V with ν degrees of freedom. This states the actual coefficient and model, rather than interpreting 2/3 as a differently normalized mixture parameter. The old printed t2/t5 labels were not used to describe the executable t3/t10 generators.

D.5. Seeds and uncertainty.

The base cell seeds are 1100+j for Table 11, j=0,…,11; 3100+j for Table 12, j=0,…,10; and 2100+j for the separate sensitivity panel. Replication r=0,…,399 uses data seed 100003 × base seed + r and full-ACES bootstrap seed 1000000 × base seed + r+3. ACES-R enumeration uses no inner random seed. Models and order are fixed by cell_specs() and written to design_and_seeds.csv.

For a rejection count k among R = 400 samples, the printed estimate is k/R and its estimated Monte Carlo standard error is $\sqrt{\{\hat{p}(1-\hat{p})/R\}}$. The maximum possible value is 0.025, and at $\hat{p} = 0.05$ it is about 0.0109. Rates and standard errors are rounded to four decimals only at presentation. For the paired difference, the standard error uses the sample variance of the within-sample difference of the two rejection indicators, not an independence approximation.

Tables 11 and 12 contain 23 paired cells and 9200 generated samples. The sensitivity study adds 12 cells and 4800 samples. The archive consequently contains 14000 sample rows and 23200 actual test p-values. All primary comparisons are raw rejection proportions; they are not equal-size power comparisons or an efficiency theorem. The procedures can have different effective finite-sample null rejection probabilities. ACES-R is a distinct finite-sample-valid core that incurs information loss through blocking, not a validated version of full ACES. Stable calibration under a symmetric heavy-tailed null is not by itself evidence of better power under heavy-tailed asymmetry, and the suggested niche remains exploratory.

D.6. Frozen results.

The following table includes the exact rejection counts underlying every primary printed cell. The shipped CSV files additionally include all standard errors, seeds and design metadata.

Table D.2. Exact primary rejection counts and rates for Tables 11and 12.

Panel / model	N	ACES-R count	ACES count	ACES-R rate	ACES rate
Table11 / Normal	100	22	10	0.0550	0.0250
Table11 / Logistic	100	20	15	0.0500	0.0375
Table11 / Laplace	100	17	9	0.0425	0.0225
Table11 / Student-t(3)	100	20	15	0.0500	0.0375
Table11 / Uniform	100	17	20	0.0425	0.0500
Table11 / Symmetric mixture	100	18	25	0.0450	0.0625
Table11 / Skew-normal(3)	100	126	191	0.3150	0.4775
Table11 / Lognormal(0.5)	100	322	391	0.8050	0.9775
Table11 / Gamma(4)	100	229	299	0.5725	0.7475
Table11 / Beta(2,5)	100	158	229	0.3950	0.5725
Table11 / Contaminated normal	100	131	182	0.3275	0.4550
Table11 / Weibull(1.5)	100	296	362	0.7400	0.9050
Table12 / Student-t(3) null	50	9	16	0.0225	0.0400
Table12 / Student-t(3) null	100	22	13	0.0550	0.0325
Table12 / Student-t(3) null	200	23	9	0.0575	0.0225
Table12 / Student-t(10) null	100	18	13	0.0450	0.0325
Table12 / Cauchy null	50	13	17	0.0325	0.0425
Table12 / Cauchy null	100	22	14	0.0550	0.0350
Table12 / Cauchy null	200	17	19	0.0425	0.0475
Table12 / Skew-t(3)	50	27	43	0.0675	0.1075
Table12 / Skew-t(3)	100	66	79	0.1650	0.1975
Table12 / Skew-t(3)	200	124	140	0.3100	0.3500
Table12 / Skew-t(10)	100	26	22	0.0650	0.0550

D.7. Verification files.

Verification/replay_comparison.json records the hashes and agreement of the revision replay outputs. Verification/tests_final.txt records the regression suite executed during final preparation. Verification/document_checks.json records the final manuscript table check, clean/blue content equivalence, mathematical-content preservation and figure preservation. The archive checksum manifest detects accidental file mixing. Historical and current execution records are labelled separately; reaggregation of stored results is not described as a fresh simulation.

Appendix E. Computational supplement

Reproducibility, calibration checks, and source verification.

Companion to “Cancellation-Free Distributional Symmetry Testing via Reliability-Inspired Consecutive-System Entropy Scores”. This supplement separates exact replay of the recovered numerical record from post-audit verification experiments and documents the computational evidence

retained in the accompanying reproducibility archive.

S1. Recovery, inputs and exact replay.

Recovered numerical record.

The recovered archive supplies the ACES and benchmark functions, the 48-cell simulation driver and outputs, the null-only window script, the application arrays, and the historical repair plotting code. These files are retained unchanged under provenance/recovered_v6. The audited working implementation is code/aces_v12.py.

