Powerful statistical tests for ordered data

Jürgen Köfinger^{*} and Gerhard Hummer[†]

Ordered data are ubiquitous, yet statistical tests tend to ignore this order or do not make full use of it. Here we develop powerful non-parametric tests for the statistics of the sign order of the residuals based on their run-length distributions. Compared to Pearson's χ^2 test, the $h^\pm$ -statistic and its combination with the χ^2 -statistic, $(\chi^2, h^\pm)$, have superior statistical power over orders of magnitudes of the numbers of data points.

One-dimensional ordered data abound, most prominently as a function of time. Nevertheless, widely used statistical tests are blind to order. Pearson's¹ χ^2 , for example, depends only on the absolute values of the residuals $|r_i| = |f_i - d_i|$ between model f_i and data d_i for $i = 1, \dots, N$. Yet, practitioners do not take a reasonable χ^2 value and the resulting high score in a goodness-of-fit test at face value. In a careful inspection of the residuals, they search for recognizable patterns to pinpoint systematic deviations or hidden correlations in the data. Such systematic deviations indicate that one or more of the underlying assumptions might be wrong. This intuition is partly quantified in the runs test of Wald and Wolfowitz², which uses the number of runs as a test statistic. A run is defined by all consecutive residuals of equal sign $s_i = \text{sign}(r_i)$. Similarly, the length of the longest run of consecutive signs³, as used in the CorMap tool⁴, for example, only relies on the signs of the residuals. Sign-based tests have the advantage that they are independent of models and estimates for the magnitude of the statistical errors. However, the runs test and the longest run test both probe only a small part of the information on systematic deviations contained in the ordered signs. Thus, they perform poorly as tests for randomness in many cases of interest (see Fig. S1).

Here we introduce two test statistics based on the full distribution of run lengths. For the h and $h^\pm$ statistic, we collect the length of runs of positive and negative signs in a single histogram and in separate histograms, respectively (see Fig. 1a-d). We derived analytical expressions $p(h)$ and $p(h^\pm)$ for the probabilities of these statistics (Online Methods). For notational simplicity, we use $p(x)$ to indicate both probabilities and probability densities, which are distinguished by the argument. The h and $h^\pm$ statistics are independent of the error model, making them robust with respect to both the magnitude and detailed distribution of the residuals and thus independent of the χ^2 statistic. Therefore, they can be combined with Pearson's χ^2 statistic into the (χ^2, h) and $(\chi^2, h^\pm)$ statistics, with overall probabilities that factorize into those of

the individual statistics, i.e., $p(\chi^2, h) = p(\chi^2)p(h)$ and $p(\chi^2, h^\pm) = p(\chi^2)p(h^\pm)$ (Online Methods). Despite being non-parametric, the h and $h^\pm$ tests have high statistical power when applied alone, and even higher when applied in combination with a χ^2 test, as we show in the following.

As a measure of surprise for how untypical or extreme an observed run length distribution is, we use their probabilities $p(h)$ and $p(h^\pm)$, similar to the Fisher-Irwin test.^{6,7} Typical distributions close to the expected run length distribution⁵ $\propto 2^{-l}$, where l is the run length, have high probability. Atypical distributions h and $h^\pm$ with, for example, many long runs will have low probabilities. The cumulative distribution function $\text{cdf}(\mathcal{I})$ of the Shannon information⁸ $\mathcal{I} = -\ln p$ corresponding to the discrete probabilities p then defines a P-value. We combine h and $h^\pm$ with χ^2 by following Jaynes⁹, extending the definition of typicality to probability densities $p(x)$ of random continuous variables x . We define the Shannon information as $\mathcal{I}(x) = \ln m(x)/p(x)$, where $m(x)$ is the invariant measure. With this definition, both Shannon information and P-values are invariant under variable transformations. For the dimensionless χ^2 statistic, we use $m(\chi^2) = \text{const.}$, which does not effect the ranking of states. Thus, we write in the following $\mathcal{I} = -\ln p$ for probabilities and probability densities. However, if a variable transformation is applied, $m(x)$ has to be taken into account. With this definition, we can calculate P-values for the multidimensional (χ^2, h) and $(\chi^2, h^\pm)$ statistics. For χ^2 we note that in this way the double-sided P-value is approached with increasing data size, where $p(\chi^2)$ becomes more symmetric.

The P-values for the h and $h^\pm$ statistic can be calculated by exact enumeration of the partitions of the integer N or, approximately, from the probabilities of run length distributions estimated from randomly generated sign configurations. For the (χ^2, h) and $(\chi^2, h^\pm)$ statistics, we can use the exact χ^2 statistic or draw residuals from the normal distribution to calculate χ^2 , and multiply their probabilities with $p(h)$ and $p(h^\pm)$, respectively. We calculate the Shannon information values and collect them in cumulative distribution functions (Online Methods and Fig. S2).

For all test statistics considered here, shifted gamma distributions with three parameters accurately describe the distribution of the Shannon information for a given number of data points (Figs. S3, S4, and S5). Using B-splines¹⁰ to represent the dependence of the gamma distribution parameters on the number of data points N , we can calculate P-values for all of our test statistics from $N = 50$ to $N = 100000$ accurately (see Figs. S5 and S6). In the Supporting Information, we show that

FIG. 1. The run length distributions $h^\pm = (h^+, h^-)$ of the signs of the residuals, and their combination with Pearson's χ^2 -statistic, $(\chi^2, h^\pm)$, sensitively detect systematic deviations not detected by common alternative tests. (a-d) Illustration of run length distributions underlying the $h^\pm$ test statistic. (a) To assess the agreement of a model (blue line) with the data (orange symbols with errorbars), we (b) plot the residuals scaled by the inverse standard error of the mean. (c) The signs of the residuals show runs of consecutive sign values (red: positive, black: negative). (d) The multiplicities of the run lengths l of positive (h^+ , red circles) and negative signs (h^- , black squares) fluctuate around their most probable values⁵ $2^{-l}/(1 - 2^{-N/2})N/4$ (gray bars). (e-j) We resolve six different models (top panels) with $N = 500$ data points and add Gaussian noise with zero mean and unit standard deviation. In panel (e-j), we show top to bottom the models, the residuals, the signs of the residuals, and a bar plot of the P-values for the χ^2 , $(\chi^2, h^\pm)$, $h^\pm$ test, the runs test of Wald and Wolfowitz², r , and the longest run test, $l_{\max}$ ³.

the gamma distribution form of the Shannon information distribution is a general feature.

The $h^\pm$ test can detect systematic deviations masked by common tests, as illustrated in Fig. 1e-j, which show examples for the true model 1 and five alternative models 2-6. The true model is exactly zero such that the models 2-6 in Fig. 1f-j also represent the residuals without noise. We added uncorrelated Gaussian noise with zero mean and unit standard deviation to models 1-6 to create $N = 500$ ordered data values each and then applied the different statistical tests. The χ^2 , $(\chi^2, h^\pm)$ and $h^\pm$ tests as well as the runs test² r and the longest run test³ $l_{\max}$ all give a high P-value to the true model. However, for the alternative models 2-5, the $(\chi^2, h^\pm)$ and $h^\pm$ tests, which account for data order, consistently outperform both the χ^2 test and the other order-based tests, with P-values that are at least two orders of magnitude

smaller in all but one case. $h^\pm$ also dominates in the combined $(\chi^2, h^\pm)$ test, with similarly low P-values for $h^\pm$ and $(\chi^2, h^\pm)$ tests.

