

## Article

# A General Framework for Neutrality Tests Based on the Site Frequency Spectrum

Sebastián E. Ramos-Onsins <sup>1,\*</sup>, Giacomo Marmorini <sup>2,3,†</sup>, Guillaume Achaz <sup>4</sup> and Luca Ferretti <sup>5,\*</sup><sup>1</sup> Centre for Research in Agricultural Genomics (CRAG) CSIC-IRTA-UAB-UB, 08193 Bellaterra, Spain<sup>2</sup> Department of Physics and Mathematics, Aoyama Gakuin University, Sagamihara 252-5258, Japan; giacomo@yukawa.kyoto-u.ac.jp<sup>3</sup> Department of Physics, Nihon University, Tokyo 156-8550, Japan<sup>4</sup> Institut de Systématique, Evolution, Biodiversité, UMR 7205, MNHN and Centre Interdisciplinaire de Recherche en Biologie, UMR 7241, Collège de France, 75231 Paris, France; guillaume.achaz@mnhn.fr<sup>5</sup> Pandemic Sciences Institute and Big Data Institute, Nuffield Department of Medicine, University of Oxford, Oxford OX1 3AZ, UK

\* Correspondence: sebastian.ramos@cragenomica.es (S.E.R.-O.); luca.ferretti@bdi.ox.ac.uk (L.F.)

† These authors contributed equally to this work.

**Abstract:** One of the main necessities for population geneticists is the availability of sensitive statistical tools that enable to accept or reject the standard Wright–Fisher model of neutral evolution. A number of statistical tests have been developed to detect specific deviations from the null frequency spectrum in different directions (e.g., Tajima’s D, Fu and Li’s F and D tests, Fay and Wu’s H). A general framework exists to generate all neutrality tests that are linear functions of the frequency spectrum. In this framework, it is possible to develop a family of optimal tests with almost maximum power against a specific alternative evolutionary scenario. In this paper we provide a thorough discussion of the structure and properties of linear and nonlinear neutrality tests. First, we present the general framework for linear tests and emphasise the importance of the property of scalability with the sample size (that is, the interpretation of the tests should not depend on the sample size), which, if missing, can lead to errors in interpreting the data. After summarising the motivation and structure of linear optimal tests, we present a more general framework for the optimisation of linear tests, leading to a new family of tunable neutrality tests. In a further generalisation, we extend the framework to nonlinear neutrality tests and we derive nonlinear optimal tests for polynomials of any degree in the frequency spectrum.

**Keywords:** coalescent theory; site frequency spectrum; neutrality test; statistical power; summary statistics



**Citation:** Ramos-Onsins, S.E.; Marmorini, G.; Achaz, G.; Ferretti, L. A General Framework for Neutrality Tests Based on the Site Frequency Spectrum. *Genes* **2023**, *14*, 1714. <https://doi.org/10.3390/genes14091714>

Academic Editor: Nico M. Van Straalen

Received: 21 July 2023

Revised: 25 August 2023

Accepted: 26 August 2023

Published: 28 August 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Since the development of molecular genetics techniques allowed to obtain nucleotide sequences for the study of populations genetics [1], a number of neutrality tests have been developed with the objective to facilitate the interpretation of an increasing volume of molecular data. Statistical tests for neutrality have been generated exploiting the different properties of the stationary neutral model. Examples of tests are the HKA [2], which takes advantage of the polymorphism/divergence relationship across independent loci in a multi-locus framework, and the Lewontin–Krakauer test [3], which looks for an unexpected level of population differentiation in a locus in relation to other loci. Additionally, another family of popular tests are the ones related to linkage disequilibrium, such as the one developed by [4], which detects long haplotypes at unusual elevated frequencies in candidate regions.

An important family of these tests, often used as summary statistics, is built on the frequency spectrum of biallelic polymorphisms. This family includes the well known tests by Tajima [5], Fu and Li [6], and Fay and Wu [7]. If an outgroup is available, these tests are based on the unfolded spectrum  $\xi_i$ , that is, the number of segregating sites with a derived

allele frequency of  $i$  in a sample of (haploid) size  $n$ . Without an outgroup, it is not possible to distinguish derived and ancestral alleles and the only available data correspond to the folded spectrum  $\eta_i$ , that is, the number of segregating sites with a minor allele frequency of  $i$ . The quantities  $\xi_i$  and  $\eta_i$  are all positive and the range of allele frequencies is  $1 \leq i \leq n - 1$  for the unfolded spectrum,  $1 \leq i \leq \lfloor n/2 \rfloor$  for the folded spectrum. The average spectra for the standard Wright–Fisher neutral model are given by:

$$E(\xi_i) = \frac{1}{i} \theta L, \quad E(\eta_i) = \frac{n}{i(n-i)(1 + \delta_{i,n-i})} \theta L, \quad (1)$$

where  $L$  is the length of the sequence and  $\theta = 2p\mu N_e$ , where  $\mu$  is the mutation rate per base,  $p$  is the ploidy and  $N_e$  is the effective population size. Note that we define  $\theta$  as the rescaled mutation rate per base and not per sequence. Apart from this, we follow the notation of [8,9]. Note that the spectra are proportional to  $\theta$ .

A general framework for linear tests was presented in [8]. The general tests proposed there were of the form:

$$T_\Omega = \frac{\sum_{i=1}^{n-1} i \Omega_i \xi_i}{\sqrt{\text{Var}\left(\sum_{j=1}^{n-1} j \Omega_j \xi_j\right)}}, \quad T_\Omega^* = \frac{\sum_{i=1}^{\lfloor n/2 \rfloor} i \Omega_i^* \eta_i}{\sqrt{\text{Var}\left(\sum_{j=1}^{\lfloor n/2 \rfloor} j \Omega_j^* \eta_j\right)}} \quad (2)$$

that are centred (i.e., they have a null expectation value) if the weights  $\Omega_i, \Omega_i^*$  satisfy the conditions  $\sum_{i=1}^{n-1} \Omega_i = 0$  and  $\sum_{i=1}^{\lfloor n/2 \rfloor} \Omega_i^* = 0$ . This framework allows the construction of many new neutrality tests and has been used to obtain optimal tests for specific alternative scenarios [10]. However, the original framework does not take into account the dependence of the tests on the sample size, as emphasised in [10]. It is important to choose this dependence in order to obtain results that are as independent as possible on sample size. Moreover, the framework presented in [8] covers just a large subfamily of neutrality tests based on the frequency spectrum, that is, the class of tests that are linear functions of the spectrum. This subfamily contains almost all the tests that can be found in the literature with the exception of the  $G_{\xi}, G_{\eta}$  tests of Fu [11], which are second order polynomials in the spectrum whose form is related with Hotelling statistics. Since these  $G_{\xi}, G_{\eta}$  tests were shown to be quite effective in some scenarios, it would be interesting to build a general framework for these quadratic (and more generally nonlinear) tests. New optimal tests based on extensions of the site frequency spectrum [12,13] were recently applied successfully to the detection of selection pressure on human chromosomal inversions [14].

In this work we provide a detailed study of the properties of the whole family of tests based on the site frequency spectrum, i.e., focus on the structure and the properties of neutrality tests that consider the frequency of the variants per position. Technical details and proofs can be found in Appendices A and B.

We begin with the discussion of the most interesting case, i.e., linear tests. Achaz [8] developed a general framework for constructing linear tests comparing two different estimators of variability, which are based on linear combinations of the frequency, each one containing different weights. We summarise his approach in the Section 2.1.

In Section 2.2, we discuss the importance of scaling the weights with the aim to avoid a dependence with the sample size. This allows us to interpret the information of the test in the same way for any sample size analysed. We provide a thorough analysis of a simple proposal for the scaling of the tests with the sample size. Different scaling strategies (including alternatives to scaling) are analysed and evaluated, demonstrating the importance but also the suitability of different weighting methods depending on the nature of the statistic.

In Section 2.3, we present and expand the theory of optimal tests, i.e., tests that have maximum power to detect an alternative scenario versus a null scenario, introduced in [10]. We show that generic linear tests cannot detect arbitrary deviations from the neutral spectrum, and why tests must be optimised with respect to a specific alternative scenario. A

geometrical interpretation of neutrality test is developed, showing how these tests depend on the differences between the null and the alternative scenario, clarifying the theoretical basis for these tests.

In Section 2.4, we extend this approach to nonlinear tests that consider different moments of the frequency spectrum combined in generic polynomials or in power series. In contrast to linear tests, nonlinear tests are dependent on the level of variability  $\theta$  and thus need an accurate estimation of  $\theta$  to have good statistical properties (e.g., unbiased, high statistical power). These tests are classified in strongly and weakly centred (that is, having null expectation value). Strongly centred tests are those tests that are centred for any value for any estimate of  $\theta$ , even if this estimate is far from the actual value, and thus are more robust to deviations of  $\theta$  estimates than weakly centred tests (simulations shown already in the Results section (Section 3)). Instead, weakly centred tests are only centred if the  $\theta$  estimate is equal to the actual value, but this feature also makes the tests more powerful if the inference of  $\theta$  is reliable.

In Section 3.1 in the Results, we see some consequences of the framework outlined in Methods; it is possible to optimise neutrality tests following a general maximisation approach that depends on the proxy used for the power to reject the null model in the alternative scenario, which depends on the mean spectra, the covariance matrices, and the critical  $p$ -value used. Moreover, we show that under some conditions, the power maximisation can be always achieved by tuning a parameter in a new family of linear “tunable” optimal tests developed here, which depend only on the mean spectra of the alternative model.

In Section 3.2, we present the results for the power to detect deviations from the neutral spectrum in coalescent simulations. We show how linear optimal tests have more power than usual neutrality tests such as Tajima’s  $D$ , nonlinear optimal tests are more powerful than linear ones, and weakly centred tests are more powerful than strongly centred ones if  $\theta$  is known.

In summary, our research augments the understanding of neutrality tests, encompassing weight normalisation, optimal test formulation, and both linear and nonlinear paradigms. The outcome not only enriches theoretical foundations but also provides novel methodologies for increasing power and effectiveness of neutrality tests across diverse scenarios.

## 2. Material and Methods

### 2.1. Linear Neutrality Tests

#### General Framework

As discussed by Achaz [8], the general form for linear tests based on the unfolded spectrum can be written as:

$$T_{\Omega} = \frac{\sum_{i=1}^{n-1} i\Omega_i\tilde{\zeta}_i}{\sqrt{\text{Var}\left(\sum_{j=1}^{n-1} j\Omega_j\tilde{\zeta}_j\right)}} \quad (3)$$

where  $\Omega_i$  is a set of weights satisfying the condition:

$$\sum_{i=1}^{n-1} \Omega_i = 0. \quad (4)$$

This is the most general form if we require that the test is centred and with variance 1, that is,  $E(T_{\Omega}) = 0$  and  $\text{Var}(T_{\Omega}) = 1$ . The condition of intendedness can be obtained substituting the spectrum with its average in the standard neutral model, given by the Equation (1).

Alternatively, it is sufficient to choose any pair of unbiased estimators of  $\theta$  based on the unfolded spectrum:

$$\hat{\theta}_\omega = \frac{1}{L} \sum_{i=1}^{n-1} i\omega_i \zeta_i \quad , \quad \hat{\theta}_{\omega'} = \frac{1}{L} \sum_{i=1}^{n-1} i\omega'_i \zeta_i \quad (5)$$

with weights  $\omega_i, \omega'_i$  that obey the conditions:

$$\sum_{i=1}^{n-1} \omega_i = 1 \quad , \quad \sum_{i=1}^{n-1} \omega'_i = 1 \quad (6)$$

to obtain a new test for neutrality:

$$T_\Omega = \frac{\hat{\theta}_\omega - \hat{\theta}_{\omega'}}{\sqrt{\text{Var}(\hat{\theta}_\omega - \hat{\theta}_{\omega'})}} = \frac{\sum_{i=1}^{n-1} i(\omega_i - \omega'_i)\zeta_i}{\sqrt{\text{Var}\left(\sum_{j=1}^{n-1} j(\omega_j - \omega'_j)\zeta_j\right)}} = \frac{\sum_{i=1}^{n-1} i\Omega_i \zeta_i}{\sqrt{\text{Var}\left(\sum_{j=1}^{n-1} j\Omega_j \zeta_j\right)}} \quad (7)$$

that is equivalent to the definition (3) with  $\Omega_i = \omega_i - \omega'_i$ . Therefore a test  $T_\Omega$  is defined by real vectors  $\Omega$  or  $\omega, \omega'$  satisfying the above normalisation conditions.