The complete main experiment was re-executed with the original implementation: 16 models, three sample sizes, and 1000 simulated samples per cell, with 399 bootstrap draws per test. All 192 rejection proportions in Tables 1–4 were reproduced exactly. A second complete execution with the audited code reproduced all 192,000 individual p-values and all 192 proportions without a single different 5% rejection decision.

The 60 null-only window configurations were also re-executed with their original per-cell seeds and sign-before-radius draw order. All 60 rejection proportions matched the recovered detailed output; their summaries reproduce Table 5 at its printed precision. Agreement with saved output establishes replay, not correctness of every modeling assumption or implementation choice.

Preserved source observations

The 45 repair observations are measured in hours. They agree value-for-value with Dataset 2 in Section 6.3 of Chaudhary and Gupta's author paper [60], which explicitly attributes that sample to Qiu and Jia [57]. Direct access to the original Qiu/Jia full text was not obtained. The corroboration route is therefore stated rather than described as direct primary-source verification.

The 50 tensile strengths agree value-for-value and in order with the Nextel 610, lot A0168, strength column (MPa) in Table 2 of Wilson [58]. This check used indexed author-posted full text [58], rather than a newly downloaded publisher PDF. Row-wise comparisons for both datasets are included in results/provenance_checks.

Some other repair-data sources contain a 46-point variant with the additional value 2.7. That observation was not inserted into the primary 45-point analysis. A separately labelled 46-point sensitivity computation is retained in the archive; it also rejects symmetry, without changing the chosen dataset.

Scope boundary

The global mathematical results and the conditional transfer statement are retained. This computational work does not derive a primitive joint central limit theorem or a method-specific proof of symmetrized-bootstrap validity for the paired moving log-spacings. The final computational layer adds an independent empirical-characteristic-function benchmark and a Cauchy stress null; neither computation changes the theoretical scope.

S2. Implementation audit and corrections.

Signed diagnostic orientation.

The manuscript defines

$$C = log\ q.$$

The historical repair script instead plotted

$$log\ q(1-u) - log\ q(u)$$

under the label

$$C(u) - C(1-u).$$

Figure 2 now plots the defining difference

$$log\ q(u) - log\ q(1-u).$$

Its sign reverses, but the absolute discrepancy, all ACES components, and all p-values are unchanged. The discrete weights are labelled W_i to distinguish them from the continuously normalized weight function.

Relative zero-span floor and exceptional cases

The working code replaces only zero quantile-density estimates by the stated relative floor. An absolute machine-epsilon floor tied to a minimum scale of one was removed because it destroyed invariance at very small measurement units. Tests verify the ACES discrepancy after scales of 10^{-30} and 10^{30} and ordinary affine transformations. Constant original samples and nonfinite inputs are errors; constant bootstrap draws are redrawn and counted. A nonpositive bootstrap standard deviation is reported as an error, not silently replaced by one.

Benchmark source convention

Inspection of the CRAN lawstat source [61] shows that the observed sample is sorted and centered before its statistic is evaluated. The recovered Python code used the uncentered observed input. This matters for the Mira formula with odd sample sizes or tied medians. The working code now uses observed-sample centering; resampled statistics are not additionally recentered. An independent scalar evaluation agrees with the corrected observed statistic.

For the 45 repair observations, the Mira statistic changes from 3.3720671268 to 3.4253182325. Under the original application seed stream, the empirical p-value remains 0.0030015008. The tensile statistic and all 192,000 main-study p-values are unchanged. No executed cross-language or RNG-parity check was performed; this verification is a source/formula audit followed by Python replay.

Why the column is labelled Mira-G

The recovered implementation uses $max(S, 10^{-12})$ inside the Mira denominator; the inspected R formula does not contain this protection. In a dedicated audit of 812 input samples, each benchmark involved 5,788,463 statistic evaluations over its candidate resample sizes. The Mira guard activated 6,506 times, all at negative S; no MGG or CM denominator guard activated. The reported Mira column is therefore labelled Mira-G, an explicitly specified numerical variant. It is not presented as the unmodified method.

A triggered sensitivity check of this numerical guard is reported in Table S6. Its results do not license relabelling Mira-G as unmodified Mira. Both the source URL and the exact independent audit script are included so the distinction remains inspectable.

S3. Independent larger-bootstrap null checks

The additional null experiment was fixed before its results were inspected: the same six null families, $N = 100, R = 1000$, and $B = 1999$. Each model uses an independent seed, 920000 through 920005. The samples differ from those in the main $B = 399$ experiment; differences between the two experiments must not be attributed solely to the bootstrap count.

Table S1. ACES null rejection rates with 1999 bootstrap draws.