The $h^\pm$ and $(\chi^2, h^\pm)$ tests have superior statistical power over orders of magnitudes in the number of data points (Fig. 2). We performed a systematic comparison to the statistical power of the χ^2 -test for the models depicted in Fig. 1e-j by varying the standard errors of the mean and evaluating the power correlation for selected numbers of data points $N = 500, 5000$, and 50000 at significance level $\alpha = 0.05$ (see Fig. 2a-c). For all models considered here, the h and $h^\pm$ tests have superior statistical power compared to the runs test of Wald and Wolfowitz² for $N \gtrsim 50$ (Figs. S7 and S8). We observe that the h and $h^\pm$ test outperform Schilling's longest-run test³ for $N = 50$ for three out of the five alternative models and for $N \gtrsim 100$ for all models (Figs. S9 and S10).

FIG. 2. The $h^\pm$ and $(\chi^2, h^\pm)$ tests have superior statistical power compared to the χ^2 -statistic over orders of magnitude in the number of data points. (a-c) Power correlation for the $h^\pm$ -statistic (left column) and $(\chi^2, h^\pm)$ -statistic (right column) with the power of the χ^2 -statistics for (a) $N = 500$, (b) $N = 5000$, (c) $N = 50000$ data points for the six models shown in Fig.1 and a significance level $\alpha = 0.05$. Points of the same color represent different noise values, with noise increasing from right to left and top to bottom. The lines in the correlations plot are fits based on an empirical equation describing the dependence of the statistical power on the noise (Online Methods). Note that for the true model m_1 (blue circles) the statistical power for both tests is equal to the significance level α , corresponding to a flat P-value distribution. (d-g) For these six models (colors) and the $h^\pm$ and $(\chi^2, h^\pm)$ tests, we show the β -risk ratios with respect to the χ^2 test as functions of the number of data points. We fixed the β -risk of the χ^2 -test as $\beta(N; \chi^2) = 4\alpha$ for the significance level $\alpha = 0.05, 0.01, 0.005, 0.001$ and evaluated the β -risk of our test statistics for the corresponding noise levels. The black horizontal lines indicate a β -risk ratio of one where both tests have equal power. Due to numerical limitations some β -risk values are zero and thus not shown in these log-log plots.

The h and $h^\pm$ tests alone or in combination with χ^2 also outperform the χ^2 test alone.

The relative performance of run-length distribution-based tests improves with increasing data size N . The β -risk of incorrectly accepting the wrong model is given by one minus the statistical power. We use β to quantify the dependence of the statistical power on the number of data points. The logarithm of the β -risk conveniently resolves the interesting region of the statistical power close to one, where tests can detect statistical differences reliably. Specially, we are interested in the β -risk of a test T evaluated for noise levels at which the χ^2 test is considered to be powerful. For a given significance level α , model m , and number of data points N , we thus deter-

mine the value of the standard deviation $\sigma = \tilde{\sigma}$ such that the β -risk of the χ^2 test has a fixed value of 4α , i.e., $\beta(N; \chi^2) \equiv 1 - \text{pow}(N, \chi^2, \tilde{\sigma}, m, \alpha) = 4\alpha$, where $\text{pow}(\cdot)$ is the statistical power (Online Methods). We then use this value $\tilde{\sigma}$ to evaluate the β -risk for a test T as $\beta(N; T) \equiv 1 - \text{pow}(N, T, \tilde{\sigma}, m, \alpha)$. The β -risk ratio $\beta(N; T) / \beta(N; \chi^2) = \beta(N; T) / (4\alpha)$ is thus the factor by which the β -risk changes if we use test T instead of the χ^2 test for noise levels where the χ^2 test has a β -risk of 4α . β -risk ratios larger/smaller than one correspond to the test T being less/more powerful than the χ^2 test.

We find that the $(\chi^2, h^\pm)$ -statistic has lower β -risk than the χ^2 statistic already for a few tens of data points (Fig. 2d-g). The β -risk ratio decreases with increasing

number of data points and seems to reach a plateau for models 5 and 6. For large enough system sizes, the β -risk ratio reaches values below ~ 0.1 for all considered α values and models. For $\alpha \geq 0.01$, the β -risk ratio reaches values < 0.01 for all models. In comparison, the (χ^2, h) test is less powerful resulting in a higher β -risk ratio (see Fig. S11).

This trend of essentially monotonically decreasing β -risk ratio with increasing data size is also observed for the $h^\pm$ -statistic. Note that the β -risk ratios of the h test and of the (χ^2, h) test show a non-monotonic dependence on N with a minimum at $N \approx 3000$ for $\alpha = 0.05$ (see Fig. S11). Although the $h^\pm$ statistic relies on the signs of the residuals only, it performs significantly better than the χ^2 -statistic over large ranges of data sizes. We observe that the number of data points, where the β -risk ratio becomes one and both tests have equal power, moves to larger sizes for larger α -values. However, uncertainties in the standard errors, which we assume here to be known exactly, will degrade the performance of the χ^2 test but not of the $h^\pm$ test.

The $h^\pm$ statistic can easily be extended to ordered data

subject to correlated noise. Interpreting the signs of the residuals as spins, the uncorrelated and correlated cases correspond to one-dimensional Ising models without and with spin-spin couplings. Constant nearest-neighbor couplings correspond to the simplest case of exponentially decaying correlations. Partition functions and run length distributions h and $h^\pm$ can then readily be calculated using the machinery of statistical mechanics (Online Methods). Proceeding as before in the case of uncorrelated noise, one can then calculate P-values as measures of surprise.

An open source Python 3 implementation called `hplusminus` is available free of charge at <https://github.com/bio-phys/hplusminus>.

ACKNOWLEDGMENTS

We thank Drs. Ruth Pfeiffer, Mitchell H. Gail, and Roberto Covino for helpful discussions. This work was supported by the Max Planck Society.