## 2.2. Sample Size Independent Tests

### 2.2.1. Scaling of Weights with Sample Size

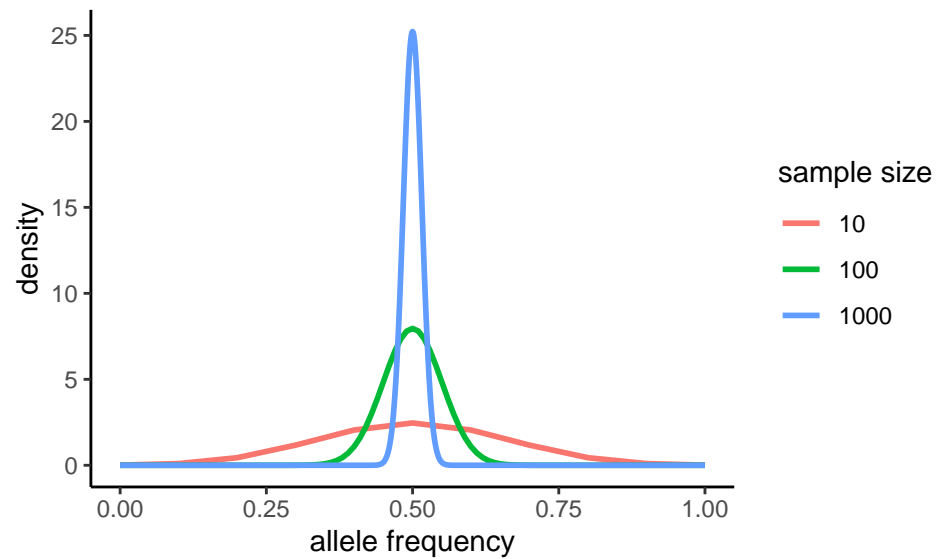
In this section we would like to remark that there are conditions that have to be imposed on the weights  $\Omega_i$  or  $\omega_i, \omega'_i$  to ensure that these tests are consistent and meaningful in their interpretation. In fact, the values (and even the number!) of these weights depend explicitly on sample size  $n$ . Since every conceivable test should be applied to samples of different sizes, then its definition involves a whole family of weights  $\{\Omega_i^{(n)}\}$  or  $\{\omega_i^{(n)}, \omega'_i{}^{(n)}\}$  with  $n = 2, 3, \dots, \infty$  and to define a test it is necessary to specify how these weights scale with  $n$ .

As an example of the weird effects of some choices of scaling, we consider the test for admixture of [8]. The weights of this test are  $\omega_i = \binom{n}{i} 2^{-n} (1 - 2^{-n+1})^{-1}$  and  $\omega'_i = 1/(n-1)$ . Suppose that the population under study shows an excess of alleles of frequency  $f$  between 0.3 and 0.4. The average weight of these frequencies, rescaled by the sample size, is 0.5 for  $n = 10$ , but it reduces to  $-0.75$  for  $n = 100$  and to  $-1.0$  for  $n = 1000$ . These weights are largely different, even in sign, therefore a strong excess of alleles in this range of frequency would show itself as either a positive or a negative value for this test, depending on the sample size! The reason can be understood by noticing that for  $n$  large, the binomial can be approximated by a Gaussian function of the allele frequencies  $f = i/n$  centred in  $f = 1/2$  and with variance  $1/4n$ . Therefore this weight function has a strong dependence on  $n$  when considered as a function of  $f$  and  $n$ . The changes of this weight function with sample size are apparent in the plot of Figure 1, which shows the actual function (rescaled by sample size) for  $n = 101,001,000$ .

In this example it is apparent that the interpretation of the results of this test depends on  $n$ . This means that the calibration of the test should be different for each possible sample size.

The weight-consistency requirement that we propose is that the result of the test should be almost independent of sample size. This requirement is equivalent to a condition on the scaling of the weights  $\Omega_i^{(n)}$  with  $n$ . Our proposal for a reasonable requirement on this scaling is the following: the relative weight of different frequencies in the population should remain approximately constant while varying the size of the sample. This condition ensures that at least for sufficiently large  $n$ , the average values of the test on samples of different size from the same population should be approximately independent on sample size, i.e., that the test should be weight-consistent. Note that this requirement has to do with the interpretation of the test, rather than with the usual definition of statistical consistency.

Tests that are not weight-consistent could be statistically consistent, but the interpretation of results in the left and right tails would not be assured to be independent of sample size.



**Figure 1.** Illustrative example of the dependence of the weight on sample size: weight  $\Omega$  as a function of  $i/n$  for the test for admixture by Achaz, plotted for different sample size  $n = 10$  (blue), 100 (red), 1000 (green).

To determine the scaling, we note that in limit of large  $n$ , the frequency spectrum approaches a continuum and we can define the weights as functions  $\Omega(f)$  or  $\omega(f), \omega'(f)$  with  $f \in (0,1)$  and  $\int_0^1 df \Omega(f) = 0, \int_0^1 df \omega(f) = \int_0^1 df \omega'(f) = 1$ . Since the ratio of the derived allele count and the sample size  $i/n$  is an unbiased estimator of the frequency  $f$  of the allele in the population (because  $E(i) = nf$ ), a simple scaling that satisfies the above requirement is:

$$\Omega_i^{(n)} \simeq \Omega(i/n) \quad \text{or} \quad \omega_i^{(n)} \simeq \omega(i/n), \omega_i'^{(n)} \simeq \omega'(i/n) \tag{8}$$

as proposed by some of the authors in [10].

In order to have the above approximate scaling while obeying the condition  $\sum_{i=1}^{n-1} \Omega_i = 0$ , there are two simple weight-consistent forms for the weights:

$$\Omega_i^{(n)} = \Omega\left(\frac{i}{n}\right) - \frac{1}{n-1} \sum_{j=1}^{n-1} \Omega\left(\frac{j}{n}\right) \tag{9}$$

where the last term is a (typically small) correction that enforce centredness of the test, or:

$$\Omega_i^{(n)} = \omega_i^{(n)} - \omega_i'^{(n)} = \frac{\omega\left(\frac{i}{n}\right)}{\sum_{j=1}^{n-1} \omega\left(\frac{j}{n}\right)} - \frac{\omega'\left(\frac{i}{n}\right)}{\sum_{j=1}^{n-1} \omega'\left(\frac{j}{n}\right)} \tag{10}$$

where the denominators are normalisation factors.

Typically, this second form (10) for the scaling is more consistent in practice and it is implicitly assumed for most of the existing tests. However, the above expressions give similar numerical results for most choices of the functions  $\Omega(f) = \omega(f) - \omega'(f)$ . In fact, if  $\Omega(f)$  is a limited and piecewise-continuous function, the difference between (9) and (10) is of order  $O(\Omega)/n$  (since it is a factor coming from the discretisation of the frequencies) and it does not have a relevant impact on the results of the test. Therefore, in these cases the two scaling relations (9) and (10) are practically equivalent.

Note that all the tests involving the Watterson estimator (that corresponds to  $\omega(f) \sim 1/f$ ) have additional subtleties that are discussed in the next section.

Example: Fay and Wu’s  $H$  test

This test was proposed in [7] to look for an excess of high-frequency derived alleles as a signal of selection. It can be defined by the weight functions  $\omega(f) = 2(1 - f)$  and  $\omega'(f) = 2f$  or alternatively  $\omega(f) = 1$  and  $\omega'(f) = 2f$ . The weights can be found following Equation (10). The resulting test is:

$$T_H = \frac{\frac{1}{n-1} \sum_{i=1}^{n-1} i \tilde{\zeta}_i - \frac{2}{n(n-1)} \sum_{i=1}^{n-1} i^2 \tilde{\zeta}_i}{\sqrt{\text{Var}\left(\frac{1}{n-1} \sum_{j=1}^{n-1} j \tilde{\zeta}_j - \frac{2}{n(n-1)} \sum_{j=1}^{n-1} j^2 \tilde{\zeta}_j\right)}} \tag{11}$$

The scaling defined in Equation (9), with weight function  $\Omega(f) = \omega(f) - \omega'(f) = 1 - 2f$ , gives precisely the same result.

Example:  $F(r, r')$  tests of Fu [15]

This large class of test is based on the comparison of two estimators with weights:

$$\omega_i = \frac{i^{-r}}{\sum_{j=1}^{n-1} j^{-r}} \quad , \quad \omega'_i = \frac{i^{-r'}}{\sum_{j=1}^{n-1} j^{-r'}} \tag{12}$$

that in the case  $r, r' < 1$  correspond precisely to the scaling (10) suggested above, with weight functions  $\omega(f) = (1 - r)f^{-r}$  and  $\omega'(f) = (1 - r')f^{-r'}$ . This can be easily verified by multiplying both the numerator and the denominator of  $\omega_i, \omega'_i$  by a factor  $(1 - r)/n^{-r}, (1 - r')/n^{-r'}$  respectively. The test by Fay and Wu corresponds actually to  $F(0, -1)$ .

The cases with  $r \geq 1$  or  $r' \geq 1$  involve weight functions with divergent integrals and will be discussed in the next section.

Note that the same weight functions with the scaling (9) would give rise to a slightly different test with weights:

$$\Omega_i = (1 - r) \left(\frac{i}{n}\right)^{-r} - (1 - r') \left(\frac{i}{n}\right)^{-r'} - \left( \frac{(1 - r) \sum_{j=1}^{n-1} j^{-r}}{(n - 1)n^{-r}} - \frac{(1 - r') \sum_{j=1}^{n-1} j^{-r'}}{(n - 1)n^{-r'}} \right) \tag{13}$$

that is not weight-consistent for weights of low frequency alleles, i.e., with  $i/n \lesssim n^{2/\max(r,r')}$ , and is therefore less interesting.

Example: Test for bottleneck of Achaz [8]

This test is another example of a test with an unwanted scaling:

$$\omega_i = \frac{e^{-\alpha i}}{\sum_{j=1}^{n-1} e^{-\alpha j}} \quad , \quad \omega'_i = \frac{1}{n - 1} \tag{14}$$

The weight function for this test is  $e^{-\alpha n f} \alpha n / (1 - e^{-\alpha n}) - 1$  that depends strongly on  $n$ , therefore this test is not weight-consistent in the above sense.

It is easy to build an equivalent test with the correct scaling by choosing the functions  $\omega(f) = \beta e^{-\beta f} / (1 - e^{-\beta}), \omega'(f) = 1$ . The resulting weights with the scaling (10) are:

$$\omega_i = \frac{e^{-\beta i/n}}{\sum_{j=1}^{n-1} e^{-\beta j/n}} = \frac{1 - e^{-\beta/n}}{1 - e^{-\beta(1-1/n)}} e^{-\beta(i-1)/n} \quad , \quad \omega'_i = \frac{1}{n - 1} \tag{15}$$

as discussed before. The optimal value reported in [8] is  $\alpha \simeq 0.9$  for  $n = 30$ . This value corresponds to  $\beta \simeq 27$ .

The test can also be implemented by choosing the scaling (9) and the weight function  $\Omega(f) = \omega(f) - \omega'(f) = \beta e^{-\beta f} / (1 - e^{-\beta}) - 1$ . The resulting weights are:

$$\begin{aligned} \Omega_i &= \frac{\beta e^{-\beta i/n}}{1 - e^{-\beta}} - 1 - \frac{1}{n-1} \left( \frac{\beta(1 - e^{-\beta(1-1/n)})}{(1 - e^{-\beta})e^{\beta/n}(1 - e^{-\beta/n})} - (n-1) \right) = \\ &= \frac{\beta(1 - e^{-\beta(1-1/n)})}{(1 - e^{-\beta})e^{\beta/n}(1 - e^{-\beta/n})} \cdot \left( \frac{1 - e^{-\beta/n}}{1 - e^{-\beta(1-1/n)}} e^{-\beta(i-1)/n} - \frac{1}{n-1} \right) \end{aligned} \quad (16)$$

that are equivalent to the weights (15) up to an irrelevant multiplicative factor (see Theorem A1). Therefore, in this case the two choices of scaling give precisely the same result.

### 2.2.2. Divergent Weights

As discussed above, the two choices of scaling in Equations (9) and (10) do not usually obtain sensibly different numerical results. However, there are important choices of  $\Omega(f)$  for which this approximate equivalence between (9) and (10) does not hold. These critical cases correspond to functions that diverge as  $1/f$  or faster near  $f = 0$  (or  $f = 1$ ). This divergence is not a real feature of the distribution, because the integral has a natural cutoff at the scale of the inverse population size  $f_{min} = 1/N$  (more precisely, the effective population size  $1/N_e$ , but this does not affect the discussion). However, in this case the integral  $\int_{1/N}^1 df \Omega(f)$  has a strong dependence on the cutoff  $1/N$  and therefore the function  $\Omega(f)$  itself should depend strongly on  $N$  to ensure proper normalisation.

If this dependence is contained in a multiplicative term in front of  $\omega(f)$  or  $\omega'(f)$  or both, then the second term in Equation (9) is not a small correction of order  $1/n$  as it happens with simple functions  $\Omega(f)$ , but rather it represents a relevant correction with a strong dependence on sample size  $n$  and population size  $N$ . The denominators in Equation (10) also show a strong dependence on  $n$  (that could not be avoided anyway) but not on  $N$ , and therefore this second scaling form should be used. The dependence on sample size is as strong as the dependence of the divergent integral from the cutoff. This can be easily understood by noticing that the sample size  $n$  plays the role of the cutoff in the sum over the frequencies that are present in the sample, which is the same role played by the population size  $N$  for the whole population; more formally, the denominator in Equation (10) can be bounded from above and from below by the divergent integral, and therefore the divergence of the denominator as  $n \rightarrow \infty$  will be the same as the divergence of the integral as its inverse cutoff (that is,  $N$ ) goes to infinity). For functions diverging as  $f^{-k}$  with  $k \geq 1$ , the dependence on  $n$  goes as  $n^{1-k}$  if  $k > 1$  or  $\log(n)$  for  $k = 1$ . This case always occurs when the test is built by comparing an estimator of  $\theta$  with the Watterson estimator, which corresponds to  $\omega(f) \sim 1/f$  and therefore has a logarithmic dependence on  $n$  given by the usual harmonic factor  $a_n = \sum_{j=1}^{n-1} 1/j \simeq \log(n) + \gamma + O(1/n)$ . A well-known example of this case is Tajima's  $D$  [5].