Null model	Rate	MCSE	95% Wilson interval	Floor uses
Normal	0.035	0.0058	[0.025, 0.048]	0
Logistic	0.038	0.0060	[0.028, 0.052]	0
Laplace	0.036	0.0059	[0.026, 0.049]	0
Student-t(3)	0.037	0.0060	[0.027, 0.051]	0
Uniform	0.044	0.0065	[0.033, 0.059]	0
Symmetric mixture	0.066	0.0079	[0.052, 0.083]	0

Intervals are pointwise binomial Wilson intervals for 1000 independent outer replications; they do not establish uniform size over the composite null. The five unimodal models remain conservative, with rates 0.035–0.044. The symmetric mixture gives 0.066, with a pointwise interval of approximately [0.052, 0.083]. This is evidence of residual over-rejection in that setting, not exact nominal calibration.

No zero-span floor activation or constant-resample redraw occurred in these six runs. This records what happened in the sampled configurations; it is not a proof that empirical-bootstrap ties are harmless for other sample sizes, rounding levels, or distributions.

Interpretation

The original mixture rates of 0.051, 0.053, and 0.059 remain reproducible for the original experiment. The new rate 0.066 is retained alongside them, rather than used to replace a less favorable observation or to retune the spacing multiplier. The narrative consequently describes comparative finite-sample performance, not general multimodal invariance or exact size control.

These checks assess finite-sample calibration only. They do not establish the high-level joint convergence condition in the manuscript, and increasing the number of bootstrap draws reduces Monte Carlo granularity without repairing every approximation error in the null resampling law.

S4. Reference-specific size-matched diagnostic

Two reference laws were fixed: standard normal, and

$$0.5N(-1.5,1) + 0.5N(1.5,1).$$

For each reference, 4000 independent samples of size 100 were generated and all four implemented procedures were evaluated with $B = 399$. Master calibration seeds 910001 and 910002 each spawn four disjoint 1000-sample streams. These training observations are not reused for size validation or power evaluation.

For each method and reference, choose the smallest p-value cutoff t whose training cumulative proportion reaches 0.05. If L training p-values are below t and E equal t, reject at equality with probability

$$\rho = (0.05K - L)/E$$

where $K = 4000$. Thus the conditional expected training rejection fraction is 5%. Before identifying ties, probabilities are mapped to their exact tail-count lattice: multiples of $1/400$ for ACES and $2/399$ for the benchmarks. This prevents floating-point variants of the same count from being treated as different probabilities.

Table S2. Reference-specific cutoffs and boundary randomization.

Reference	Method	p cutoff	Tail count	ρ
Normal	ACES	0.070000	28	0.166667
Normal	MGG	0.055138	11	0.866667
Normal	CM	0.050125	10	0.391304
Normal	Mira-G	0.045113	9	0.285714
Mixture	ACES	0.055000	22	0.090909
Mixture	MGG	0.010025	2	0.017857
Mixture	CM	0.005013	1	0.309859
Mixture	Mira-G	0.025063	5	0.038462

The count column is authoritative: divide the ACES count by 400, or multiply a benchmark count by 2/399, to recover the cutoff without decimal-rounding ambiguity. The working script retains the full precision and the exact training counts.

Independent evaluation design

Achieved size is evaluated on the corresponding 1000-sample $N = 100$ main-study cell, whose seed stream is disjoint from calibration. Normal-reference power is evaluated for the prespecified skew-normal(3), Gamma(4), and contaminated-normal alternatives. The mixture-reference alternative is $0.75N(-1.5,1) + 0.25N(1.5,1)$, generated independently with seed 910003. Evaluation boundary randomization uses its own seeds 950000 onward.

This procedure is an oracle reference-specific diagnostic. It is not recommended as a universal replacement for composite-null calibration when the symmetric parent law is unknown. Estimated cutoffs retain training uncertainty; validation rates, rather than assumed exact size, are reported next.

S4. Independent achieved size and power (continued).

Table S3. Reference-calibrated rejection proportions on independent evaluation samples.

Ref.	Evaluation model	Target	ACES	MGG	CM	Mira-G
N	Normal	Size	0.054	0.050	0.046	0.049
M	Symmetric mixture	Size	0.059	0.044	0.048	0.043
N	Skew-normal(3)	Power	0.553	0.427	0.395	0.355
N	Gamma(4)	Power	0.782	0.602	0.584	0.537
N	Contaminated normal	Power	0.515	0.435	0.413	0.383
M	Asymmetric mixture (0.75/0.25)	Power	0.911	0.515	0.466	0.685

N denotes the normal reference and M the symmetric-mixture reference. Every evaluation uses 1000 samples of size 100 and 399 inner bootstrap draws. All four alternatives fixed in the protocol are reported. No model was omitted after its results were observed.