-
- * Department of Theoretical Biophysics, Max Planck Institute of Biophysics, Max-von-Laue-Straße 3, 60438 Frankfurt am Main, Germany; Author to whom correspondence should be addressed. Electronicmail: juergen.koefinger@biophys.mpg.de
 - † Department of Theoretical Biophysics, Max Planck Institute of Biophysics, Max-von-Laue-Straße 3, 60438 Frankfurt am Main, Germany; Institute for Biophysics, Goethe University, Max-von-Laue-Straße 9, 60438 Frankfurt am Main, Germany
 - ¹ K. Pearson, *Philosophical Magazine* **50**, 157 (1900).
 - ² A. Wald and J. Wolfowitz, *Annals Math. Statistics* **11**, 147 (1940).
 - ³ M. F. Schilling, *The College Mathematics Journal* **21**, 196 (1990).
 - ⁴ D. Franke, C. M. Jeffries, and D. I. Svergun, *Nature Methods* **12**, 419+ (2015).
 - ⁵ S. I. Denisov and P. Hänggi, **71**, 046137, publisher: American Physical Society.
 - ⁶ R. A. Fisher, *Journal of the Royal Statistical Society* **98**, 39 (1935).
 - ⁷ J. O. Irwin, *Metron* **12**, 83–94 (1935).
 - ⁸ C. E. Shannon, *The Bell System Technical Journal* **27**, 379 (1948).
 - ⁹ E. T. Jaynes, in *Statistical Physics* (K. Ford (ed.), Benjamin, New York, 1963), Chap. Information Theory and Statistical Mechanics, p. p. 181.
 - ¹⁰ H. Prautzsch, W. Boehm, and M. Paluszny, *Bézier and B-Spline Techniques, Mathematics and Visualization* (Springer, Berlin and Heidelberg, 2002).

ONLINE METHODS

The h -statistic. The histogram of run lengths for a given configuration is given by the vector of integers $h \equiv (h_1, h_2, \dots, h_N)$, where $0 \leq h_i \leq N$ is the number of runs of length i in a sign configuration. We collect the lengths of runs of positive signs and the lengths of runs of negative signs in the single histogram h . Note that $\sum_{i=1}^N h_i = n$, where n is the number of runs, and that $\sum_{i=1}^N i h_i = N$, where N is the number of data points. This means that h is a partition of the integer N .

We assume that a configuration of N signs is generated by randomly drawing positive and negative signs with equal probability. Then, the conditional probability to observe n runs is given by

$$p(n|N) = \frac{1}{2^{N-1}} \binom{N-1}{n-1} \quad (1)$$

The binomial coefficient $\binom{N-1}{n-1}$ is the multiplicity of dividing N sites in n segments and 2^{N-1} is the normalization constant. Here and in the following we use the convention that the binomial coefficient $\binom{N}{n} = 0$ for $n < 0$ and $n > N$.

We calculate the number of possibilities to arrange n runs. If all runs were of different length then there would be $n!$ possibilities. However, we have h_i runs of length i such that we have to correct $n!$ by dividing out $h_i!$ for all i , i.e., the multiplicity of h is given by the multinomial coefficient $n! / \prod_{i=1}^N h_i!$. The normalization constant is given by the total number of configurations with n segments, i.e., by $\binom{N-1}{n-1}$. Consequently, the probability to observe a run length distribution h for a given number of data points N and number of runs n is given by

$$p(h|n, N) = \frac{1}{\binom{N-1}{n-1}} \frac{n!}{\prod_{i=1}^N h_i!}. \quad (2)$$

Thus, we obtain for the joint probability $p(h) \equiv p(h, n|N) = p(h|n, N)p(n|N)$ of observing a run length distribution h with n runs of finite length

$$p(h) = \frac{1}{2^{N-1}} \frac{n!}{\prod_i h_i!} \quad (3)$$

The $h^\pm$ -statistic. The $h^\pm = (h^+, h^-)$ statistic is given by the pair of run length histograms of runs with positive signs, $h^+ = \{h_1^+, \dots, h_N^+\}$, and runs with negative signs, $h^- = \{h_1^-, \dots, h_N^-\}$. Note that $\sum_{i=1}^N h_i^S = n^S$ and $\sum_{i=1}^N i h_i^S = N^S$, where n^S is the number of runs with sign $S = +, -$ and N^S is the number of positive ($S = +$) and negative ($S = -$) signs.

We consider run length distributions for positive and negative signs separately. Of n runs, n^+ runs have sign $s = +1$ and n^- runs have sign $s = -1$, i.e., $n = n^+ + n^-$. Positive and negative runs alternate such that $|n^+ - n^-| = 0$ if n is even and $|n^+ - n^-| = 1$ if n is odd. The conditional probability to observe n^+ runs of sign $+1$,

and consequently $n^- = n - n^+$ runs of sign -1 , is given by

$$p(n^+|n \text{ even}, N) = \delta\left(n^+, \frac{n}{2}\right) \quad (4)$$

for even n and by

$$p(n^+|n \text{ odd}, N) = \frac{1}{2} \left[\delta\left(n^+, \frac{n+1}{2}\right) + \delta\left(n^+, \frac{n-1}{2}\right) \right] \quad (5)$$

for odd n . Here, $\delta(x, y) = 1$ if $x = y$ and zero otherwise. In summary,

$$p(n^+|n, N) = \begin{cases} p(n^+|n \text{ even}, N) & \text{for } n \text{ even} \\ p(n^+|n \text{ odd}, N) & \text{for } n \text{ odd} \end{cases} \quad (6)$$

The conditional probability to observe $N^+ = N - N^-$ positive signs is determined by the product of the multiplicities of distributing N^+ positive signs on n^+ runs and N^- negative signs on n^- runs, i.e.,

$$p(N^+|n^+, N) = \frac{1}{Z} \binom{N^+ - 1}{n^+ - 1} \binom{N^- - 1}{n^- - 1}$$

for $n^+ > 0$ and $n^- > 0$ and by $p(N^+|n^+, N) = 0$ else. The normalization constant Z is given by

$$Z = \sum_{N^+ = n^+}^{N - n^+} \binom{N^+ - 1}{n^+ - 1} \binom{N^- - 1}{n^- - 1} = \left[\binom{N - n^+ - 1}{n^- - 1} {}_2F_1(n^+, n^+ + n^- - N, 1 + n^+ - N, 1) \right]^{-1} \quad (7)$$

where ${}_2F_1(\cdot, \cdot, \cdot)$ is the ordinary hypergeometric function.

The conditional probability to observe a run length histogram $h^S \equiv (h_1^S, h_2^S, \dots, h_N^S)$, where $S = +, -$, is

$$p(h^S|n^S, N^S, N) = \binom{n^S}{h_1^S \dots h_{N^S}^S} \left[\binom{N^S - 1}{n^S - 1} \right]^{-1} \quad (8)$$

for $n^S > 0$ and by $p(h^S|n^S, N^S, N) = 1$ otherwise.

Consequently, the probability to observe the two run length histograms h^+ and h^- is given by

$$p(h^\pm) = \times p(h^+|n^+, N^+, N) p(h^-|n^-, N^-, N) \times p(N^+|n^+, N) p(N^-|n^-, N) p(n|N) \quad (9)$$

In the Supporting Information, we present a corresponding expression for $p(h^\pm)$ generalized to asymmetric probabilities for the signs.

Combination with χ^2 -statistic. Pearson's χ^2 -test is based on the probability density of χ^2 given by

$$p(\chi^2|k) = \frac{(\chi^2)^{k/2-1} e^{-\chi^2/2}}{2^{k/2} \Gamma(\frac{k}{2})} \quad (10)$$

where k is the number of degrees of freedom. For $k \geq 2$, $p(\chi^2|k)$ has a single peak at $k - 2$. We introduce $p(\chi^2) = p(\chi^2|N)$.