If the dependence of  $\Omega(f)$  on  $N$  is contained in an additive term that does not depend on  $f$ , it is the correction in (9) that does not depend on  $N$  and therefore the first scaling form is more appropriate. We do not know examples of tests of this kind in the literature, even if the test by Zeng et al. [16] can be interpreted also in this way.

#### Example: Tajima's $D$ test

This is the most known test for neutrality based on the frequency spectrum. It is given by the difference between the Tajima estimator  $\Pi$  [17] based on the nucleotide pairwise diversity  $\Pi$  and the Watterson estimator  $\theta_W$  [18] based on the number  $S$  of segregating sites, therefore it can be defined by the weight functions  $\omega(f) = 2(1 - f)$  for  $\Pi$  and  $\omega'(f) = 1/f \log(N)$  for the Watterson estimator. The latter function has an integral that diverges logarithmically near  $f = 0$ , and the corresponding dependence on  $N$  is contained

in the factor  $1/\log(N)$  that multiplies  $\omega'(f)$ , therefore the scaling (10) should be used. The result is the usual test:

$$T_D = \frac{\sum_{i=1}^{n-1} \frac{2i(n-i)}{n(n-1)} \xi_i - S/a_n}{\sqrt{\text{Var}\left(\sum_{j=1}^{n-1} \frac{2j(n-j)}{n(n-1)} \xi_j - S/a_n\right)}} = \frac{\Pi - S/a_n}{\sqrt{\text{Var}(\Pi - S/a_n)}} \quad (17)$$

Example: Test of Zeng et al. [16]

This test was proposed to look for an excess of high-frequency derived alleles compared to low-frequency alleles. It is defined by the weight functions  $\omega(f) = 1/f \log(N)$  and  $\omega'(f) = 1$ , the former corresponding to the Watterson estimator. Proceeding as in the above example, the result is:

$$T_E = \frac{S/a_n - \sum_{i=1}^{n-1} \frac{i}{(n-1)} \xi_i}{\sqrt{\text{Var}\left(S/a_n - \sum_{j=1}^{n-1} \frac{j}{(n-1)} \xi_j\right)}} \quad (18)$$

Note that, exceptionally, the scaling of this test can also be defined by (9), without modifying the result. This is a consequence of the two equivalent forms for the weight function,  $\Omega(f) = 1/f \log(N) - 1$  or  $\Omega(f) = 1/f - \log(N)$ .

### 2.2.3. Weights of Singletons

The above scaling (8) is valid in principle for all weights. However, in practice there is an important exception, that is, the weight  $\Omega_1$  of singletons. This is due to the fact that for  $n \ll N$ , the number of derived singletons  $\xi_1$  is the only estimator that is affected by very rare derived alleles (and often by sequencing errors, see [19]). More precisely,  $\xi_1$  is actually the only estimator sensitive to the deviations from neutrality in alleles of frequency  $1/N < f < 1/n$ , which represent a vast majority of the SNPs in the population and can contain interesting biological information. Therefore, if the contribution of these alleles is relevant for the test, we can enhance (or reduce) the weight  $\Omega_1$  by adding a factor  $\Omega_{ds}$ .

In the approach detailed in the previous sections, this additional contribution to  $\Omega_1$  is needed to take into account a contribution  $\Delta\Omega(f)$  to  $\Omega(f)$  of the form  $\Delta\Omega(f) = \Omega_{ds}I(f < \phi)/\phi$  with  $\phi \ll 1$ . As far as the maximum sample size never exceeds in practice  $n_{max} \ll 2/\phi$ , this function weights positively only alleles that appear as singletons.

Similarly,  $\omega_1$  and  $\omega'_1$  can be enhanced by  $\omega_{ds}$ ,  $\omega'_{ds}$  that correspond to contributions  $\Delta\omega(f) = \omega_{ds}I(f < \phi)/\phi$ ,  $\Delta\omega'(f) = \omega'_{ds}I(f < \phi)/\phi$ . The test of Fu and Li [6] fall into this case.

A similar argument applies also to the weights of the number of ancestral singletons, that is,  $\Omega_{n-1}$ ,  $\omega_{n-1}$ ,  $\omega'_{n-1}$  that can be enhanced by factors  $\Omega_{as}$ ,  $\omega_{as}$ , and  $\omega'_{as}$  respectively. However, this case is more rare, the only interesting example being the tests of Achaz [19] that avoid sequencing errors by neglecting both derived and ancestral singletons.

Summarising the results up to this section, a test  $T_\Omega$  is completely defined by a function  $\Omega(f)$  and two parameters  $\Omega_{ds}$ ,  $\Omega_{as}$  (that could depend on  $n$ ) satisfying the conditions:

$$\Omega_{ds} + \Omega_{as} + \int_0^1 df \Omega(f) = 0 \quad (19)$$

and determining the weights through the formula:

$$\Omega_i^{(n)} = \Omega\left(\frac{i}{n}\right) + \Omega_{ds}\delta_{i,1} + \Omega_{as}\delta_{i,n-1} - \frac{1}{n-1} \left( \Omega_{ds} + \Omega_{as} + \sum_{j=1}^{n-1} \Omega\left(\frac{j}{n}\right) \right) \quad (20)$$



or by a pair of functions  $\omega(f), \omega'(f)$  and parameters  $\omega_{ds}, \omega'_{ds}, \omega_{as}, \omega'_{as}$  satisfying:

$$\omega_{ds} + \omega_{as} + \int_0^1 df \omega(f) = \omega'_{ds} + \omega'_{as} + \int_0^1 df \omega'(f) = 1 \tag{21}$$

and resulting in this formula for the scaling of the weights:

$$\Omega_i^{(n)} = \frac{\omega_{ds}\delta_{i,1} + \omega_{as}\delta_{i,n-1} + \omega\left(\frac{i}{n}\right)}{\omega_{ds} + \omega_{as} + \sum_{j=1}^{n-1} \omega\left(\frac{j}{n}\right)} - \frac{\omega'_{ds}\delta_{i,1} + \omega'_{as}\delta_{i,n-1} + \omega'\left(\frac{i}{n}\right)}{\omega'_{ds} + \omega'_{as} + \sum_{j=1}^{n-1} \omega'\left(\frac{j}{n}\right)} \tag{22}$$

As showed in the examples above and below, most of the tests in the literature have this general scaling, with the only exceptions the ones contained in [8] that are not weight-consistent in the above sense.

Example: Fu and Li’s  $F$  test

This test looks for an excess of very rare derived alleles as a possible signature of negative selection [6]. The only nonzero weights are  $\omega_{ds} = 1$  and  $\omega'(f) = 1/f \log(N)$ , while  $\omega(f) = \omega'_{ds} = \omega_{as} = \omega'_{as} = 0$ . The resulting test is:

$$T_F = \frac{\xi_1 - S/a_n}{\sqrt{\text{Var}(\xi_1 - S/a_n)}} \tag{23}$$

Note that this test has both singleton weights and a divergent weight function.

Example: Error-corrected tests of Achaz [19]

This class of tests is an attempt to correct for sequencing errors and biases in the data by removing the alleles where most of the problems manifest themselves, i.e., singletons (both ancestral and derived). With a slight generalisation of the proposal in [19], the weights of the singletons are chosen in such a way to cancel precisely the contributions of the weight functions:

$$\Omega_{ds} = -\Omega\left(\frac{1}{n}\right), \Omega_{as} = -\Omega\left(1 - \frac{1}{n}\right) \tag{24}$$

or:

$$\omega_{ds} = -\omega\left(\frac{1}{n}\right), \omega_{as} = -\omega\left(1 - \frac{1}{n}\right), \omega'_{ds} = -\omega'\left(\frac{1}{n}\right), \omega'_{as} = -\omega'\left(1 - \frac{1}{n}\right) \tag{25}$$

therefore, the final weights of derived or ancestral singletons are zero. These corrections can be applied in principle to any weight function.

#### 2.2.4. Alternative Choices of Scaling

The choice of scaling discussed in the previous sections represents a quite simple and effective way to fix the dependence on  $n$  of a newly devised test. However, other choices are possible whose weights differ from the above ones for small  $n$ . The reason is that for  $n$  not too large, both the variance of order  $f(1-f)/n \simeq i(n-i)/n^3$  in the estimation of the frequency  $f = i/n$  and the related uncertainty about how the frequencies are actually weighted in the test become important. This uncertainty originates from the (binomial) sampling of individuals from the population and there is some degree of arbitrariness in deciding how to account for it. Moreover, tests that take it into account could be inconsistent in the above sense.

A possible choice of scaling that uses the binomial sampling is the following: considering  $\omega(f)$ ,  $\omega'(f)$  as frequency distributions, the weights  $\omega_i$ ,  $\omega'_i$  are assigned from  $\omega(f)$ ,  $\omega'(f)$  through the same binomial sampling that is done for allele spectra, that is:

$$\omega_i = \frac{\int_0^1 df \binom{n}{i} f^i (1-f)^{n-i} \omega(f)}{\int_0^1 df (1-f^n - (1-f)^n) \omega(f)} \tag{26}$$

$$\omega'_i = \frac{\int_0^1 df \binom{n}{i} f^i (1-f)^{n-i} \omega'(f)}{\int_0^1 df (1-f^n - (1-f)^n) \omega'(f)} \tag{27}$$

A simple example of this scaling (but with an highly divergent weight function) is given by the test for admixture [8] discussed previously. Optimal tests also follow this scaling.

Example: Test for admixture of Achaz [8]

This test is apparently not consistent and it does not follow the scaling (8). However, it follows another scaling related to the allele sampling. To understand this, consider the weight functions  $\omega(f) = \delta(f - 1/2)$ ,  $\omega'(f) = 1$  where  $\delta(f - 1/2)$  is a Dirac delta function centred in 1/2. (The Dirac delta  $\delta(f - a)$  is a function whose value is 0 if  $f \neq a$  and  $+\infty$  if  $f = a$ . The integral  $\int \delta(f - a)g(f)df$  is  $g(a)$  if  $a$  is inside the range of integration and 0 otherwise. Technically speaking it is not a mathematical function, but a distribution, i.e., an element of a dual space of regular functions.) If we scale the weights according to (26) and (27), that is:

$$\omega_i = \frac{\int_0^1 df \binom{n}{i} f^i (1-f)^{n-i} \omega(f)}{\int_0^1 df (1-f^n - (1-f)^n) \omega(f)} = \frac{\binom{n}{i} 2^{-n}}{1 - 2^{-n+1}} \tag{28}$$

$$\omega'_i = \frac{\int_0^1 df \binom{n}{i} f^i (1-f)^{n-i} \omega'(f)}{\int_0^1 df (1-f^n - (1-f)^n) \omega'(f)} = \frac{1}{n-1} \tag{29}$$

then the corresponding test is precisely the one proposed by Achaz. Note that the strong dependence of the test from sample size does not come only from the choice of scaling, but also from the weight function chosen, that is highly divergent.

### 2.3. Optimal Tests

#### 2.3.1. On the Existence of Generic Tests

An interesting question on the way to build good linear tests is the following: do there exist generic tests? A completely generic test for neutrality should be able to detect any deviation from the spectrum of the null model that is sufficiently large. Unfortunately, these tests do not exist. In fact, for every test defined by a set of weights  $\Omega_i$  it is possible to find a spectrum  $\xi_i = \alpha/ia_n + (1-\alpha)\Delta_i$  that is maximally different from the standard spectrum at least in a range of frequencies and is nevertheless undetectable by the test because its average value on this spectrum is zero. This is expressed in a more formal way in the following theorem, which shows that even the complete lack of alleles in some range of frequencies could not be always detected.

**Theorem 1.** For every set of  $n$  real weights  $\Omega_i$  with  $\sum_i \Omega_i = 0$ , there is a set of  $n$  real numbers  $\Delta_i \neq \text{const}/i$  and a parameter  $\alpha \in [0, 1]$  that satisfy the conditions:

$$\sum_i i\Omega_i\Delta_i = 0 \quad , \quad \min_{i \in [1, n-1]} \left( \alpha \frac{1}{ia_n} + (1-\alpha)\Delta_i \right) = 0 \tag{30}$$

The above limitation is not a consequence of the small sample size. This can be seen for example in the framework of the scaling theory discussed in this paper. In fact, for large sample sizes, the weights can be approximated by a weight function  $\Omega(f)$ . In this context it is possible to prove the next theorem, that is a continuous equivalent of the previous one.

**Theorem 2.** For every piece-wise continuous weight function  $\Omega(f) \in L^1_{[1/N,1]}$  such that  $\int_{1/N}^1 \Omega(f)df = 0$ , there is a smooth function  $\Delta(f) \neq \text{const}/f$  and a parameter  $\alpha \in [0, 1]$  that satisfy the conditions:

$$\int_{1/N}^1 df f\Omega(f)\Delta(f) = 0 \quad , \quad \inf_{f \in [0,1]} \left( \alpha \frac{1}{f \log(N)} + (1 - \alpha)\Delta(f) \right) = 0 \quad (31)$$

Note that in principle this problem can be solved using multiple tests. In fact, multiple tests should be able to detect any strong deviation from the null spectrum, provided that the number of these tests is large enough, as can be seen from the following theorem.

**Theorem 3.** Given at least  $n - 2$  linearly independent sets of  $n - 1$  real weights  $\Omega_i$  with  $\sum_i \Omega_i = 0$ , it is not possible to find a set of real numbers  $\Delta_i \neq \text{const}/i$  such that  $\sum_i i\Omega_i\Delta_i = 0$ .

This last theorem is only a formal result and the requirement of  $n - 2$  independent tests is too strong. In practice, a small (but good) set of tests can detect most of the reasonable and interesting deviations for realistic spectra.