Under the normal reference, achieved validation sizes are 0.054, 0.050, 0.046, and 0.049 for ACES, MGG, CM, and Mira-G. Under the mixture reference they are 0.059, 0.044, 0.048, and 0.043. The procedures are therefore closer to a common target in this experiment, but their achieved sizes are estimated and are not identical.

ACES has the largest calibrated rejection rate for each of the four selected alternatives. For example, the skew-normal(3) values are 0.553 for ACES and 0.427 for MGG; the asymmetric-mixture values are 0.911 and 0.515. These findings support comparative power within the declared reference-specific design, not uniform superiority over all alternatives or symmetric null shapes.

Uncertainty and reporting

The maximum binomial Monte Carlo standard error for a 1000-sample rejection proportion is about 0.0158. The machine-readable table includes each rate's MCSE and pointwise Wilson interval, conditional on the estimated cutoff. A second file reports paired ACES-minus-competitor differences and their standard errors on the common evaluation samples. These conditional summaries do not include the uncertainty from estimating cutoffs on the finite training panel.

The archive also retains the raw, uncalibrated rejection rates on exactly the same evaluation samples. The calibrated comparison is additional evidence; Tables 1–4 are not retrospectively relabelled as size-adjusted power.

S5. Application replay and numerical sensitivity.

Table S4. Application p-values replayed using the original seed stream.

Dataset	N	ACES	MGG	CM	Mira-G
Repair times	45	0.0005	0 (empirical)	0 (empirical)	0.0030
Tensile strength	50	0.6975	0.8914	0.8884	0.8764

These values use $B = 1999$ and one Generator stream initialized at 20260619: repair first, then tensile; within each sample, ACES, MGG, CM, then Mira-G. A zero empirical benchmark tail frequency is not an exact zero probability. ACES uses the +1 correction; its repair result has zero exceedances and

$$p=1/2000.$$

Table S5. Independent ACES checks at two bootstrap counts.

Dataset	B	Seed	p	Exceedances	Floor uses
Repair times	1999	930000	0.0005	0	0
Repair times	9999	930000	0.0002	1	0
Tensile strength	1999	930001	0.6995	1398	0
Tensile strength	9999	930001	0.7010	7009	0

Each row was repeated at zero-span floor multipliers 0.5, 1, and 2 using identical random draws. The p-values and component standardization were unchanged across those multipliers. The reason is explicit: no local quantile-density estimate was zero in the original samples or any of these bootstrap batches. These examples therefore do not validate a floor rule under heavy rounding.

The larger-bootstrap repair estimate is 0.0002 with one exceedance among 9999 draws. The tensile estimate is 0.7010. Both decisions agree with the replay, without interpreting those numerical differences as a change in the scientific conclusion.

The 46-point repair-variant check is deliberately separate: the extra value 2.7 was added only for that sensitivity, using $B = 9999$ and seed 930046. The primary data files, main results, and Figure 2 continue to use the documented 45 observations.

S5. Complementary local diagnostic (continued).

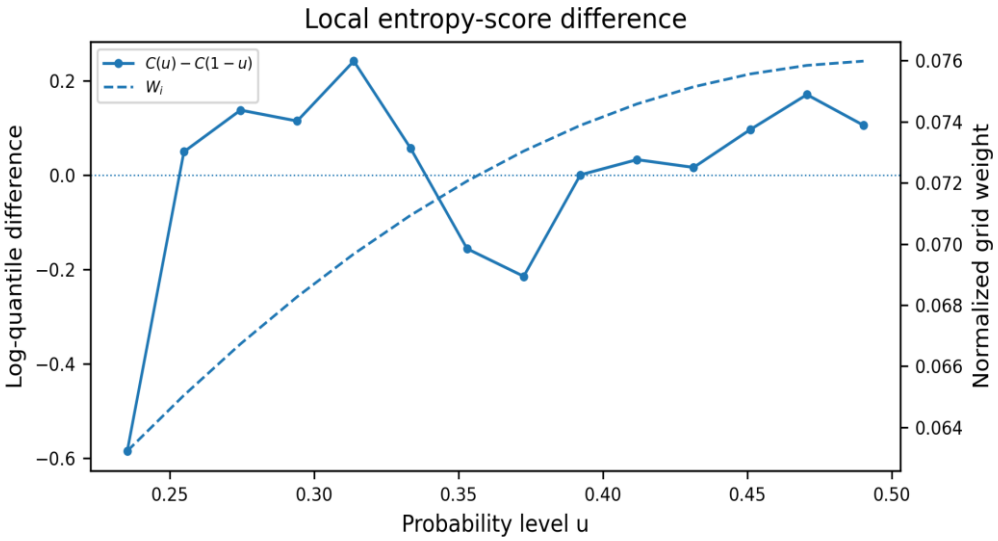


Figure S1. Signed complementary log-quantile-density difference for the tensile strengths and normalized discrete weights. This is a descriptive curve, not a simultaneous confidence band or an independent test at each probability level.