P-value calculation using the Shannon information distribution. We assume that our discrete statistic of state i is called x_i and has a probability p_i . For a given value x_k we now calculate the probability that a state x_j randomly sampled from the discrete distribution $\{x_i, p_i\}$ has a probability $p_j \leq p_k$, or equally that $\mathcal{I}_j \geq \mathcal{I}_i$ where $\mathcal{I}_i = -\ln p_i$ is the Shannon information.⁸ The probability distribution of the Shannon information is then given by counting all states that have the same probability, i.e.,

$$p(\mathcal{I}_k) = p_k \sum_i \delta[\mathcal{I}_k - \mathcal{I}_i] \quad (11)$$

where the sum extends over all states and where $\delta[y] = 1$ if $y = 0$ and zero else. Thus, the P-value P is given by the cumulative distribution function of the Shannon information,

$$P(\mathcal{I}_k) = \text{cdf}(\mathcal{I}_k) = \sum_i \Theta[\mathcal{I}_k - \mathcal{I}_i] p_i \quad (12)$$

where $\Theta[y] = 1$ if $y \geq 0$ and zero otherwise.

For example, for the probability distribution given by eq 9, we obtain

$$\text{cdf}(\mathcal{I}) = \sum_{h^\pm} p(h^\pm) \Theta[-\ln p(h^\pm|N) - \mathcal{I}] \quad (13)$$

where we sum over all run length distribution $h^\pm$ and where $\mathcal{I}$ is the Shannon information of the sample. Using that $p(h^\pm) = \frac{1}{2^N} \sum_{s_1=\pm 1} \dots \sum_{s_N=\pm 1} \delta(h^\pm(\{s_i\}) - h^\pm)$, where we sum over all 2^N sign configurations $\{s_i\}$, the above equation can be rewritten as

$$\text{cdf}(\mathcal{I}) = \frac{1}{2^N} \sum_{s_1=\pm 1} \dots \sum_{s_N=\pm 1} \Theta[-\ln p(h^\pm(\{s_i\})) - \mathcal{I}] \quad (14)$$

where $\Theta[y] = 1$ if $y \geq 0$ and zero else.

For continuous random variables x and constant invariant measure $m(x)$, we obtain

$$p(\mathcal{I}) = \int p(x) \delta[\mathcal{I}(x) - \mathcal{I}] dx \quad (15)$$

where $\delta(\cdot)$ is Dirac's δ -function and

$$P(\mathcal{I}) = \text{cdf}(\mathcal{I}) = \int p(x) \Theta[\mathcal{I}(x) - \mathcal{I}] dx \quad (16)$$

We obtain that

$$\text{cdf}(\mathcal{I}) = \sum_i \frac{e^{-\mathcal{I}(x_i)}}{|\mathcal{I}'(x_i)|} \quad (17)$$

where x_i are the solutions to $\mathcal{I}(x) = -\ln p(x)$ and $\mathcal{I}'(x)$ is the first derivative of $\mathcal{I}(x)$.

In practice, we generate M sign configurations randomly, calculate the probabilities according to eq 9, and rank them from smallest to largest, i.e., $\{\mathcal{I}_i\} =$

$\{\mathcal{I}_1, \mathcal{I}_2, \dots, \mathcal{I}_M\}$ with $\mathcal{I}_i \leq \mathcal{I}_{i+1}$. We then calculate the P-value of for the Shannon information $\mathcal{I}$ of a sign configuration by counting all sampled values of the Shannon information $\mathcal{I}_i \geq \mathcal{I}$ and dividing by the number of samples M , i.e.,

$$\text{pv}(\mathcal{I}) = \frac{k}{M} \quad (18)$$

with k given such that $\mathcal{I}_k \geq \mathcal{I}$ and $\mathcal{I}_{k+1} < \mathcal{I}$. Note that by calculating the Shannon information instead of probabilities we avoid numerical problems arising when evaluating eqs 3 and 9 directly.

For continuous random variable and constant invariant measure $m(x)$, we use the probability density to rank samples according to typicality and to calculate the P-value as in the case for discrete random variable as explained above.

We illustrate in the Supporting Information that the Shannon information distributions of various probability distributions closely follow gamma distributions. The Shannon information distributions of the Gaussian distribution and the exponential distribution are given exactly by shifted gamma distributions. For the truncated Gaussian distribution we obtain a truncated and shifted gamma distribution. The Shannon information distribution of the χ^2 distribution can be accurately approximated by a shifted gamma distribution and this approximation becomes more accurate for increasing number of degrees of freedom (see Fig. S5)

Statistical power. The cumulative P-value distribution function defines the statistical power of a test statistic for a given model m , i.e.,

$$\text{pow}(N, T, \sigma, m, \alpha) = \int_0^\alpha p(\text{pv}|N, T, \sigma, m) d\text{pv} \quad (19)$$

where pv denotes the P-value and where $p(\text{pv}|N, T, \sigma, m)$ is the distribution function of P-values generated by all noise realizations with standard deviation σ . The statistical power quantifies how strongly the P-value distribution function is peaked at $P = \alpha$. If m is the true model then the P-value distribution is flat and $\text{pow}(N, T, \sigma, m, \alpha) = \alpha$.

Empirically, we find that the statistical power as a function of the noise is well described by

$$\text{pow}(\sigma; \alpha, k_i, b_i) = \alpha^{\exp\left[-\frac{k_i}{(\sigma - b_i)^2}\right]} \quad (20)$$

for $\sigma \geq b_i$, where k_i and b_i are positive constants for a given test statistic T_i , model, and α value. For $\sigma \leq b_i$, we set $\text{pow}(\sigma; \alpha, k_i, b_i) = 1$.

From eq 20, we can express σ and obtain for a given value of the statistical power $\overline{\text{pow}}$

$$\sigma(\overline{\text{pow}}) = \sqrt{\ln \frac{k_i}{\ln \frac{\alpha}{\overline{\text{pow}}}}} + b_i \quad (21)$$

We can use this expression to describe the correlation of the statistical power of a test T_j with a test T_i by

inserting eq 21 into eq 20

$$\begin{aligned} \ln \text{pow}(\sigma(\overline{\text{pow}}); \alpha, k_j, b_j) \\ = \exp \left[-\frac{k_j}{(\sigma(\overline{\text{pow}}) - b_j)^2} \right] \ln \alpha \end{aligned} \quad (22)$$