The above theorems can be extended to the folded spectrum. In this section and the next ones, we will consider only tests based on the unfolded spectrum. The generalisation of the discussion to the folded spectrum is usually straightforward after substituting  $\xi_i$  ( $i = 1 \dots n - 1$ ) with  $\eta_i$  ( $i = 1 \dots \lfloor n/2 \rfloor$ ).

### 2.3.2. Optimal Tests and Their Geometric Structure

From the theorems of the previous section, it is clear that a single test cannot detect all the possible deviations occurring in complicated evolutionary scenarios. However, it is still possible to optimise neutrality tests for a specific alternative evolutionary scenario. A simple optimality condition has been proposed by some of the authors in [10] in order to maximise the power of the test to detect a fixed alternative scenario. We use a different notation  $E()$  and  $\mathcal{E}()$  for the expected value with respect to the null scenario (neutral model)  $E()$  and the alternative scenario  $\mathcal{E}()$ . If the null spectrum is  $E(\xi_i) = \theta L \xi_i^0$  and the expected spectrum of the alternative scenario is  $\mathcal{E}(\xi_i) = \theta L \bar{\xi}_i$  (note that it may include an average of frequencies in a number of different related scenarios, e.g., averaging over some of the parameters of the scenario), the condition for optimal tests is the maximisation of the average result of the test under the alternative scenario:

$$\mathcal{E}(T_\Omega) = \frac{\sum_{i=1}^{n-1} \Omega_i \theta L \bar{\xi}_i / \xi_i^0}{\sqrt{\text{Var}\left(\sum_{j=1}^{n-1} \Omega_j \bar{\xi}_j / \xi_j^0\right)}} \quad (32)$$

This condition is based on the observation that the tests have mean zero and variance 1; therefore, if the distributions of the results of the tests are similar, the maximisation of the average value of the test should correspond to the maximisation of the average power of the test. It is also possible to maximise directly the power of the test, taking into account the different distribution of the results under the null and the alternative model; this possibility will be pursued in Section 3.1.

Interestingly, optimal tests show a geometric structure which becomes apparent after defining the scalar product between spectra:

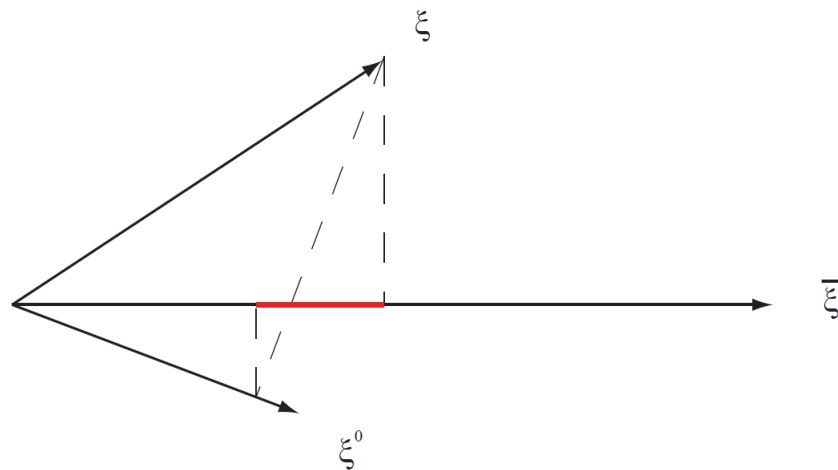
$$\langle\langle \xi', \xi'' \rangle\rangle \equiv \sum_{i,j} \xi'_i c_{ij}^{-1} \xi''_j \quad (33)$$

where  $c_{ij}^{-1}$  is the inverse of the covariance matrix  $\text{Cov}(\xi_i, \xi_j)$ . Since  $\text{Cov}(\xi_i, \xi_j)$  is symmetric and positive, its inverse is also symmetric and positive, i.e., it is a positive bilinear form, therefore, the above expression defines a scalar product. Then, the optimal test for an alternative spectrum  $\bar{\xi}$  can be written in the elegant form:

$$T_O = \frac{\langle \langle \bar{\xi}, \bar{\xi} \rangle \rangle - \langle \langle \bar{\xi}, \xi^0 \rangle \rangle \langle \langle \xi^0, \bar{\xi} \rangle \rangle / \langle \langle \xi^0, \xi^0 \rangle \rangle}{\sqrt{\langle \langle \bar{\xi}, \bar{\xi} \rangle \rangle - \langle \langle \xi^0, \bar{\xi} \rangle \rangle^2 / \langle \langle \xi^0, \xi^0 \rangle \rangle}} \tag{34}$$

This expression can be easily obtained as a special case of the general formula (50) that we will discuss later in the context of nonlinear tests. A direct proof of this result can be found in [10] after substituting the scalar products with the definition (33).

The numerator of the test is actually the matrix element between  $\bar{\xi}$  and  $\xi$  of the linear operator  $1 - P_{\xi^0}$ , where  $P_{\xi^0}$  is the projection operator along  $\xi^0$ . In other words, it is proportional to the difference between the length of the projection of  $\xi$  on  $\bar{\xi}$  and the length of the projection on  $\bar{\xi}$  of the spectrum obtained by the projection of  $\xi$  on  $\xi^0$ , as illustrated in Figure 2.



**Figure 2.** Geometrical representation of the numerator of the optimal test  $T_O$  in (34). The length of the red line segment corresponds to the value of the numerator.

From this geometrical interpretation it is clear that if the spectrum  $\xi$  corresponds to the null spectrum  $\theta L \xi^0$ , then the two projections are equal and the result of the test is zero. On the other side, if the spectrum is the alternative spectrum  $\theta L \bar{\xi}$ , then the value of the test is:

$$T_O^{(max)} = \theta L \sqrt{\langle \langle \bar{\xi}, \bar{\xi} \rangle \rangle - \frac{\langle \langle \xi^0, \bar{\xi} \rangle \rangle^2}{\langle \langle \xi^0, \xi^0 \rangle \rangle}} \tag{35}$$

which is the maximum value over all possible tests in the alternative scenario. The same expression, but with a minus sign, corresponds to the minimum value.

The denominator of the test is the square root of the matrix element of the linear operator  $1 - P_{\xi^0}$  between  $\bar{\xi}$  and itself. Note that both the numerator and the denominator of the test do not change by adding any (possibly negative) multiple of  $\xi^0$  to  $\bar{\xi}$ , because  $\xi^0$  lies in the kernel of  $1 - P_{\xi^0}$ . This means that optimal tests depend only on the expected deviations from the null spectrum in the alternative scenario. The result of the test is maximum when the deviations of the data from the null spectrum correspond exactly to the expected ones, and it is minimum when they are opposite to the expected ones.

2.4. Beyond Linear Neutrality Tests

2.4.1. Quadratic and Nonlinear Tests

Almost all the neutrality tests proposed in the literature are linear in the spectrum  $\xi_i$ . As far as we know, there is only one exception, namely the  $G_{\xi}$  test of Fu [11]. This test is a quadratic polynomial reminiscent of Hotelling’s  $t^2$  statistics for the different components of the spectra:

$$G = \sum_{i,j=1}^{n-1} c_{ij}^{-1} (\xi_i - \theta L \xi_i^0) (\xi_j - \theta L \xi_j^0) \tag{36}$$

where  $c_{ij}^{-1}$  is the inverse of the covariance matrix  $\text{Cov}(\xi_i, \xi_j)$ . Actually, the test proposed by Fu is an approximation to this test with a different normalisation, namely:

$$G_{\xi} = \frac{1}{n-1} \sum_{i=1}^{n-1} \frac{(\xi_i - \theta L \xi_i^0)^2}{\text{Var}(\xi_i)} \tag{37}$$

In this approximation, the off diagonal terms in the covariance can be neglected [9,11]. For large samples, the distribution of the results of the test  $G$  tends to a  $\chi^2$  distribution with  $n - 1$  degrees of freedom.

Fu’s approach cannot be extended to general quadratic or higher order tests, because the distribution of the results of the test would be generally unknown and not positive definite. For this reason we propose to rescale the tests to have zero mean and variance 1. With this normalisation, we expect that the distribution would asymptotically converge to a Gaussian  $N(0, 1)$  for all tests. As an example, the (re)normalised version of Fu’s test would be:

$$T_G = \frac{\sum_{i,j=1}^{n-1} c_{ij}^{-1} (\xi_i - \theta L \xi_i^0) (\xi_j - \theta L \xi_j^0) - (n-1)}{\sqrt{\text{Var}\left(\sum_{i,j=1}^{n-1} c_{ij}^{-1} (\xi_i - \theta L \xi_i^0) (\xi_j - \theta L \xi_j^0) - (n-1)\right)}} \tag{38}$$

Since the only difference between this test and the original one is the normalisation and a shift by a constant factor  $n - 1$ , the power of the test is the same.

Now we present a systematic discussion of nonlinear tests that are generic polynomials (or eventually power series) in the spectrum  $\xi_i$ . All the tests are rescaled to be centred (i.e., to have zero mean) and have variance 1. We denote by  $\mu_{ijk\dots}$  the moments of the spectrum under the null model, that is,  $\mu_{ijk\dots} = E(\xi_i \xi_j \xi_k \dots)$ . With this definition,  $\mu_i = \theta L \xi_i^0$ . Note that all these moments depend on  $\theta$ . In the approximation of unlinked (independent) sites and small  $\theta$ , the second moments are equal to  $\mu_{ij} = \theta L \xi_i^0 \delta_{ij} + \theta^2 L^2 \xi_i^0 \xi_j^0$ .

The weights of general nonlinear tests can depend explicitly on  $\theta$ , as seen in the previous example. To compute the values of the tests, the (unknown) parameter  $\theta$  is substituted with an estimator  $\hat{\theta}$ . Unlike the linear case, in this case there are two different classes of tests, related to the dependence on  $\hat{\theta}$  of the centredness: *strongly centred* and *weakly centred* tests.

Strongly centred tests are tests that are always centred for any value of  $\hat{\theta}$ , even if it is different from the actual value of  $\theta$ . The general form for strongly centred tests is:

$$T_{\Omega} = \frac{\sum_{i=1}^{n-1} \Omega_i^{(1)} \xi_i + \sum_{i,j=1}^{n-1} \Omega_{ij}^{(2)} \xi_i \xi_j + \sum_{i,j,k=1}^{n-1} \Omega_{ijk}^{(3)} \xi_i \xi_j \xi_k + \dots}{\sqrt{\text{Var}\left(\sum_{i=1}^{n-1} \Omega_i^{(1)} \xi_i + \sum_{i,j=1}^{n-1} \Omega_{ij}^{(2)} \xi_i \xi_j + \sum_{i,j,k=1}^{n-1} \Omega_{ijk}^{(3)} \xi_i \xi_j \xi_k + \dots\right)}} \tag{39}$$

with the real symmetric weights  $\Omega_{ijk\dots}^{(n)}$  satisfying the set of conditions:

$$0 = \sum_{i=1}^{n-1} \Omega_i^{(1)} \mu_i^{(m)} + \sum_{i,j=1}^{n-1} \Omega_{ij}^{(2)} \mu_{ij}^{(m)} + \dots \quad , \quad m = 1, 2, 3 \dots \tag{40}$$

where we denote by  $\mu_{ijk\dots}^{(p)}$  the  $p$ -th term of the Taylor expansion with respect to  $\theta L$  of  $\mu_{ijk\dots}$  (in other words,  $\mu_{ijk\dots} = \sum_p \theta^p L^p \mu_{ijk\dots}^{(p)}$ , where the coefficients  $\mu_{ijk\dots}^{(p)}$  are independent on  $\theta$ ). The sum can be limited to polynomials of some finite order in  $\xi_i$  or it can be a (convergent) power series. If we introduce the notation  $\mathbf{I} = ijk\dots$  to denote a group of  $n_{\mathbf{I}}$  indices, we can rewrite the test in the simpler form:

$$T_{\Omega} = \frac{\sum_{\mathbf{I}} \Omega_{\mathbf{I}}^{(n_{\mathbf{I}})} (\xi \dots \xi)_{\mathbf{I}}}{\sqrt{\text{Var}(\sum_{\mathbf{I}} \Omega_{\mathbf{I}}^{(n_{\mathbf{I}})} (\xi \dots \xi)_{\mathbf{I}})}} \tag{41}$$

with the conditions:

$$0 = \sum_{\mathbf{I}} \Omega_{\mathbf{I}}^{(n_{\mathbf{I}})} \mu_{\mathbf{I}}^{(m)} \quad , \quad m = 1, 2, 3 \dots \tag{42}$$

If we constrain these tests to be first order polynomials in  $\xi_i$ , we recover the linear case with  $\Omega_i^{(1)} = \Omega_i / \xi_i^0$ . Note that linear tests are always strongly centred. In fact, in the infinite site model the spectrum is always proportional to  $\theta$ , which consequently factorises out by linearity and therefore has no effect on the centredness.