S6. Triggered Mira-G guard sensitivity.

After the denominator audit, a separate numerical sensitivity check compared variance floors 10^{-12}, 10^{-8}, and 10^{-4}. Identical data and resampling streams were reused across floors within each setting. This check was triggered by the audit and is distinguished from the prespecified larger-bootstrap and reference-calibration experiments.

Table S6. Mira-G rejection rates across three variance floors.

Model	N	Floor 1e−12	Floor 1e−8	Floor 1e−4
Normal	20	0.029	0.029	0.029
Symmetric mixture	20	0.053	0.053	0.053
Normal	100	0.055	0.055	0.055
Symmetric mixture	100	0.085	0.085	0.085
Skew-normal(3)	20	0.080	0.080	0.080
Skew-normal(3)	100	0.384	0.384	0.384

Each setting uses 1000 samples and 399 bootstrap draws. Rejection rates at 5% are identical across these floors in all six settings. This finite check supports numerical stability for the tested range only; the code is still a guarded variant and not an executed reproduction of the unmodified R procedure.

S7. Reproduction and limits of verification.

Run commands from the unzipped archive root after installing the recorded dependencies. The environment.json file records the versions actually used, and requirements.txt gives the tested Python-

package environment. The R source-check script is optional and is explicitly marked as not executed in the archived verification session.

```
python -m unittest discover -s tests -v
python code/run_main_simulation.py --version v6 --workers 4
python code/run_main_simulation.py --version v12 --workers 4
python code/run_window_verification.py
python code/run_additional_experiments.py --workers 4
python code/summarize_verification.py
python code/verify_applications.py
python code/check_mira_guard_sensitivity.py
python code/verify_source_values.py
python code/make_figures.py
python code/run_v13_extropy_comparison.py --workers 6
python code/run_v13_ablations.py --workers 6
python code/run_v13_largeN_rounding.py --mode both --workers 6
python code/run_v13_portfolio_calibration.py --workers 8
python code/run_v14_hkm_comparison.py --workers 6 --cauchy-R 1000
```

Main replay drivers reuse a complete matching cell file when it already exists. Delete or move the corresponding results directory for a fresh computation; do not mix different R/B settings in one directory. Every numerical table can be traced to a CSV and an explicit RNG protocol. SHA256SUMS.json records the packaged file hashes.

What these checks do not establish

They do not prove a primitive joint null central limit theorem, primitive symmetrized-bootstrap validity, exact distribution-free finite-sample calibration, or general robustness to multimodality and heavy rounding. The additional verification program includes direct predecessor comparison, score/weight/architecture ablations, larger-N experiments, rounding stress tests, balanced portfolio calibration, the HKM characteristic-function benchmark, and a Cauchy stress null. These computations strengthen the empirical map of the method but do not substitute for the open paired-overlapping-spacing limit theory.

S8. Direct consecutive-system extropy comparison.

The source audit for [30] identified two different window statements in the published power section. The direct comparison therefore reports both floor(sqrt(N)) and floor(N/2)-1, each recalibrated from 50,000 draws under the published symmetric GLD Case 1 reference. The primary manuscript uses the sqrt(N) reading only as the most explicit Monte Carlo convention; neither reading is represented as original-code parity.

Table S7. N=100 size and power for ACES and the two extropy window readings.

Model	ACES	Extropy sqrt(N)	Extropy N/2-1
Normal	0.027	0.045	0.059
Logistic	0.039	0.074	0.134
Laplace	0.043	0.103	0.201
Student-t(3)	0.030	0.137	0.365
Uniform	0.043	0.232	0.022
Symmetric mixture	0.061	0.191	0.040
Skew-normal(3)	0.471	0.678	0.638
Skew-normal(6)	0.820	0.961	0.948
Lognormal(0.5)	0.963	0.998	0.998
Lognormal(1.0)	1.000	1.000	1.000
Gamma(2)	0.976	0.999	0.999
Gamma(4)	0.761	0.947	0.962
Beta(2,5)	0.569	0.880	0.886
Contaminated normal	0.449	0.548	0.615
Weibull(1.5)	0.912	0.996	1.000
Reflected Weibull(1.5)	0.914	0.996	0.999

Raw extropy power is intentionally reported alongside its null distortion. No size-adjusted superiority claim is made from Table S7.

S9. Score, weight, and architecture ablations.