Correlated noise. We now introduce sign correlations by biasing configurations represented by histograms. The histograms h and the pair of histograms $h^\pm = (h^+, h^-)$ are coarse-grained representations of the sign configuration. h and $h^\pm$ are complete descriptions of the configuration space in the sense that each sign configuration can be assigned to exactly one histogram h and exactly one pair of histograms $h^\pm$. A physically appealing way to introduce correlations is to interpret the signs as spins of a one-dimensional Ising model. We define energy potentials $E(h)$ and $E(h^\pm)$ that depend on h and $h^\pm$, respectively. The probability to observe the histogram h then becomes

$$p(h|J) \propto p(h) \exp[-E(h)] \quad (23)$$

and the probability to observe $h^\pm$ is given by

$$p(h^\pm|J) \propto p(h^\pm) \exp[-E(h^\pm)] \quad (24)$$

For nearest neighbor coupling J we have $E = -J \sum_{i=1}^{N-1} s_i s_{i+1} = -J(N-1-2(n-1)) = -J(N-2n+1)$. For this simplest correlated model, the Boltzmann distribution is given by $\exp[-E]/\cosh^{N-1}(J)$ and the sign-sign correlations decay exponentially in the thermodynamic limit, $\langle s_i s_j \rangle \propto \exp[-|i-j|/\xi]$ with $\xi = 1/\ln \tanh(J)$. We can also add a field coupling to the spins, which then represents asymmetric error distributions. With these newly defined probabilities, we can proceed as in the case for uncorrelated noise and calculate P-values as measures for typicality.

Calculation details. We define six models for the difference between the true model and alternative models as shown in Fig. 1e-j. For the true model m_1 , this difference is zero. The five alternative models $m_2, \dots, m_6$ have been generated using step functions and they have been least-square fitted to m_1 . We then resolve these models with equally spaced data points. We used two sets data points. To highlight results for individual value of N , we chose the numbers of data points as $N = i \times 10^j$, where $i = 1, 2, \dots, 9$ and $j = 1, \dots, 4$, and $N = 10^5$. To study the size dependence of the statistical power and the β -risk we use an approximately logarithmic scale for N , i.e., $N = \text{round}(10^{1+0.1i})$ with $i = 0, 1, \dots, 40$ such that $N = 10, 13, 16, \dots, 79433, 100000$. To the models we added uncorrelated Gaussian noise $\mathcal{N}(0, \sigma)$ with

zero mean and standard deviation σ . For the σ -values we chose 199 equidistant values with spacing 0.0025 in the closed interval $[0.005, 0.5]$ and scaled these values by $1/0.11$ for models 1 and 2 and by $1/0.1, 1/0.095, 1/0.08$ and $1/0.09$ for models 3, 4, 5, and 6, respectively.

For the P-value calculation, we have to calculate the distribution of the test statistics for the true model. To do so, we added normally distributed noise to the models, evaluated for all noise realization the statistics and their probabilities, and recorded them in lists. We evaluated the χ^2 -, h -, and $h^\pm$ -statistic for 10^6 samples. For the (χ^2, h) - and $(\chi^2, h^\pm)$ -statistic, we take advantage of the independence of the χ^2 statistic and the h and $h^\pm$ -statistic. We generated 5000 χ^2 values and evaluated their log-probabilities. We also generated 20000 sign configurations and evaluated h and $h^\pm$, and the respective log-probabilities for those. We then summed up the Shannon information values corresponding to the products $p(\chi^2)p(h)$ and $p(\chi^2)p(h^\pm)$, sorted the list, and sampled only every 100th value to reduce the size. We then formed cumulative histograms of the Shannon information for each statistic (see Fig. S2 for an example with $N = 500$).

We perform least-square fits of the cumulative gamma distribution function to each of the Shannon information distribution functions and extracted the shape parameter α , the inverse scale parameter β , and the shift or location parameter $\mathcal{I}_0$. We fitted B-splines¹⁰ to these parameters as functions of the logarithm of the number of data points N (see Fig. S5).

We used the Z-score approximation of the runs test of Wald and Wolfowitz and calculate two-tailed P-values for $Z = (r - \mu_r)/\sigma_r$ where r is the number of runs and where $\mu_r = 1 + (2N^+N^-)/N$ and $\sigma_r^2 = (2N^+N^-(2N^+N^- - N))/(N^2(N-1))$.

We calculated the cumulative probabilities $\text{cdf}(l_{\max})$ for the longest run $l_{\max}$ following Ref. 3. We then calculate a two-tailed P-value by $\text{P-value}(l_{\max}) = 2\text{cdf}(l_{\max})$ if $\text{cdf}(l_{\max}) < 0.5$ and $\text{P-value}(l_{\max}) = 2[1 - \text{cdf}(l_{\max})]$ else.

To calculate the statistical power for the considered models, we generated for each model and σ -value 5000 samples. We evaluated the P-values for these samples and collected the negative logarithm of the P-values in histograms using 160 bins in the range of 0 to 40. We obtained the statistical power by evaluating the interpolating function of the cumulative P-value histogram for the given significance level α .

For all considered statistics, system sizes, and noise levels, and for given α -value, we fitted eq 20 to the statistical power as a function of σ using the fit parameters k_i and b_i . We use these fitted values of the parameters to reproduce the correlations of the statistical powers of two test statistics using eq 22 as shown in Fig. 2a-c.

Supporting information: Powerful statistical tests for ordered data

Jürgen Köfinger^{1,*} and Gerhard Hummer^{1,2}

¹*Department of Theoretical Biophysics, Max Planck Institute of Biophysics,
Max-von-Laue-Straße 3, 60438 Frankfurt am Main, Germany*

²*Institute for Biophysics, Goethe University, Max-von-Laue-Straße 9, 60438 Frankfurt am Main, Germany*

I. GENERALIZATION TO ASYMMETRIC SIGN PROBABILITIES

In the derivation of eq 9 in the Online Methods, we assumed that positive and negative signs are equally likely. Here, we assume signs have probabilities p_+ to be positive and $p_- = 1 - p_+$ to be negative. For these asymmetric probabilities, we derive the probability of observing distributions of run length of positive and negative signs corresponding to eq 9. We write this probability as

$$\begin{aligned} p(h^\pm|p_+) = & \quad (1) \\ & \times p(h^+|n^+, N^+, N) p(h^-|n^-, N^-, N) \\ & \times p(n^+, n^-|N^+, N) p(N^+|N) \end{aligned}$$

where $p(h^+|n^+, N^+, N)$ and $p(h^-|n^-, N^-, N)$ are given by eq 8. In the following, we derive expressions for $p(n^+, n^-|N)$ and $p(N^+|N)$.