Weakly centred tests are tests that are centred but not strongly centred, i.e., they are centred if and only if  $\hat{\theta} = \theta$ . The general form for weakly centred tests is:

$$T_{\Gamma} = \frac{\gamma + \sum_{i=1}^{n-1} \Gamma_i^{(1)} \xi_i + \sum_{i,j=1}^{n-1} \Gamma_{ij}^{(2)} \xi_i \xi_j + \sum_{i,j,k=1}^{n-1} \Gamma_{ijk}^{(3)} \xi_i \xi_j \xi_k + \dots}{\sqrt{\text{Var}(\gamma + \sum_{i=1}^{n-1} \Gamma_i^{(1)} \xi_i + \sum_{i,j=1}^{n-1} \Gamma_{ij}^{(2)} \xi_i \xi_j + \sum_{i,j,k=1}^{n-1} \Gamma_{ijk}^{(3)} \xi_i \xi_j \xi_k + \dots)}} \tag{43}$$

with the condition:

$$0 = \gamma + \sum_{i=1}^{n-1} \Gamma_i^{(1)} \mu_i + \sum_{i,j=1}^{n-1} \Gamma_{ij}^{(2)} \mu_{ij} + \sum_{i,j,k=1}^{n-1} \Gamma_{ijk}^{(3)} \mu_{ijk} + \dots \tag{44}$$

where the  $\Gamma_{ijk\dots}$  are real symmetric weights, possibly dependent on  $\theta$ . We can simplify these expressions using the same notation as above, obtaining the simpler form:

$$T_{\Gamma} = \frac{\gamma + \sum_{\mathbf{I}} \Gamma_{\mathbf{I}}^{(n_{\mathbf{I}})} (\xi \dots \xi)_{\mathbf{I}}}{\sqrt{\text{Var}(\gamma + \sum_{\mathbf{I}} \Gamma_{\mathbf{I}}^{(n_{\mathbf{I}})} (\xi \dots \xi)_{\mathbf{I}})}} \tag{45}$$

with the condition:

$$0 = \gamma + \sum_{\mathbf{I}} \Gamma_{\mathbf{I}}^{(n_{\mathbf{I}})} \mu_{\mathbf{I}} \tag{46}$$

Additionally, for this class of tests the sum can be limited to polynomials of fixed order or extended to power series. Note that the rescaled version of the  $G$  test by Fu presented above belongs to this class.

The important difference between strongly and weakly centred tests is related to the robustness with respect to a biased estimation of  $\theta$ . Since the class of weakly centred tests is much larger than the class of strongly centred ones, it should be easier to find powerful tests in the former class than in the latter. However, even if weakly centred tests could be more powerful, they would not be centred in scenarios where the value of  $\theta$  could not be estimated precisely. On the other side, strongly centred tests are robust with respect to a bad estimation of  $\theta$  and therefore they would be preferable in scenarios where an unbiased estimation of  $\theta$  is troublesome.

The scaling rule (8) can be generalised to nonlinear tests in terms of functions  $\Omega^{(n_I)}(f_1, f_2 \dots f_{n_I})$  for strongly centred and  $\Gamma^{(n_I)}(f_1, f_2 \dots f_{n_I})$  for weakly centred tests:

$$\Omega_I^{(n_I)} \simeq \frac{1}{n^{n_I}} \Omega^{(n_I)} \left( \frac{i}{n}, \frac{j}{n}, \frac{k}{n} \dots \right) \tag{47}$$

$$\Gamma_I^{(n_I)} \simeq \frac{1}{n^{n_I}} \Gamma^{(n_I)} \left( \frac{i}{n}, \frac{j}{n}, \frac{k}{n} \dots \right) \tag{48}$$

Fixing the precise scaling is more ambiguous than in the linear case because there are many different ways to preserve centredness. For this reason, the choice of scaling would be different for strongly and weakly centred tests and will not be discussed here.

All the possible nonlinear neutrality tests based on the frequency spectrum fall into one of the two classes presented in this section and have the form (41) and (42), or (45) and (46). Since both these classes contain an infinite number of possible choices of weights, the only reasonable criterion to study general nonlinear tests is to select the most powerful or interesting ones. Apart from the Hotelling choice of Fu [11], the most interesting choice is apparently the subclass of nonlinear optimal tests, which will be discussed in the next sections.

#### 2.4.2. Strongly Centred Optimal Tests

As discussed for the linear case, optimal tests depend on the expected alternative scenario. In the nonlinear case, in principle it would be possible to find generic optimal tests, but there is no clear framework to obtain them. For this reason we limit our study to the case of optimal tests for a specific alternative scenario. We denote by  $\tilde{\mu}_{ijk\dots} = \mathcal{E}(\xi_i \xi_j \xi_k \dots)$  the moments of the alternative spectrum for this scenario.

Since we use the same normalisation for linear and nonlinear tests, the optimality condition corresponds to the maximisation of the expected value of the test under the alternative scenario:

$$\mathcal{E}(T_\Omega) = \frac{\sum_I \Omega_I^{(n_I)} \tilde{\mu}_I}{\sqrt{\text{Var}(\sum_I \Omega_I^{(n_I)} (\xi \dots \xi)_I)}} \tag{49}$$

and can be justified as in the linear case.

We denote by  $\tilde{\mathbf{I}}$  the ordered sequence of the indices contained in  $\mathbf{I} = ijk\dots$  and by  $\sigma(\mathbf{I})$  the number of distinct permutations of the sequence  $\mathbf{I}$ , i.e., the total number of permutations divided by the number of permutations that leave  $\mathbf{I}$  invariant. The main result for the optimal weights is presented in this theorem.

**Theorem 4.** *The maxima of  $\mathcal{E}(T_\Omega)$  correspond to the weights:*

$$\Omega_I^{(n_I)} = \frac{1}{\sigma(\mathbf{I})} \left( \sum_{\tilde{\mathbf{L}}} C_{\tilde{\mathbf{I}}\tilde{\mathbf{L}}}^{-1} \tilde{\mu}_{\tilde{\mathbf{L}}} - \sum_k \sum_l \sum_{\tilde{\mathbf{L}}} C_{\tilde{\mathbf{I}}\tilde{\mathbf{L}}}^{-1} \mu_{\tilde{\mathbf{L}}}^{(k)} \mathcal{M}_{kl} \sum_{\tilde{\mathbf{J}}, \tilde{\mathbf{K}}} \mu_{\tilde{\mathbf{J}}}^{(l)} C_{\tilde{\mathbf{J}}\tilde{\mathbf{K}}}^{-1} \tilde{\mu}_{\tilde{\mathbf{K}}} \right) \tag{50}$$

where the matrices  $C_{\tilde{\mathbf{I}}\tilde{\mathbf{J}}}^{-1}$  and  $\mathcal{M}_{kl}$  satisfy the identities:

$$\sum_{\tilde{\mathbf{K}}} C_{\tilde{\mathbf{I}}\tilde{\mathbf{K}}}^{-1} (\mu_{\tilde{\mathbf{K}}\tilde{\mathbf{J}}} - \mu_{\tilde{\mathbf{K}}}\mu_{\tilde{\mathbf{J}}}) = \delta_{\tilde{\mathbf{I}}\tilde{\mathbf{J}}} \tag{51}$$

$$\sum_r \mathcal{M}_{kr} \sum_{\tilde{\mathbf{I}}, \tilde{\mathbf{L}}} \mu_{\tilde{\mathbf{I}}}^{(r)} C_{\tilde{\mathbf{I}}\tilde{\mathbf{L}}}^{-1} \mu_{\tilde{\mathbf{L}}}^{(l)} = \delta_{kl} \tag{52}$$

Moreover, the variance of the corresponding unnormalised test under the null model is equal to its expected value under the alternative model:

$$\text{Var}\left(\sum_{\mathbf{I}} \Omega_{\mathbf{I}}^{(n_1)} (\xi \dots \xi)_{\mathbf{I}}\right) = \sum_{\mathbf{I}} \Omega_{\mathbf{I}}^{(n_1)} \bar{\mu}_{\mathbf{I}} \tag{53}$$

Note that in general all the weights of the above optimal solution (50) are nonzero, therefore the maximum average value of the test for optimal tests built on polynomials of degree  $d$  increases with the degree  $d$ . This suggests that optimal tests of higher degree should be more powerful than linear optimal tests.

We provide explicit formulae for the above weights for the optimal quadratic test in the independent sites approximation. Given  $E(\xi_i) = \mu_i$  and  $\mathcal{E}(\xi_i) = \bar{\mu}_i$ , the relevant weights  $\Omega_{\mathbf{I}}^{(n_1)}$  are:

$$\Omega_i^{(1)} = (\Sigma_{\mu} + 2 - \Sigma_{\bar{\mu}}) \left( \frac{\bar{\mu}_i - \Sigma_{\bar{\mu}}}{\mu_i - \Sigma_{\mu}} \right) - \frac{1}{2} \left( \frac{\bar{\mu}_i^2 - \Sigma_{\bar{\mu}}^2}{\mu_i^2 - \Sigma_{\mu}^2} \right) \tag{54}$$

$$\Omega_{ii}^{(2)} = - \left( \frac{\bar{\mu}_i - \Sigma_{\bar{\mu}}}{\mu_i - \Sigma_{\mu}} \right) + \frac{1}{2} \left( \frac{\bar{\mu}_i^2 - \Sigma_{\bar{\mu}}^2}{\mu_i^2 - \Sigma_{\mu}^2} \right) \tag{55}$$

$$\Omega_{ij}^{(2)} = \frac{1}{2} \left[ \left( \frac{\bar{\mu}_i \bar{\mu}_j - \Sigma_{\bar{\mu}}^2}{\mu_i \mu_j - \Sigma_{\mu}^2} \right) - \left( \frac{\bar{\mu}_i - \Sigma_{\bar{\mu}}}{\mu_i - \Sigma_{\mu}} \right) - \left( \frac{\bar{\mu}_j - \Sigma_{\bar{\mu}}}{\mu_j - \Sigma_{\mu}} \right) \right] \tag{56}$$

where  $\Sigma_{\mu} = \sum_{i=1}^{n-1} \mu_i$  and  $\Sigma_{\bar{\mu}} = \sum_{i=1}^{n-1} \bar{\mu}_i$ . All these formulae are also valid for the folded spectrum if the appropriate  $\mu_i$  and  $\bar{\mu}_i$  are used. These results are discussed in Appendix A.7.

For optimal tests of higher degree, explicit expressions become cumbersome and the numerical implementation of the test (50) and the matrices (51) and (52) is more convenient.

### 2.4.3. Weakly Centred Optimal Tests

In this case the optimality condition corresponds to the maximisation of the expression:

$$\mathcal{E}(T_{\Gamma}) = \frac{\gamma + \sum_{\mathbf{I}} \Gamma_{\mathbf{I}}^{(n_1)} \bar{\mu}_{\mathbf{I}}}{\sqrt{\text{Var}\left(\gamma + \sum_{\mathbf{I}} \Gamma_{\mathbf{I}}^{(n_1)} (\xi \dots \xi)_{\mathbf{I}}\right)}} \tag{57}$$

with the same condition:

$$0 = \gamma + \sum_{\mathbf{I}} \Gamma_{\mathbf{I}}^{(n_1)} \mu_{\mathbf{I}} \tag{58}$$

The simplest case corresponds to a first order polynomial:

$$T_{\Gamma} = \frac{\gamma + \sum_{i=1}^{n-1} \Gamma_i^{(1)} \xi_i}{\sqrt{\gamma^2 + 2\gamma \sum_{j=1}^{n-1} \Gamma_j^{(1)} \mu_j + \sum_{j=1}^{n-1} \sum_{k=1}^{n-1} \Gamma_j^{(1)} \Gamma_k^{(1)} \mu_{jk}}} \tag{59}$$

whose maximum corresponds to the optimal weights:

$$\Gamma_i^{(1)} = \sum_{j=1}^{n-1} c_{ij}^{-1} (\bar{\mu}_j - \mu_j) \quad , \quad \gamma = - \sum_{j=1}^{n-1} \sum_{k=1}^{n-1} \mu_j c_{jk}^{-1} (\bar{\mu}_k - \mu_k) \tag{60}$$



where  $c_{ij}^{-1}$  is the inverse matrix of the covariance matrix  $c_{ij} = \mu_{ij} - \mu_i\mu_j$ . Since  $\gamma \neq 0$  for this optimal test, the value of this test for the specific scenario for which it is built is larger than the value of the corresponding linear optimal test. In fact the maximum of the test is:

$$\mathcal{E}(T_\Gamma) = \sqrt{\sum_{j=1}^{n-1} \sum_{k=1}^{n-1} (\bar{\mu}_j - \mu_j) c_{jk}^{-1} (\bar{\mu}_k - \mu_k)} \tag{61}$$

that should be compared to the maximum of the optimal test for the linear case, which can be rewritten as:

$$\mathcal{E}(T_\Omega)_{linear} = \sqrt{\sum_{j=1}^{n-1} \sum_{k=1}^{n-1} (\bar{\mu}_j - \mu_j) c_{jk}^{-1} (\bar{\mu}_k - \mu_k) - \frac{\left(\sum_{j=1}^{n-1} \sum_{k=1}^{n-1} \mu_j c_{jk}^{-1} (\bar{\mu}_k - \mu_k)\right)^2}{\sum_{j=1}^{n-1} \sum_{k=1}^{n-1} \mu_j c_{jk}^{-1} \mu_k}} \tag{62}$$

The comparison shows clearly that nonlinear optimal tests are always more powerful than linear optimal tests for the same scenario.

The form of the results for the general case is similar to this simple case.