All variants are evaluated on the same simulated sample and bootstrap draws within each cell. The baseline remains the original ACES max over beta={1,1.5,2} with the (3,1) system weight. Alternative designs are diagnostics only.

Table S8. N=100 ablation rejection rates for all asymmetric alternatives.

Alternative	Baseline	beta=1	beta=1.5	beta=2	Uniform	Arch.(5,1)
Skew-normal(3)	0.467	0.440	0.467	0.470	0.480	0.513
Skew-normal(6)	0.813	0.763	0.797	0.820	0.820	0.843
Lognormal(0.5)	0.957	0.927	0.943	0.967	0.960	0.973
Lognormal(1.0)	1.000	1.000	1.000	1.000	1.000	1.000
Gamma(2)	0.973	0.957	0.967	0.977	0.977	0.983
Gamma(4)	0.740	0.673	0.720	0.763	0.760	0.803
Beta(2,5)	0.543	0.500	0.527	0.557	0.550	0.607
Contaminated normal	0.423	0.390	0.410	0.433	0.457	0.487
Weibull(1.5)	0.923	0.833	0.897	0.933	0.937	0.950
Reflected Weibull(1.5)	0.907	0.847	0.887	0.917	0.923	0.943

The quadratic component wins the standardized baseline maximum most often under the stronger alternatives, but no score or weight is uniformly best. The paired-difference CSV in the archive gives common-random-number standard errors for every comparison.

S10. Larger-N and rounding diagnostics.

Integer rounding is deliberately severe and often activates the floor tens of thousands of times across the bootstrap batches. The 0.1-unit and two-significant-digit scenarios are far less disruptive in the sampled configurations. These results do not extend the continuous theory to discrete data.

Table S9. Larger-N ACES rejection rates; mixture cells use 1000 outer replications and all other cells use 500.

Model	N	Rate	Wilson lo	Wilson hi	Bootstrap floor rows
Normal	200	0.038	0.024	0.059	0
Logistic	200	0.042	0.028	0.063	0
Laplace	200	0.086	0.064	0.114	0
Student-t(3)	200	0.032	0.020	0.051	0
Uniform	200	0.046	0.031	0.068	0
Symmetric mixture	200	0.054	0.042	0.070	0
Skew-normal(3)	200	0.774	0.735	0.808	0
Lognormal(1.0)	200	1.000	0.992	1.000	0
Gamma(2)	200	1.000	0.992	1.000	0
Weibull(1.5)	200	1.000	0.992	1.000	0
Normal	500	0.036	0.023	0.056	0
Logistic	500	0.032	0.020	0.051	0
Laplace	500	0.046	0.031	0.068	0
Student-t(3)	500	0.046	0.031	0.068	0
Uniform	500	0.028	0.017	0.046	0
Symmetric mixture	500	0.047	0.036	0.062	0
Skew-normal(3)	500	1.000	0.992	1.000	0
Lognormal(1.0)	500	1.000	0.992	1.000	0
Gamma(2)	500	1.000	0.992	1.000	0
Weibull(1.5)	500	1.000	0.992	1.000	0

Table S10. N=100 rounding stress test. Floor counts are totals across the 300 outer samples and their bootstrap batches.

Model	Rounding	Rate	Observed floor rows	Bootstrap floor rows
Normal	0dp	0.007	101	44946
Laplace	0dp	0.023	112	47795
Uniform	0dp	0.020	8	8704
Symmetric mixture	0dp	0.047	0	4
Skew-normal(3)	0dp	0.150	298	115063
Lognormal(1.0)	0dp	1.000	176	44266
Gamma(2)	0dp	0.770	37	5741
Weibull(1.5)	0dp	0.460	298	116982
Normal	1dp	0.033	0	0
Laplace	1dp	0.043	0	0
Uniform	1dp	0.043	0	0
Symmetric mixture	1dp	0.053	0	0
Skew-normal(3)	1dp	0.483	0	0
Lognormal(1.0)	1dp	1.000	0	0
Gamma(2)	1dp	0.987	0	0
Weibull(1.5)	1dp	0.880	0	0
Normal	2sig	0.053	0	0
Laplace	2sig	0.047	0	0
Uniform	2sig	0.060	0	0
Symmetric mixture	2sig	0.060	0	0
Skew-normal(3)	2sig	0.463	0	0
Lognormal(1.0)	2sig	1.000	0	0
Gamma(2)	2sig	0.967	0	0
Weibull(1.5)	2sig	0.920	0	0

S11. Balanced portfolio equal-target calibration.

The portfolio uses 4000 training samples in total, balanced across Normal, Student-t(3), Uniform, and the symmetric mixture at N=100. The training set is disjoint from every evaluation cell. A single cutoff per method and target level is learned from the pooled portfolio, with randomized treatment of cutoff ties.