The probability that we find N^+ positive signs is given by the binomial distribution as

$$p(N^+|N) = \binom{N}{N^+} p_+^{N^+} p_-^{N-N^+}. \quad (2)$$

Next we calculate the joint probability $p(n^+, n^-|N^+, N)$ to find $n_\pm$ runs with signs ± 1 . To take into account that in an alternating sequence of positive and negative runs, n^+ and n^- cannot deviate from each other by more than one, we introduce $\Delta = n^+ - n^- = -1, 0, 1$. For the two ordered states with all signs positive ($N^+ = N$) or all signs negative ($N^+ = 0$), the probabilities are equal one,

$$p(n^+ = 1, \Delta = 1|N^+ = N, N) = 1 \quad (3)$$

and

$$p(n^+ = 0, \Delta = -1|N^+ = 0, N) = 1 \quad (4)$$

and zero otherwise. For $N^+ \neq 0, N$, the conditional probability is given by

$$p(n^+, \Delta|N^+, N) = \frac{1 + \delta(\Delta, 0)}{Z} \binom{N^+ - 1}{n^+ - 1} \binom{N^- - 1}{n^+ - \Delta - 1} \quad (5)$$

For $\Delta = 0$, $1 + \delta(\Delta, 0) = 2$ which accounts for the two possible arrangements of runs for $n^+ = n^-$, starting the sequence of runs either with a positive or a negative run. The normalization constant Z is given by the double sum

$$Z = \sum_{n^+=n_{\min}^+}^{n_{\max}^+} \sum_{\Delta=\Delta_{\min}(n^+)}^{\Delta_{\max}(n^+)} [1 + \delta(\Delta, 0)] \binom{N^+ - 1}{n^+ - 1} \binom{N^- - 1}{n^+ - \Delta - 1} \quad (6)$$

where the lower limit for the number of positive runs n^+ is given by

$$n_{\min}^+ = \begin{cases} 0 & \text{for } N^+ = 0, \text{ and} \\ 1 & \text{otherwise.} \end{cases} \quad (7)$$

and the upper limit by

$$n_{\max}^+ = \min(N^+, N^- + 1). \quad (8)$$

The lower limit for Δ , and therefore the upper limit for the number of negative runs $n^- = n^+ + \Delta$, is given by

$$\Delta_{\min} = \begin{cases} 1 & \text{for } n^+ = N^- + 1, \\ 0 & \text{for } n^+ = N^-, \text{ and} \\ -1 & \text{otherwise,} \end{cases} \quad (9)$$

and the upper limit by

$$\Delta_{\max} = \begin{cases} -1 & \text{for } n^+ = 0, \\ 0 & \text{for } n^+ = 1, \text{ and} \\ 1 & \text{otherwise.} \end{cases} \quad (10)$$

We show numerically that for $p^+ = p^- = 1/2$, eq 1 gives the same results as eq 9. For $N = 10$, we generate 10^6 sign configurations, for which we evaluate $p(h^\pm)$ given by eq 9 of the Online Methods and $p(h^\pm|p_+ = 1/2)$ given by eq 1 presented here. Additionally, we determine the probabilities of the pairs of distributions (h^+, h^-) by counting. To do so, we convert (h^+, h^-) into unique strings. We find that both expressions agree well with each other and with the numerically determined probabilities (see Fig. S12).

II. THE PROBABILITY DISTRIBUTION OF THE SHANNON INFORMATION

We can calculate the Shannon information distribution $p(\mathcal{I})$ of a probability distribution $p(x)$ of a continuous scalar variable x directly via

$$p(\mathcal{I}) = \sum_i \frac{p^2(x_i)}{|p'(x_i)|} = \sum_i \frac{e^{-\mathcal{I}(x_i)}}{|\mathcal{I}'(x_i)|} \quad (11)$$

where the x_i are solutions to

$$\mathcal{I} = -\ln p(x) \quad (12)$$

We illustrate next that the Shannon information distribution follows a gamma distribution in many important

FIG. S1. **The runs test r of Wald and Wolfowitz¹ and the longest-run test $l_{\max}$ by Schilling² perform worse than the χ^2 test for all data sizes.** For the models shown in Fig. 1f-j in the main text (colors), we show the β -risk relative to the β -risk of the χ^2 statistic as a function of the number of data points. We fixed the β -risk of the χ^2 -test as $\beta(\chi^2) = 4\alpha$ (black horizontal lines) for the significance levels $\alpha = 0.05, 0.01, 0.005, 0.001$ (a-d) and evaluated the β -risk of the alternative test statistics for the corresponding noise levels. In each panel a-d, the top plot shows results for the runs test of Wald and Wolfowitz and the bottom plot shows results for Schilling's longest-run test. Note that we evaluated the longest run test of Schilling for data sizes up to $N = 1000$ because the recursive calculation of the probability² fails for larger sizes.

cases. In the following four examples, $p(\mathcal{I})$ is either exactly given by, proportional to, or approximated by a shifted γ -distribution,

$$\gamma(\mathcal{I} - \mathcal{I}_o, \alpha, \beta) = \frac{\beta^{-\alpha} (\mathcal{I} - \mathcal{I}_o)^{\alpha-1} e^{-\beta(\mathcal{I} - \mathcal{I}_o)}}{\Gamma(\alpha)} \quad (13)$$

The shift parameter $\mathcal{I}_o$ is given by the largest value of the probability density.

Exponential distribution For the exponential distribution

$$p(x) = k e^{-kx} \quad (14)$$

we obtain

$$p(\mathcal{I}) = \frac{e^{-\mathcal{I}}}{k} = e^{-(\mathcal{I} + \ln k)} = \gamma(\mathcal{I} - \mathcal{I}_o; 1, 1) \quad (15)$$

where $\mathcal{I}_o = -\ln k$ and $\mathcal{I}_o \leq \mathcal{I} \leq -\infty$.

Gaussian distribution For the Gaussian distribution

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (16)$$

we obtain

$$p(\mathcal{I}) = \gamma\left(\mathcal{I} - \mathcal{I}_o; \frac{1}{2}, 1\right) \quad (17)$$

where $\mathcal{I}_o = \ln \sqrt{2\pi\sigma^2}$

Truncated Gaussian distribution For the truncated Gaussian distribution given by

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2} \operatorname{erf}\left(\frac{w}{\sqrt{2\sigma^2}}\right)} e^{-\frac{x^2}{2\sigma^2}} \quad (18)$$

for $-w < x < w$ and zero otherwise, we obtain that the probability distribution of the Shannon information is proportional to a shifted gamma-distribution, i.e.,

$$p(\mathcal{I}) = \operatorname{erf}^{-1}\left(\frac{w}{\sqrt{2\sigma^2}}\right) \gamma\left(\mathcal{I} - \mathcal{I}_o; \frac{1}{2}, 1\right) \quad (19)$$

where $\mathcal{I}_o = \ln \sqrt{2\pi\sigma^2} / \operatorname{erf}\left(\frac{w}{\sqrt{2\sigma^2}}\right)$ and $\mathcal{I}_o \leq \mathcal{I} \leq -\ln p(w)$.