**Theorem 5.** *The maxima of  $\mathcal{E}(T_\Gamma)$  correspond to the weights:*

$$\Gamma_{\mathbf{I}}^{(n_1)} = \frac{1}{\sigma(\mathbf{I})} \sum_{\mathbf{J}} C_{\mathbf{IJ}}^{-1} (\bar{\mu}_{\mathbf{J}} - \mu_{\mathbf{J}}) \quad , \quad \gamma = - \sum_{\mathbf{J}} \mu_{\mathbf{J}} \sum_{\mathbf{K}} C_{\mathbf{JK}}^{-1} (\bar{\mu}_{\mathbf{K}} - \mu_{\mathbf{K}}) \tag{63}$$

where  $C_{\mathbf{IJ}}^{-1}$  satisfied the identity:

$$\sum_{\mathbf{K}} C_{\mathbf{IK}}^{-1} (\mu_{\mathbf{KJ}} - \mu_{\mathbf{K}}\mu_{\mathbf{J}}) = \delta_{\mathbf{IJ}} \tag{64}$$

Moreover, the variance of the corresponding unnormalised test under the null model is equal to its expected value under the alternative model:

$$\text{Var} \left( \gamma + \sum_{\mathbf{I}} \Gamma_{\mathbf{I}}^{(n_1)} (\xi \dots \xi)_{\mathbf{I}} \right) = \sum_{\mathbf{I}} \Gamma_{\mathbf{I}}^{(n_1)} \bar{\mu}_{\mathbf{I}} + \gamma \tag{65}$$

Also in this case, the power of optimal tests based on polynomials of higher degree increases with the degree of the polynomial.

It is possible to give explicit expressions of the above matrix and moments for the optimal quadratic test. The formulae for the weights  $\Gamma_{\mathbf{I}}^{(n_1)}$  for the unfolded spectrum are:

$$\Gamma_i^{(1)} = (\Sigma_\mu + 2 - \Sigma_{\bar{\mu}}) \left( \frac{\bar{\mu}_i}{\mu_i} - 1 \right) - \frac{1}{2} \left( \frac{\bar{\mu}_i^2}{\mu_i^2} - 1 \right) \tag{66}$$

$$\Gamma_{ii}^{(2)} = \frac{1}{2} \left( \frac{\bar{\mu}_i}{\mu_i} - 1 \right)^2 \tag{67}$$

$$\Gamma_{ij}^{(2)} = \frac{1}{2} \left( \frac{\bar{\mu}_i}{\mu_i} - 1 \right) \left( \frac{\bar{\mu}_j}{\mu_j} - 1 \right) \tag{68}$$

$$\gamma = \frac{1}{2} (\Sigma_\mu - \Sigma_{\bar{\mu}}) (\Sigma_\mu + 2 - \Sigma_{\bar{\mu}}) \tag{69}$$

These results are valid in the independent sites approximation. They are also valid for the folded spectrum if the appropriate  $\mu_i$  and  $\bar{\mu}_i$  are used. An expression for the denominator of the test in the independent sites approximation can be found in Appendix A.7.

### 2.5. Empirical Simulations and Code Used for Studying the Statistical Power of Neutrality Tests

We obtained the frequency spectrum of samples for different subdivision and expansion models using one million iterations. Simulation parameters, as well as all the data frequencies from the simulations are included in Zenodo (DOI: 10.5281/zenodo.8279694). Linear and nonlinear tests were calculated using our own code, also available in the same link. Additionally, linear and nonlinear optimal tests are also included in the program *mstatspop* (<https://github.com/cragenomica/mstatspop>, accessed on 15 August 2023), to facilitate the calculation of these tests by the users.

## 3. Results

### 3.1. General Optimisation of Linear Tests

The condition for optimal tests is the maximisation of  $\mathcal{E}(T_\Omega)$  under the alternative scenario. However, a better approach would be the maximisation of the power of the test to reject the neutral model in the alternative scenario, given a choice of significance level  $\alpha$ . This approach requires the knowledge of the form of the probability distributions  $p(T_\Omega = t|H_0)$ ,  $p(T_\Omega = t|H_1)$  where  $H_0$  and  $H_1$  are the null and alternative model, or equivalently of all the moments of the spectrum  $E(\xi_i \xi_j \xi_k \dots)$  and  $\mathcal{E}(\xi_i \xi_j \xi_k \dots)$ .

Since this information is usually not available in analytic form and hard to obtain computationally, we limit to the case where the distribution can be well approximated by a Gaussian both for the null and for the expected model. Then, the only information needed are the spectra  $\mu_i = E(\xi_i)$ ,  $\bar{\mu}_i = \mathcal{E}(\xi_i)$  and their covariance matrices  $c_{ij} = E(\xi_i \xi_j) - E(\xi_i)E(\xi_j)$ ,  $\bar{c}_{ij} = \mathcal{E}(\xi_i \xi_j) - \mathcal{E}(\xi_i)\mathcal{E}(\xi_j)$ .

We expect that both in this approximation and in the general case, the tests with maximum power will depend on the significance level chosen, therefore limiting the interest of these test and the possibilities of comparison between results of the test on samples from different experiments.

We call  $\tau = \text{erf}^{-1}(1 - 2\alpha)$  the z-value corresponding to the critical p-value  $\alpha$ . In the Gaussian approximation, the power is given by the following expression:

$$\text{Power} = \frac{1}{2} \left[ 1 + \text{erf} \left( \frac{\sum_j \bar{\mu}_j \Omega_j - \tau \sqrt{\sum_{j,k} c_{jk} \Omega_j \Omega_k}}{\sqrt{\sum_{j,k} \bar{c}_{jk} \Omega_j \Omega_k}} \right) \right] \quad (70)$$

then its maximisation is equivalent to the maximisation of:

$$\frac{\sum_j \bar{\mu}_j \Omega_j - \tau \sqrt{\sum_{j,k} c_{jk} \Omega_j \Omega_k}}{\sqrt{\sum_{j,k} \bar{c}_{jk} \Omega_j \Omega_k}} \quad (71)$$

In the general case, the weights corresponding to the maximum depend explicitly on  $\tau$  and therefore on  $\alpha$ . This dependence is expected but unwanted, since the interpretation of the test depends explicitly on the critical p-value chosen.

There is only one case with weights independent on  $\tau$ , that is the case of  $\bar{c}_{ij}$  (approximately) proportional to  $c_{ij}$ . In this case the maximisation of the power of the test is (approximately) equivalent to the maximisation of the average result of the test, which is precisely the condition for optimal tests in the sense of [10]. In fact, in this case, optimal tests correspond precisely to an approximation of the likelihood-ratio tests under the assumption of Gaussian likelihood functions, and are therefore approximately the most powerful tests because of the Neyman–Pearson lemma.

As a side note, there is also a regime of values of  $\alpha$  such that the weights corresponding to maximum power are independent of  $\alpha$ , that is, the regime  $\tau(\alpha) \gg 1$ . In this case the power is an increasing function of  $\sum_{j,k} \bar{c}_{jk} \Omega_j \Omega_k / \sum_{l,m} c_{lm} \Omega_l \Omega_m$  and the weights are simply given by the null eigenvector (or linear combination of null eigenvectors) of the matrix  $\bar{c}_{ij} - \chi c_{ij}$ , where  $\chi$  is uniquely defined by the requirement that  $\bar{c}_{ij} - \chi c_{ij}$  be a negative

semidefinite matrix with at least a null eigenvalue. However, this regime is uninteresting because such small significance levels are practically useless (if  $\tau \sim 10$ , the corresponding critical  $p$ -value is  $\alpha \sim 10^{-20}$ ). However, it could be possible to build interesting tests with higher power by selecting linear combinations of the weights of the two  $\alpha$ -independent tests discussed above, that is, optimal tests and tests that maximise the alternative/null variance ratio  $\sum_{j,k} \bar{c}_{jk} \Omega_j \Omega_k / \sum_{l,m} c_{lm} \Omega_l \Omega_m$ .

### 3.1.1. Generalised Optimality Conditions

In a more general linear framework, the (normalised) test has the form  $T = \sum_{i=1}^{n-1} \Omega_i \xi_i / \xi_i^0$ . We denote its moments under the standard neutral model (SNM) and the alternative model by:

$$\mathcal{E}_0 = E(T)_{SNM} = \theta L \sum_{i=1}^{n-1} \Omega_i, \quad \mathcal{V}_0 = \text{Var}(T)_{SNM} = \sum_{i,j=1}^{n-1} \Omega_i \Omega_j c_{ij} / \xi_i^0 \xi_j^0 \tag{72}$$

$$\mathcal{E} = \mathcal{E}(T)_{alt} = \theta L \sum_{i=1}^{n-1} \Omega_i \bar{\xi}_i / \xi_i^0, \quad \mathcal{V} = \text{Var}(T)_{alt} = \sum_{i,j=1}^{n-1} \Omega_i \Omega_j \bar{c}_{ij} / \xi_i^0 \xi_j^0 \tag{73}$$

where  $E(\xi_i)_{SNM} = \theta L \xi_i^0$ ,  $\mathcal{E}(\xi_i)_{alt} = \theta L \bar{\xi}_i$ ,  $\text{Cov}(\xi_i, \xi_j)_{SNM} = c_{ij}$  and  $\text{Cov}(\xi_i, \xi_j)_{alt} = \bar{c}_{ij}$ . For the standard neutral model  $\xi_i^0 = 1/i$ .

The optimality condition can be the maximisation of a general function  $M(\mathcal{E}_0, \mathcal{V}_0, \mathcal{E}, \mathcal{V})$  of these quantities with respect to the weights  $\Omega_i$ . Since we are dealing with linear combination of the frequency spectrum with mean  $\mathcal{E}_0 = 0$  and variance  $\mathcal{V}_0 = 1$ , the function effectively depends on  $\mathcal{E}, \mathcal{V}$  only. Maximisation of power (in the Gaussian approximation), given a significance level  $\alpha$ , is equivalent to the maximisation of:

$$M(\mathcal{E}, \mathcal{V}) = \frac{\mathcal{E} - \Phi^{-1}(1 - \alpha)}{\sqrt{\mathcal{V}}} \tag{74}$$

If there is some knowledge of the variance of the alternative spectrum, or at least of the contribution for unlinked sites, maximisation of power is not the only possible optimisation. In other words, it is possible to optimise with respect to other functions  $M(\mathcal{E}, \mathcal{V})$  of the alternative mean  $\mathcal{E}$  and variance  $\mathcal{V}$  of the test, in order to obtain, e.g., more robust or conservative tests.

For example:

- “Optimal tests”: if  $\mathcal{V}$  is assumed to be completely unknown, the best we can do is minimise the  $p$ -value of the expected alternative spectrum, that is, in the Gaussian approximation  $p = 1 - \Phi((\mathcal{E} - \mathcal{E}_0) / \sqrt{\mathcal{V}_0})$  where  $\Phi(z)$  is the cumulative distribution function for the standard normal distribution. This is equivalent to the maximisation of:

$$M(\mathcal{E}_0, \mathcal{V}_0, \mathcal{E}, \mathcal{V}) = \frac{\mathcal{E} - \mathcal{E}_0}{\sqrt{\mathcal{V}_0}} \tag{75}$$

- Most powerful one-tail tests: in the Gaussian approximation, the power of the right tail of test to reject the neutral Wright–Fisher model given a significance level  $\alpha$  is  $\text{Power} = 1 - \Phi((\mathcal{E}_0 + \tau \sqrt{\mathcal{V}_0} - \mathcal{E}) / \sqrt{\mathcal{V}})$ , where  $\tau = \Phi^{-1}(1 - \alpha)$ . Maximising this power is equivalent to the maximisation of:

$$M(\mathcal{E}_0, \mathcal{V}_0, \mathcal{E}, \mathcal{V}) = \frac{\mathcal{E} - \mathcal{E}_0 - \tau \sqrt{\mathcal{V}_0}}{\sqrt{\mathcal{V}}} \tag{76}$$

If  $\mathcal{V} \propto \mathcal{V}_0$ , this is equivalent to the previous case and we retrieve the optimal tests of [10] that are therefore the most powerful tests in this approximation.

- Most powerful two-tails tests: in the Gaussian approximation, the power of both tails of the test to reject the neutral Wright–Fisher model is:

$$M(\mathcal{E}_0, \mathcal{V}_0, \mathcal{E}, \mathcal{V}) = 1 - \Phi\left(\frac{\mathcal{E}_0 + \tau' \sqrt{\mathcal{V}_0} - \mathcal{E}}{\sqrt{\mathcal{V}}}\right) + \Phi\left(\frac{\mathcal{E}_0 - \tau' \sqrt{\mathcal{V}_0} - \mathcal{E}}{\sqrt{\mathcal{V}}}\right) \tag{77}$$

where  $\tau' = \Phi^{-1}(1 - \alpha/2)$  and  $\alpha$  is the significance level. If  $\mathcal{E}_0 = 0$ , this maximisation is equivalent to the maximisation of  $(\mathcal{E} - \mathcal{E}_0 - \tau' \sqrt{\mathcal{V}_0}) / \sqrt{\mathcal{V}}$  similar to the previous case.

- Penalisations for high or low variance, depending on an extra parameter  $\nu$ , such as:

$$M(\mathcal{E}_0, \mathcal{V}_0, \mathcal{E}, \mathcal{V}) = \frac{\mathcal{E} - \mathcal{E}_0}{(\sqrt{\mathcal{V}_0} + \sqrt{\mathcal{V}})^\nu} \tag{78}$$

### 3.1.2. Tunable Optimal Tests

Interestingly, it turns out that the choice of an optimisation criterion—i.e., of a function  $M(\mathcal{E}, \mathcal{V})$  to be maximised—can be performed simply by tuning a parameter  $\lambda$  in a simple class of tests that we call “tunable optimal tests”.