Table S11. Portfolio cutoffs and boundary randomization probabilities.

Target	Method	Cutoff	rho	Training N
0.010	ACES	0.013	0.600	4000
0.010	MGG	0.000	0.635	4000
0.010	Extropy	0.000	0.273	4000
0.025	ACES	0.033	0.154	4000
0.025	MGG	0.005	0.810	4000
0.025	Extropy	0.002	0.667	4000
0.050	ACES	0.055	0.800	4000
0.050	MGG	0.030	0.083	4000
0.050	Extropy	0.006	1.000	4000
0.100	ACES	0.107	1.000	4000
0.100	MGG	0.075	0.250	4000
0.100	Extropy	0.021	1.000	4000

Table S12. Portfolio-calibrated rejection rates at target 0.05 for all evaluation models.

Model	Target	ACES	MGG	Extropy
Normal	Size	0.047	0.028	0.010
Logistic	Size	0.043	0.031	0.009
Laplace	Size	0.048	0.021	0.028
Student-t(3)	Size	0.035	0.044	0.052
Uniform	Size	0.055	0.048	0.085
Symmetric mixture	Size	0.060	0.086	0.044
Skew-normal(3)	Power	0.512	0.310	0.377
Skew-normal(6)	Power	0.836	0.566	0.840
Lognormal(0.5)	Power	0.971	0.844	0.978
Lognormal(1.0)	Power	1.000	1.000	1.000
Gamma(2)	Power	0.976	0.813	0.994
Gamma(4)	Power	0.760	0.480	0.747
Beta(2,5)	Power	0.563	0.307	0.650
Contaminated normal	Power	0.438	0.275	0.242
Weibull(1.5)	Power	0.916	0.647	0.973
Reflected Weibull(1.5)	Power	0.909	0.654	0.967

Full portfolio size-power curves are stored in machine-readable form at target levels 0.01, 0.025, 0.05, and 0.10. The diagnostic is non-oracle in the sense that one fixed training portfolio is used for all evaluation models; it does not establish uniform level control over the composite symmetry null.

S12. Additional provenance and interpretation boundaries.

The additional verification layer is additive: the recovered source implementation, the audited working implementation, and all original replay outputs remain unchanged. Its scripts and results use disjoint SeedSequence streams and separate results/v13_* directories. The extropy comparator [30] is source-audited but not original-code-validated; the ablations are exploratory diagnostics rather than tuning; the larger-N and portfolio studies are independent experiments; and the rounding study is a stress test outside the intrinsic-discreteness scope of the theory. HKM and Cauchy evidence is stored separately in results/v14_hkm/.

S13. Independent HKM characteristic-function benchmark.

Henze et al. [27] propose affine-invariant omnibus symmetry tests based on the imaginary part of the empirical characteristic function and a conditional permutation procedure. The univariate implementation used here takes a=3, B=399 Rademacher sign permutations, and 32-node Gauss-Hermite quadrature for the weighted L2 permutation norm. A regression test verifies that the all-positive-sign quadrature agrees with the closed-form observed statistic to numerical tolerance, and separate tests check affine invariance and the Cauchy generator.

Table S13. Raw nominal-0.05 rejection rates for the four methods on the N=100 portfolio evaluation samples.

Model	Target	ACES	MGG	Extropy	HKM
Normal	Size	0.043	0.045	0.069	0.059
Logistic	Size	0.039	0.049	0.077	0.062
Laplace	Size	0.039	0.042	0.111	0.064
Student-t(3)	Size	0.031	0.082	0.154	0.066
Uniform	Size	0.048	0.068	0.216	0.060
Symmetric mixture	Size	0.057	0.109	0.165	0.048
Skew-normal(3)	Power	0.500	0.394	0.680	0.757
Skew-normal(6)	Power	0.823	0.638	0.958	0.966
Lognormal(0.5)	Power	0.966	0.886	0.999	0.996
Lognormal(1.0)	Power	1.000	1.000	1.000	1.000
Gamma(2)	Power	0.972	0.857	1.000	0.999
Gamma(4)	Power	0.737	0.553	0.930	0.942
Beta(2,5)	Power	0.547	0.381	0.880	0.849
Contaminated normal	Power	0.419	0.362	0.547	0.784
Weibull(1.5)	Power	0.909	0.717	0.997	0.991
Reflected Weibull(1.5)	Power	0.902	0.730	0.996	0.992

The raw table is not an equal-size ranking. In particular, the extropy and MGG columns show null-dependent size distortion. HKM is close to nominal on the six listed N=100 nulls but has the largest raw power for many smooth alternatives. The full paired p-value files are retained under results/v14_hkm/evaluation/.

S14. Four-method balanced portfolio calibration.