FIG. S2. **The P-value is given by the cumulative distribution function of the Shannon information $\text{cdf}(\mathcal{I})$.** The Shannon information is given by $\mathcal{I} = -\ln p$, where p is the probability (density) of the test statistics. Top to bottom, we show numerical results for the χ^2 , h , $h^\pm$, and $(\chi^2, h^\pm)$ statistics for $N = 500$ data points (red solid lines). The cumulative distributions functions closely follow cumulative shifted gamma distributions, which have been least-square fitted to the numerical data (black dashed line). For the values of the fit parameters see Fig. S5.

a. Multivariate normal distribution The multivariate normal distribution in n dimensions is given by

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2} \sum_{i=1}^n x_i^2} \quad (20)$$

Its maximum defines $\mathcal{I}_o = \frac{n}{2} \ln(2\pi)$. Using spherical coordinates in n dimensions and that the surface of an n -dimensional sphere is given by $S = 2\pi^{\frac{n}{2}} R^{n-1} / \Gamma(\frac{n}{2})$, with $R^2 = \sum_{i=1}^n x_i^2$, we obtain

$$p(\mathcal{I}) = \gamma \left(\mathcal{I} - \mathcal{I}_o; \frac{n}{2}, 1, \right) \quad (21)$$

χ^2 -distribution As we show next, the Shannon information distribution function of the χ^2 -distribution $p(\chi^2|N)$, defined in eq 10 of the Online Methods, closely

follows a gamma distribution even for just a few degrees of freedom. That is,

$$p(\mathcal{I}) \approx \gamma \left(\mathcal{I} - \mathcal{I}_o; \frac{1}{2}, 1 \right) \quad (22)$$

with $\mathcal{I}_o = -\ln p(N-2; N)$ and $\mathcal{I}_o < \mathcal{I} < \infty$. The values of the shape parameter α and the inverse scale parameter β are identical to the values for Gaussian distribution. The reason is that the χ^2 distribution approaches the Gaussian distribution for increasing number of degrees of freedom. Consequently, this approximation becomes more accurate for larger N .

In the following we derive the gamma distribution approximation of the Shannon information distribution for the χ^2 -distribution to highlight the underlying approximations. The derivative of the Shannon information is given by

$$\mathcal{I}'(x) = \frac{2-k}{2} \frac{1}{x} + \frac{1}{2} = \frac{1}{2} \left[1 - \frac{k-2}{x} \right] \quad (23)$$

The two solutions of $\mathcal{I} = -\ln p(x)$ are given by

$$x_n = (2-k) W_n[A(\mathcal{I})] \quad (24)$$

where we introduced

$$A(\mathcal{I}) = \frac{\left[e^{-\mathcal{I}} 2^{\frac{k}{2}} \Gamma\left(\frac{k}{2}\right) \right]^{\frac{2}{k-2}}}{2-k} \quad (25)$$

and where $n = 0, -1$ indicate the two branches $W_n(\cdot)$ of the Lambert W function. This function, also called the product logarithm, is defined as the inverse function of $f(x) = xe^x$. We thus obtain for $x_{-1} > k-2$

$$|\mathcal{I}'(x_{-1})| = \frac{1}{2} \left[1 - \frac{k-2}{x_{-1}} \right] \quad (26)$$

and for $x_0 < k-2$

$$|\mathcal{I}'(x_0)| = \frac{1}{2} \left[\frac{k-2}{x_0} - 1 \right] \quad (27)$$

such that the Shannon information distribution can be written as

$$p(\mathcal{I}) = 2e^{-\mathcal{I}} \left[\frac{1}{\frac{k-2}{x_0} - 1} - \frac{1}{\frac{k-2}{x_{-1}} - 1} \right] \quad (28)$$

Inserting, eq 24 into eq 28, we obtain

$$p(\mathcal{I}) = 2e^{-\mathcal{I}} \left[\frac{W_{-1}[A(\mathcal{I})]}{W_{-1}[A(\mathcal{I})] + 1} - \frac{W_0[A(\mathcal{I})]}{W_0[A(\mathcal{I})] + 1} \right] \quad (29)$$

Using that the derivative of the Lambert W function is given by

$$W'_n(x) = \frac{W_n(x)}{W_n(x) + 1} \quad (30)$$

FIG. S3. **The Shannon information distribution functions closely follow gamma distributions.** We plot on a semi-log scale numerical results for the Shannon information distributions functions $p(I)$ for the χ^2 , h , $h^\pm$, and $(\chi^2, h^\pm)$ for $N = 50$ (blue) $N = 100$ (orange), $N = 500$ (red), $N = 1000$ (green), $N = 10000$ (magenta), $N = 100000$ (brown) data points. Least-square fits of the gamma distribution (black dashed lines) agree excellently with the numerical results. For the values of the fit parameters see Fig. S5.

FIG. S4. **The Shannon distribution functions of a the χ^2 -statistic can be summarized by only three parameters for given N .** The shifted gamma distributions $\gamma(I - I_o, \alpha, \beta)$ have been least-square fitted to the numerically determined Shannon information distributions $p(I)$ using I_o , α , and β as fit parameters. We show as black lines the parameters for the approximate expression given by eq 22 with $\alpha = 1/2$, $\beta = 1$, and $I_o = -\ln p(N - 2|N)$, where $p(N - 2|N)$ is the χ^2 distribution given by eq 10 of Online Methods evaluated at its maximum.

we obtain

$y = -1/e$, we obtain

$$p(I) = 2e^{-I}A(I) [W'_{-1}[A(I)] - W'_0[A(I)]] \quad (31)$$

$$W'_{-1}(y) - W'_0(y) = -\frac{\sqrt{2e}}{\sqrt{\frac{1}{e} + y}} - \frac{11e^{3/2}\sqrt{\frac{1}{e} + y}}{6\sqrt{2}} - \frac{769e^{5/2}\left(\frac{1}{e} + y\right)^{3/2}}{432\sqrt{2}} + \dots \quad (32)$$

Differentiating the Taylor series of $W_{-1}(y) - W_0(y)$ at

FIG. S5. **The Shannon distribution functions of a statistic is summarized by only three parameters for given N .** The shifted gamma distributions $\gamma(\mathcal{I} - \mathcal{I}_o, \alpha, \beta)$ have been least-square fitted to the numerically determined Shannon information distributions $p(\mathcal{I})$ using $\mathcal{I}_o$, α , and β as fit parameters. Results for the h , h^+ , h^- , (χ^2, h^+) , and (χ^2, h^-) statistics (left to right) show a smooth dependence on the data size for all statistics. The black lines show smoothing B-spline fits to the parameters as a function of $\log_{10}(N)$ in the range from $N = 30$ to $N = 100000$.

FIG. S6. **Validation of the P-value calculation using Shannon information distributions represented by gamma distributions.** We use B-spline representations for the parameters of the gamma distribution α , β , $\mathcal{I}_o$ as functions of the numbers of data points N (vertical axis; see Figure S5) and compare it to our numerical results generated by sampling random sign configurations (horizontal axis). Results for $N = 50, 500, 5000$, and 50000 confirm that the B-spline representation is accurate and that we can use it to accurately calculate P-values from $N = 50$ (vertical black line) to $N = 100000$.