The family of optimised tests with tunable parameter  $\lambda$  has the simple form:

$$T = \frac{\sum_{i=1}^{n-1} i \Omega_i(\lambda) \xi_i}{\sqrt{\text{Var}(\sum_{i=1}^{n-1} i \Omega_i(\lambda) \xi_i)}} \tag{79}$$

with weights:

$$\Omega_i(\lambda) = \frac{1}{i} \left[ 1 - \frac{a_n}{1 + \lambda i \bar{\xi}_i} \left( \sum_{j=1}^{n-1} \frac{1}{j(1 + \lambda j \bar{\xi}_j)} \right)^{-1} \right] \tag{80}$$

where  $a_n = \sum_{j=1}^{n-1} 1/j$ .

The interest of this family of tests lies in the following property: for every possible choice of the optimisation function  $M(\mathcal{E}, \mathcal{V})$ , the test corresponding to the maximum value of this function belongs to this family (see Appendix B).

Usual optimal tests [10] are obtained from the maximisation of  $M(\mathcal{E}, \mathcal{V}) = \mathcal{E}$  and correspond to  $\lambda = 0^+$ , that is, to the usual weights  $\Omega_i = \bar{\xi}_i - \frac{1}{i} \frac{\sum_{j=1}^{n-1} \bar{\xi}_j}{a_n}$ .

More generally,  $\lambda$  can be obtained directly by evaluating  $\mathcal{E}, \mathcal{V}$  with the weights (80) and then looking for the maximisation of  $M(\mathcal{E}(\lambda), \mathcal{V}(\lambda))$ . Note that there could be several local maxima.

This class of tests is simple but far more flexible than usual optimal tests. However, note that the exact choice of weights depends explicitly on  $\theta$  and  $\alpha$ . Furthermore, the smoothness of the choice of weights with respect to the evolutionary parameters is not assured. In principle, a slight change of evolutionary scenario could change the optimal weights abruptly.

### 3.2. Simulations of the Power of Optimal Tests

Since a theoretical evaluation of the power of optimal tests of different degree is not possible, we evaluate numerically the power of some of these tests in different scenarios. We consider the best possible case, that is, we assume that the precise value of  $\theta$  is known. Moreover we assume unlinked sites and  $\theta \ll 1$ . In this approximation, as shown in Appendix A.7, the moments  $E(\xi_i \xi_j \xi_k \dots)$  depend only on the first moments  $\mu_i = \theta L \bar{\xi}_i^0$  and similarly  $\mathcal{E}(\xi_i \xi_j \xi_k \dots)$  depend only on  $\bar{\mu}_i = \theta L \bar{\xi}_i$ , therefore optimal tests depend only on the alternative and null average spectra.

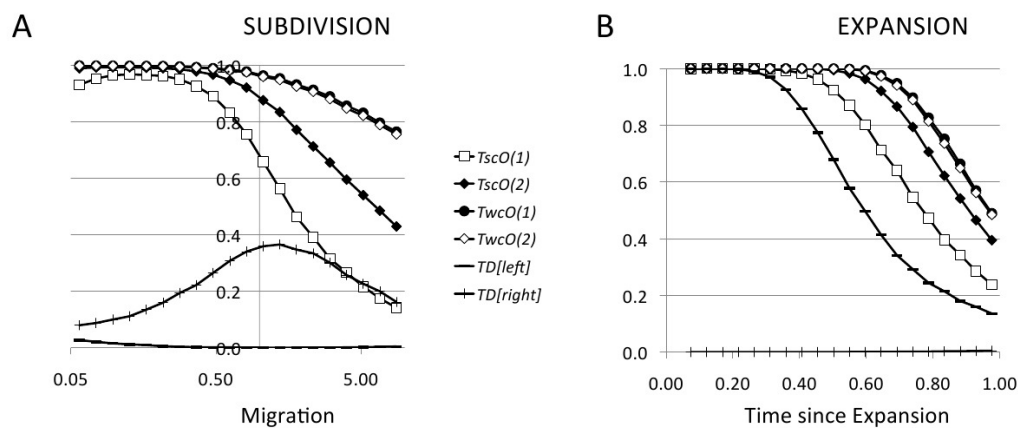
Note that for numerical simulations of optimal tests of higher degree, the numerical implementation can be made easier if all the occurrences of inverse covariance matrices  $C_{\mathbf{I}\mathbf{J}}^{-1}$  in the the above formulae are replaced with the corresponding second moments  $\mu_{\mathbf{I}\mathbf{J}}^{-1}$ ,

both in the expressions (50), (63), and in the definition (52). The test is the same because of the centredness condition, as it can be verified explicitly.

We compare four optimal tests. The first two are the linear and quadratic strongly centred optimal tests, which are denoted by  $T_{O(1)}^{sc}$  and  $T_{O(2)}^{sc}$  respectively. The third test is the weakly centred optimal test  $T_{O(1)}^{wc}$  based on a first order polynomial and presented in (59). The last optimal test  $T_{O(2)}^{wc}$  is also weakly centred and based on a quadratic polynomial. The explicit formulae for the computation of the weights of  $T_{O(2)}^{sc}$  and  $T_{O(2)}^{wc}$  were given in Equations (54)–(56) and (66)–(69).

We simulated two demographic processes: (A) subdivision, where two populations having identical size exchange individuals given a symmetric migration rate  $M$ , then individuals are sampled from one population only; (B) expansion, where the population size changes by a factor  $N_0/N = 10$  at a time  $T$  before present (in units of  $4N$  generations). For each value of the parameters  $M$  and  $T$ ,  $10^6$  simulations were performed with *mlcoalsim* v1.98b [20] for a region of 1000 bases with variability  $\theta = 0.05$  and recombination  $\rho = \infty$  and a sample size of  $n = 20$  (haploid) individuals. Confidence intervals at 95% level were estimated from  $10^6$  simulations of the standard neutral coalescent with the same parameters.

In Figure 3 we compare the power of the tests in the best possible situation, namely when  $\theta$  is known with good precision. In this condition all optimal tests should give the best results. In fact, the power of weakly centred tests ( $T_{O(1)}^{wc}$  and  $T_{O(2)}^{wc}$ ) is impressive, being around 100% for a large part of the parameter space and decreasing for large migration rates (Figure 3A) and long times (Figure 3B) as every other test, because the frequency spectrum for these cases becomes very similar to the standard spectrum. So, weakly centred tests show a very good theoretical performance, counterbalanced by their lack of robustness. The power of  $T_{O(1)}^{wc}$  and  $T_{O(2)}^{wc}$  are almost identical, therefore the contribution of the quadratic part to  $T_{O(2)}^{wc}$  is probably not relevant.



**Figure 3.** Statistical power of nonlinear optimal tests from coalescent simulations for the 5% tail, compared with Tajima's  $D$  test (for the left and the right tail). The parameters for the simulations are:  $n = 20$ ,  $\theta = 0.05$ ,  $L = 1000$  bp,  $\rho = \infty$ ; two populations considered but only one sampled (for panel A); expansion factor  $N_0/N = 10$  (for panel B).

On the other hand, strongly centred optimal tests are more powerful than Tajima's  $D$  but less powerful than weakly centred tests, as expected. However, there is a sensible difference in power between  $T_{O(1)}^{sc}$  and  $T_{O(2)}^{sc}$ : in the range of parameters where the power of weakly centred tests is around 100%, both strongly centred tests show a good performance not so far from the weakly centred ones, while in the less favourable range the quadratic test  $T_{O(2)}^{sc}$ , while performing worse than the weakly centred tests, has a power that is 20% higher than the linear test  $T_{O(1)}^{sc}$ . Taking into account the robustness of the tests, these simulations show that optimal tests like  $T_{O(2)}^{sc}$  could be an interesting alternative to the usual linear tests.

#### 4. Discussion

In this paper we have presented a systematic analysis of neutrality tests based on the site frequency spectrum. This study is intended to extend and complete the recent works in [8,10] by extending the study of the linear neutrality tests presented by Achaz. The properties of linear neutrality tests and optimal tests have been studied using this framework. A new class of “tunable” optimal tests that include usual optimal tests as a special case have been proposed, using a generalisation of the optimisation approach for linear tests.

The aim of the paper is to give mathematical guidelines to build new and more effective tests to detect deviations from neutrality. The proposed guidelines are the scaling relation (8) and the optimality condition based on the maximisation of  $\mathcal{E}(T_\Omega)$ . Both these guidelines are thoroughly explained and discussed.

One of the strengths of optimal neutrality tests (initially presented in [10]) is the fact that their weights scale automatically with the aim to avoid a dependence with the sample size, enabling interpretation of the results of the test in a sample-size-independent way. For other tests, different scaling strategies were analysed and evaluated, discussing the suitability of weighting methods and the scaling of existing neutrality tests.

The different neutral tests studied and developed in this paper have different features, which make them suitable for different purposes. For example, a general neutrality test where the weights are scaled to obtain interpretable results may be sufficient, with minimum effort, to reject the null model and to interpret results, but the statistical power when faced with a specific alternative hypothesis can be low. Instead, optimal tests become a better approach if the alternative hypothesis is clearly formulated and the data is not clearly from the null hypothesis. In respect to the differences between linear and nonlinear optimal test, while nonlinear optimal tests have been shown to be more powerful than linear ones (and weakly centred tests more powerful than strongly centred ones), power is not the only important issue: robustness must also be taken into consideration.

In fact, there are three important remarks on the relative robustness of these tests. The first one is that, as already discussed, centredness of weakly centred tests is not robust with respect to a biased estimate of  $\theta$ , therefore these tests should be preferred to strongly centred tests only in situations where the value of  $\theta$  is well known or a good estimate is available.

The second remark is that neither the weights nor the results of linear optimal tests do depend on the value of  $\theta$  and on the number of segregating sites  $S$ , while the weights of nonlinear optimal tests have an explicit dependence on  $\theta$  and their results depend not only on the spectrum but also on  $S$ ; therefore, the interpretation of the results of these tests is more complicated. However, this is not necessarily true for homogeneous tests of any degree, like the quadratic  $G_{\bar{z}}$  test by Fu. An interesting development of this work could be a study of homogeneous tests of a given degree  $k$  satisfying the optimality condition, which can be easily obtained from Equation (50) by restricting all ordered sequences of indices  $\bar{\mathbf{I}}, \bar{\mathbf{J}}, \bar{\mathbf{K}}, \bar{\mathbf{L}}$  to contain precisely  $k$  indices (along with some “traceless” condition, in case). These homogeneous optimal tests (or at least some subclass of them) should depend weakly on  $S$ .

The third remark is that linear optimal tests have two interesting properties that are not shared by other tests: they depend only on the deviations from the null spectrum and they have an easy interpretation in terms of these deviations, that is, they are positive if the observed deviations are similar to the expected ones and negative if the observed deviations are opposite to the expected ones. These features give an important advantage to linear optimal tests.

A fourth remark is that “tunable” optimal tests can be built to achieve weights that correspond to maximum power to reject the null hypothesis for a given alternative scenario but uncertain power calculations (e.g., when the variance of the alternative scenario is not known). This new class of neutrality tests is highly flexible and only depends on a single parameter and the mean alternative spectrum. Nevertheless, since the test can be used even

when the alternative scenario is not fully defined, there may be situations where optimising its power can result in the interpretability or robustness of the test being compromised.

Tests based on the frequency spectrum of polymorphic sites are fast, being based on simple matrix multiplications, and can therefore be applied to genome-wide data. Moreover, they can be used as summary statistics for Approximate Bayesian Computation or other statistical approaches to the analysis of sequence data. While linear tests are often used in this way, the nonlinear tests presented in this paper contain more information (related to the covariances and higher moments of the frequency spectrum) that could increase the power of these analyses.

**Author Contributions:** Conceptualization, S.E.R.-O., G.A. and L.F.; Methodology, S.E.R.-O. and L.F.; Software, S.E.R.-O.; Formal analysis, G.M. and L.F.; Writing—original draft, S.E.R.-O., G.M. and L.F.; Writing—review & editing, S.E.R.-O., G.M., G.A. and L.F.; Funding acquisition, S.E.R.-O. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was funded by grant CGL2009-0934 and PID2020-119255GB-I00 (MICINN, Spain) to S.R.O.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** No data was used in this paper.

**Acknowledgments:** We thank M. Perez-Enciso and J. Rozas for useful comments.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A. Technical Results

### Appendix A.1. Transformations of Weights and Invariance of Tests

We report some theorems on the invariance of the tests under affine transformations. These results can be easily proved and are implicitly used throughout this paper.

**Theorem A1.** *A test of the form (3) does not change its value if all the weights  $\Omega_i$  are rescaled by a common factor  $\lambda > 0$ , that is:*

$$\Omega_i \longrightarrow \lambda \Omega_i \quad \Rightarrow \quad T_\Omega \longrightarrow \text{sign}(\lambda) T_\Omega \quad (\text{A1})$$

Note that the invariance of the tests mean that these transformations define equivalence classes of weights, i.e., sets of different weights that actually correspond to the same test. In particular, this theorem implies that the space of possible tests, in terms of the weights  $\Omega_i$ , is not homeomorphic to  $\mathbb{R}^{n-2}$  (which would be the subspace of weights in  $\mathbb{R}^{n-1}$  that satisfy the linear condition (4)) but to its quotient with respect to the invariance (multiplicative) group  $\mathbb{R}^+$ , that is, the  $(n-3)$ -dimensional sphere  $S^{n-3} = \mathbb{R}^{n-2} / \mathbb{R}^+$ .