The four-method portfolio uses exactly the same 4000 training samples and the same independent 1000-sample evaluation cells as the earlier three-method diagnostic, but learns one

boundary-randomized cutoff for each of four methods: ACES, MGG, extropy, and HKM. New boundary-randomization streams are documented and do not overwrite the earlier result files.

Table S14. Four-method portfolio cutoffs.

Target	Method	Cutoff	rho	Training N
0.010	ACES	0.012500	0.600000	4000
0.010	MGG	0.000000	0.634921	4000
0.010	Extropy	0.000280	0.272727	4000
0.010	HKM	0.005000	0.692308	4000
0.025	ACES	0.032500	0.153846	4000
0.025	MGG	0.005013	0.809524	4000
0.025	Extropy	0.001520	0.666667	4000
0.025	HKM	0.015000	0.454545	4000
0.050	ACES	0.055000	0.800000	4000
0.050	MGG	0.030075	0.083333	4000
0.050	Extropy	0.005820	1.000000	4000
0.050	HKM	0.040000	0.300000	4000
0.100	ACES	0.107500	1.000000	4000
0.100	MGG	0.075188	0.250000	4000
0.100	Extropy	0.020560	1.000000	4000
0.100	HKM	0.090000	0.222222	4000

Table S15. Portfolio-calibrated rejection rates at target 0.05 for all evaluation models.

Model	Target	ACES	MGG	Extropy	HKM
Normal	Size	0.047	0.027	0.010	0.048
Logistic	Size	0.043	0.032	0.009	0.050
Laplace	Size	0.047	0.021	0.028	0.053
Student-t(3)	Size	0.036	0.043	0.052	0.055
Uniform	Size	0.055	0.048	0.085	0.045
Symmetric mixture	Size	0.061	0.085	0.044	0.037
Skew-normal(3)	Power	0.513	0.311	0.377	0.705
Skew-normal(6)	Power	0.838	0.567	0.840	0.955
Lognormal(0.5)	Power	0.971	0.844	0.978	0.996
Lognormal(1.0)	Power	1.000	1.000	1.000	1.000
Gamma(2)	Power	0.976	0.815	0.994	0.997
Gamma(4)	Power	0.758	0.479	0.747	0.931
Beta(2,5)	Power	0.563	0.309	0.650	0.812
Contaminated normal	Power	0.439	0.274	0.242	0.744
Weibull(1.5)	Power	0.917	0.647	0.973	0.986
Reflected Weibull(1.5)	Power	0.909	0.655	0.967	0.987

HKM has the largest calibrated rejection rate for most smooth alternatives in this experiment.

ACES exceeds MGG throughout the alternative panel and occupies a different robustness/calibration domain. These are finite-sample domain-of-advantage findings; the pooled portfolio does not establish uniform composite-null level control for any method.

S15. Cauchy stress null and final provenance.

Cauchy samples are symmetric but have no finite variance. They are therefore used only as a deliberately severe heavy-tail stress test, particularly for procedures that standardize by sample scale. Each cell uses 1000 independent samples and B=399 inner resampling or sign-permutation draws. MGG was not extended to N=500 in this stress protocol because its candidate-resampling implementation is substantially more expensive.

Table S16. Cauchy-null stress rejection rates with 95% Wilson intervals.

N	Method	Rate	95% Wilson interval	R
50	ACES	0.035	[0.025, 0.048]	1000
50	HKM	0.357	[0.328, 0.387]	1000
50	MGG	0.368	[0.339, 0.398]	1000
100	ACES	0.032	[0.023, 0.045]	1000
100	HKM	0.280	[0.253, 0.309]	1000
100	MGG	0.423	[0.393, 0.454]	1000
200	ACES	0.031	[0.022, 0.044]	1000
200	HKM	0.105	[0.087, 0.126]	1000
200	MGG	0.414	[0.384, 0.445]	1000
500	ACES	0.043	[0.032, 0.057]	1000
500	HKM	0.138	[0.118, 0.161]	1000

ACES remains between 0.031 and 0.043 across N=50–500 in this stress experiment. HKM is substantially oversized in these Cauchy cells, and MGG is substantially oversized where evaluated. The finding is reported as finite-sample robustness evidence only; it does not redefine the theoretical scope of the external comparator or establish general heavy-tail validity of ACES.

The final computational layer is additive. The recovered source, audited replay, and earlier diagnostic outputs remain unchanged. The submission archive includes code/aces_v14.py, code/run_v14_hkm_comparison.py, tests/test_v14_hkm.py, and results/v14_hkm/. After removal of exploratory asymptotic-constant checks that are not manuscript claims, the submission regression suite contains 42 passing tests in the recorded environment.



AIMS Press

© 2026 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0)