We multiply this expression by y and use only the first term, which is given by

$$-\frac{\sqrt{2ey}}{\sqrt{\frac{1}{e} + y}} = \frac{\sqrt{\frac{2}{e}}}{\sqrt{\frac{1}{e} + y}} - \sqrt{2e}\sqrt{\frac{1}{e} + y} \quad (33)$$

For $y \approx -1/e$, we can neglect the second term and we obtain

$$y [W'_{-1}(y) - W'_0(y)] \approx \sqrt{\frac{2}{e}} \left(y + \frac{1}{e} \right)^{-\frac{1}{2}} \quad (34)$$

Next, we evaluate this expression for $y = A(\mathcal{I})$. The

Taylor series expansion of $[e^{-\mathcal{I}}]^{\frac{2}{k-2}}$ at $\mathcal{I}_o$ is given by

$$[e^{-\mathcal{I}}]^{\frac{2}{k-2}} = (e^{-\mathcal{I}_o})^{\frac{2}{-2+k}} - \frac{2(e^{-\mathcal{I}_o})^{\frac{2}{-2+k}}(\mathcal{I} - \mathcal{I}_o)}{-2+k} + \dots \quad (35)$$

We use the first two terms of this series and express $A(\mathcal{I})$

FIG. S7. **The h test has larger statistical power than the runs test of Wald and Wolfowitz.** Power correlation of the h statistic (vertical axis) with the runs test of Wald and Wolfowitz (horizontal axis, r is the number of runs) for the models shown in Fig. 1 in the main text and $N = 50, 500, 5000$, and 50000 data points. The lines represent fits based on an empirical equation describing the dependence of the statistical power on the noise (Online Methods).

FIG. S8. **The $h^\pm$ test has larger statistical power than the runs test of Wald and Wolfowitz.** Power correlation of the $h^\pm$ statistic (vertical axis) with the runs test of Wald and Wolfowitz (horizontal axis, r is the number of runs) for the models shown in Fig. 1 in the main text and $N = 50, 500, 5000$, and 50000 data points. The lines represent fits based on an empirical equation describing the dependence of the statistical power on the noise (Online Methods).

as

$$\begin{aligned}
 A(\mathcal{I}) &= \frac{\left[e^{-\mathcal{I}} 2^{\frac{k}{2}} \Gamma\left(\frac{k}{2}\right) \right]^{\frac{2}{k-2}}}{2-k} \\
 &\approx \left[(e^{-\mathcal{I}_o})^{\frac{2}{-2+k}} - \frac{2(e^{-\mathcal{I}_o})^{\frac{2}{-2+k}} (\mathcal{I} - \mathcal{I}_o)}{-2+k} \right] \frac{\left[2^{\frac{k}{2}} \Gamma\left(\frac{k}{2}\right) \right]^{\frac{2}{k-2}}}{2-k} \\
 &= \left[1 - \frac{2(\mathcal{I} - \mathcal{I}_o)}{-2+k} \right] \frac{\left[2^{\frac{k}{2}} \Gamma\left(\frac{k}{2}\right) e^{-\mathcal{I}_o} \right]^{\frac{2}{k-2}}}{2-k} \quad (36)
 \end{aligned}$$

For the second factor, we obtain

$$\frac{\left[2^{\frac{k}{2}} \Gamma\left(\frac{k}{2}\right) e^{-\mathcal{I}_o} \right]^{\frac{2}{k-2}}}{2-k} = -\frac{1}{e} \quad (37)$$

such that

$$A(\mathcal{I}) \approx -\frac{1}{e} \left[1 - \frac{2(\mathcal{I} - \mathcal{I}_o)}{-2+k} \right] \quad (38)$$

FIG. S9. **The h test has larger statistical power than Schilling's longest run test in nearly all cases.** Power correlation of the h -statistic (vertical axis) with the longest run statistic, $l_{\max}$, (horizontal axis) for the models shown in Fig. 1 in the main text and $N = 50, 100, 500$ and 1000 data points. The lines represent fits based on an empirical equation describing the dependence of the statistical power on the noise (Online Methods). Only for models m_5 (pink) and m_6 (yellow) and the smallest data size $N = 50$ considered here, the longest runs test slightly outperforms the h test.

FIG. S10. **The $h^{\pm}$ test has larger statistical power than Schilling's longest run test in nearly all cases.** Power correlation of the $h^{\pm}$ -statistic (vertical axis) with the longest run statistic, $l_{\max}$, (horizontal axis) for the models shown in Fig. 1 in the main text and $N = 50, 100, 500$, and 1000 data points. The lines are fits based on an empirical equation describing the dependence of the statistical power on the noise (Online Methods). Only for models m_5 (pink) and m_6 (yellow) and the smallest data size $N = 50$ considered here, the longest runs test outperforms the h test.

Thus, using this expression for $y = A(\mathcal{I})$, we obtain

$$y [W'_{-1}(y) - W'_0(y)] \approx \sqrt{\frac{1}{e}} \left(\frac{\mathcal{I} - \mathcal{I}_o}{-2 + k} \right)^{-\frac{1}{2}} \quad (39)$$

In this approximation, $p(\mathcal{I})$ is proportional to $(\mathcal{I} -$

$\mathcal{I}_o)^{-\frac{1}{2}} e^{-\mathcal{I}}$. The normalization constant is given by

$$\int_{\mathcal{I}_o}^{\infty} (\mathcal{I} - \mathcal{I}_o)^{-\frac{1}{2}} e^{-\mathcal{I}} d\mathcal{I} = e^{-\mathcal{I}_o} \sqrt{\pi} = e^{-\mathcal{I}_o} \Gamma\left(\frac{1}{2}\right) \quad (40)$$

such that we obtain

$$p(\mathcal{I}) \approx \frac{(\mathcal{I} - \mathcal{I}_o)^{-\frac{1}{2}} \exp(-(\mathcal{I} - \mathcal{I}_o))}{\Gamma\left(\frac{1}{2}\right)} \quad (41)$$

FIG. S11. β -risk for the h -statistic and the (χ^2, h) statistic. For the models shown in Fig. 1f-j in the main text, we show the β -risk relative to the β -risk of the χ^2 statistic as a function of the number of data points. We fixed the β -risk of the χ^2 -test as $\beta(\chi^2) = 4\alpha$ (black horizontal lines) for the significance level (a-d) $\alpha = 0.05, 0.01, 0.005, 0.001$ and evaluated the β -risk of the h and (χ^2, h) test statistics for the corresponding noise levels. Due to numerical limitations some β -risk values are zero and thus not shown in these log-log plots.

equal to eq 22.

* Author to whom correspondence should be addressed. Electronic mail: juergen.koefinger@biophys.mpg.de

¹ A. Wald and J. Wolfowitz, Annals Math. Statistics **11**, 147 (1940).

² M. F. Schilling, The College Mathematics Journal **21**, 196 (1990).

FIG. S12. **Validation of eq 9 (bottom axis) and the alternative expression given by eq 1 (red) using numerical data for $N = 10$ data points.** We calculated the run length histograms h^+ and h^- for 10^6 randomly generated sign configurations. Numerically determined probability values for (h^+, h^-) (green) agree well with the analytical expressions. Note that different run length histograms can have the same multiplicities and thus the same probabilities. For the example considered here we count the 179 unique histograms $h^\pm$ expected from the theory of integer partitions and the 11 unique nonzero probability values given by sums of powers of $1/2$. The smallest probability value is given by $1/2^{10} \approx 0.98 \times 10^{-3}$.