**Theorem A2.** *A test of the form (7) does not change its value under an affine transformation of parameters  $(\lambda, \rho_i)$  on the weights  $\omega_i, \omega'_i$  with a common rescaling factor  $\lambda > 0$ , that is:*

$$\omega_i \longrightarrow \lambda \omega_i + \rho_i, \quad \omega'_i \longrightarrow \lambda \omega'_i + \rho_i \quad \Rightarrow \quad T_\Omega \longrightarrow \text{sign}(\lambda) T_\Omega \quad (\text{A2})$$

However, the estimators (5) are unbiased only if the rescaling factor satisfies the condition  $\lambda = 1 - \sum_{i=1}^{n-1} \rho_i$ .

### Appendix A.2. Generalised $D'$ Tests for Multilocus Analysis

The statistic  $D'$  [21], which is defined as the ratio of Tajima's  $D$  versus its minimum value (given a fixed number of segregating sites), has been used in the literature for multilocus analyses [21–23], arguing that the value of Tajima's  $D$  is affected by the length,

the sample size, and the number of segregating sites of each studied locus and therefore the values of each locus are not directly comparable.

The contribution of each locus to the heterogeneity is hardly known. Tajima’s  $D$  is robust to differences in the level of variability (the variance is approximately equal to one) and also quite robust against differences in sample size (as will be shown in the next part), although the quantitative values of Tajima’s  $D$  for each condition are somewhat different and therefore the comparison between values is not simple. The proposal of Schaeffer is to use the test:

$$D' = \frac{D}{\min(D)_{S=S_{obs}}} \tag{A3}$$

as a (re)normalised version of Tajima’s  $D$ .  $S_{obs}$  is the observed number of segregating sites in the sample. This proposal can be generalised for all the tests of the form (3) as follows:

$$T'_\Omega = \frac{T_\Omega}{\min(T_\Omega)_{S=S_{obs}}} = \frac{\sum_{i=1}^{n-1} i\Omega_i\zeta_i}{\min(j\Omega_j)S_{obs}} \tag{A4}$$

This appears to be the natural generalisation of  $D'$  to general linear tests. For the optimal tests, there are other possibilities.

*Appendix A.3. Tests Based on the Folded SFS*

If an outgroup is not available, then the test should be based on the folded spectrum and has the general form:

$$\begin{aligned} T_\Omega^* &= \frac{\hat{\theta}_\omega^* - \hat{\theta}_{\omega'}^*}{\sqrt{\text{Var}(\hat{\theta}_\omega^* - \hat{\theta}_{\omega'}^*)}} = \frac{\sum_{i=1}^{\lfloor n/2 \rfloor} i(n-i)(1 + \delta_{n,2i})(\omega_i^* - \omega_i^{*'})\eta_i}{\sqrt{\text{Var}\left(\sum_{j=1}^{\lfloor n/2 \rfloor} j(n-j)(1 + \delta_{n,2j})(\omega_j^* - \omega_j^{*'})\eta_j\right)}} = \\ &= \frac{\sum_{i=1}^{\lfloor n/2 \rfloor} i(n-i)(1 + \delta_{n,2i})\Omega_i^*\eta_i}{\sqrt{\text{Var}\left(\sum_{j=1}^{\lfloor n/2 \rfloor} j(n-j)(1 + \delta_{n,2j})\Omega_j^*\eta_j\right)}} \end{aligned} \tag{A5}$$

where the weights  $\Omega_i^* = \omega_i^* - \omega_i^{*'}$  satisfy the conditions:

$$\sum_{i=1}^{\lfloor n/2 \rfloor} \omega_i^* = 1 \quad , \quad \sum_{i=1}^{\lfloor n/2 \rfloor} \omega_i^{*' } = 1 \quad \Rightarrow \quad \sum_{i=1}^{\lfloor n/2 \rfloor} \Omega_i^* = 0 \tag{A6}$$

and:

$$\hat{\theta}_\omega^* = \frac{1}{L} \sum_{i=1}^{\lfloor n/2 \rfloor} \frac{i(n-i)(1 + \delta_{n,2i})}{n} \omega_i^* \eta_i \quad , \quad \hat{\theta}_{\omega'}^* = \frac{1}{L} \sum_{i=1}^{\lfloor n/2 \rfloor} \frac{i(n-i)(1 + \delta_{n,2i})}{n} \omega_i^{*' } \eta_i \tag{A7}$$

are unbiased estimators of  $\theta$ .

*Appendix A.4. Scaling of Weights in Tests without an Outgroup*

The above arguments can be repeated in a straightforward way for the tests  $T_\Omega^*$  based on the folded spectrum  $\eta_i$ . The only relevant difference is that the frequency  $f$  of the minor allele in the population is always less than 50%, that is,  $f \in (0, 1/2]$ . For consistency with the unfolded case, the weight  $\eta_{n/2}$  is reduced by a factor 2. Moreover, the additional parameters related to the weights of singletons cannot distinguish between ancestral and derived alleles and therefore reduce to  $\Omega_s^*, \omega_s^*, \omega_s^{*'}$ . These parameters, together with the functions  $\Omega^*(f), \omega^*(f),$  and  $\omega^{*' } (f)$ , should satisfy the conditions:

$$\Omega_s^* + \int_0^{1/2} df \Omega^*(f) = 0 \quad , \quad \omega_s^* + \int_0^{1/2} df \omega^*(f) = \omega_s^{*' } + \int_0^{1/2} df \omega^{*' } (f) = 1 \tag{A8}$$



The formulae that determine the scaling of the weights are:

$$\Omega_i^{*(n)} = \frac{1}{1 + \delta_{n,2i}} \Omega^* \left( \frac{i}{n} \right) + \Omega_s^* \delta_{i,1} - \frac{1}{\lfloor n/2 \rfloor} \left( \Omega_s^* + \sum_{j=1}^{\lfloor n/2 \rfloor} \frac{1}{1 + \delta_{n,2j}} \Omega^* \left( \frac{j}{n} \right) \right) \tag{A9}$$

$$\Omega_i^{*(n)} = \frac{\omega_s^* \delta_{i,1} + \omega^* \left( \frac{i}{n} \right) / (1 + \delta_{n,2i})}{\omega_s^* + \sum_{j=1}^{\lfloor n/2 \rfloor} \omega^* \left( \frac{j}{n} \right) / (1 + \delta_{n,2j})} - \frac{\omega_s^{*'} \delta_{i,1} + \omega^{*'} \left( \frac{i}{n} \right) / (1 + \delta_{n,2i})}{\omega_s^{*'} + \sum_{j=1}^{\lfloor n/2 \rfloor} \omega^{*'} \left( \frac{j}{n} \right) / (1 + \delta_{n,2j})} \tag{A10}$$

The weights of the folded versions of Tajima’s  $D$  and Fu and Li’s  $F^*$  test follow this scaling. The nonzero weight functions are  $\omega^*(f) = 1$ ,  $\omega^{*'}(f) = 1/(\log(N)f(1 - f))$  for Tajima’s  $D$  and  $\omega_s^* = 1$ ,  $\omega_s^{*'}(f) = 1/(\log(N)f(1 - f))$  for the test of Fu and Li.

Appendix A.5. Scaling of Optimal Tests

Optimal tests have weights proportional to the expected allele distribution  $\omega_i = \bar{\xi}_i / \sum_{j=1}^{n-1} \bar{\xi}_j$  and to the null allele distribution  $\omega'_i = \zeta_i^0 / \sum_{j=1}^{n-1} \zeta_j^0$ . Therefore, the weights of an optimal test follow the same scaling with sample size as the allele distributions. Denoting by  $\bar{\zeta}(f)$  and  $\zeta^0(f)$  the spectra of expected and null allele frequencies in the whole population, the spectra for the sample are obtained by binomial sampling:

$$\xi_i = \int_{1/N_e}^1 df P_{bin}(i; n, f) \zeta(f) \quad P_{bin}(i; n, f) = \binom{n}{i} f^i (1 - f)^{n-i} \tag{A11}$$

from the spectra  $\zeta(f) = \bar{\zeta}(f)$  and  $\zeta(f) = \zeta^0(f)$ , respectively. Therefore the scaling of the allele distributions is:

$$\frac{\bar{\xi}_i}{\sum_{j=1}^{n-1} \bar{\xi}_j} = \frac{\int_{1/N_e}^1 df \binom{n}{i} f^i (1 - f)^{n-i} \bar{\zeta}(f)}{\int_{1/N_e}^1 df (1 - f^n - (1 - f)^n) \bar{\zeta}(f)} \tag{A12}$$

$$\frac{\zeta_i^0}{\sum_{j=1}^{n-1} \zeta_j^0} = \frac{\int_{1/N_e}^1 df \binom{n}{i} f^i (1 - f)^{n-i} \zeta^0(f)}{\int_{1/N_e}^1 df (1 - f^n - (1 - f)^n) \zeta^0(f)} \tag{A13}$$

which is precisely the scaling (26) and (27) with weight functions  $\omega(f) \propto \bar{\zeta}(f)$  and  $\omega'(f) \propto \zeta^0(f)$ . This scaling does not correspond to the scaling (8) suggested in this paper, but it takes into account the sampling process in a straightforward way, being based on the expected and null allele distributions for the sample and therefore immediately related to the binomial sampling of alleles from the population.

Note that these weights actually follow the scaling (8) for large  $n$ , with the same weight functions  $\omega(f) \propto \bar{\zeta}(f)$  and  $\omega'(f) \propto \zeta^0(f)$ . This agrees with the fact that when the sample size is large enough, the variance of the sampling process can be safely ignored and all reasonable choices of scaling are equivalent to our proposal (8).

Appendix A.6. Generalising  $D'$  for Optimal Test

The  $D'$  statistics of [21] can be generalised for optimal tests as it was done for general linear tests. In particular, for a fixed number of segregating sites  $S_{obs}$ , the generalisation for optimal tests in the approximation of unlinked sites and  $\theta \ll 1$  is:

$$T'_O = \frac{\sum_{i=1}^{n-1} \Omega_i \bar{\xi}_i / \zeta_i^0}{\min_j (\Omega_j / \zeta_j^0) S_{obs}} = \frac{\sum_{i=1}^{n-1} (\bar{\xi}_i / S_{obs}) (\bar{\xi}_i / \sum_{j=1}^{n-1} \bar{\xi}_j) / (\zeta_i^0 / \sum_{j=1}^{n-1} \zeta_j^0) - 1}{\min_k \left( (\bar{\xi}_k / \sum_{l=1}^{n-1} \bar{\xi}_l) / (\zeta_k^0 / \sum_{l=1}^{n-1} \zeta_l^0) \right) - 1} \tag{A14}$$

which has the interesting property of depending only on the allele frequency distributions.

However, in the case of optimal tests there is another possibility, namely to define a  $\bar{D}'$  test as the ratio of the optimal test and of its average minimum, assuming that the spectrum corresponds to the average spectrum of the actual scenario. This would correspond to the form:

$$\bar{T}'_O = - \frac{\langle\langle \zeta, \bar{\zeta} \rangle\rangle - \langle\langle \zeta, \zeta^0 \rangle\rangle \langle\langle \zeta^0, \bar{\zeta} \rangle\rangle / \langle\langle \zeta^0, \zeta^0 \rangle\rangle}{S_{obs} / a_n \sqrt{(\langle\langle \bar{\zeta}, \bar{\zeta} \rangle\rangle - \langle\langle \zeta^0, \bar{\zeta} \rangle\rangle^2 / \langle\langle \zeta^0, \zeta^0 \rangle\rangle)(\langle\langle \zeta, \zeta \rangle\rangle - \langle\langle \zeta^0, \zeta \rangle\rangle^2 / \langle\langle \zeta^0, \zeta^0 \rangle\rangle)}} \tag{A15}$$

which has the interesting property of being symmetric with respect to the actual spectrum  $\zeta$  and the expected spectrum  $\bar{\zeta}$ .

*Appendix A.7. Moments of the Frequency Spectrum in the Independent Sites Approximation*

We consider the limit  $\theta \ll 1, L \rightarrow \infty$  with constant  $\theta L$ . The spectrum  $\zeta_i$  can be written as a sum of spectra for all sites:

$$\zeta_i = \sum_{s=1}^L \zeta_i(s) \tag{A16}$$

where each variable  $\zeta_i(s)$  has a Bernoulli distribution  $\zeta_i(s) \in \{0, 1\}$  with probabilities  $p(1) = \theta \zeta_i^0$  and  $p(0) = 1 - \theta \zeta_i^0$  where  $\zeta_i^0 = 1/i$  under the standard neutral model. The expectation value of  $\zeta_i$  is therefore:

$$E(\zeta_i) = \mu_i = \sum_{s=1}^L E(\zeta_i(s)) = \theta L \zeta_i^0 = \frac{\theta L}{i} \tag{A17}$$

and similarly  $\mathcal{E}(\zeta_i) = \bar{\mu}_i = \theta L \bar{\zeta}_i$  for a general model with average spectrum  $\bar{\zeta}_i$ .

In the independent sites approximation, which is equivalent to the infinite recombination limit, the variables  $\zeta_i(s)$  and  $\zeta_i(s')$  are i.i.d. random variables for  $s \neq s'$ , and more generally the random variables  $\zeta_i(s)$  and  $\zeta_j(s')$  are independent for  $s \neq s'$ . The moments for a single site  $s$  can be calculated as:

$$\begin{aligned} E(\zeta_i(s)\zeta_j(s)\zeta_k(s) \dots) &= \sum_{a,b,c,\dots \in \{0,1\}} abc \dots P(\zeta_i(s) = a, \zeta_j(s) = b, \zeta_k(s) = c, \dots) = \\ &= P(\zeta_i(s) = 1, \zeta_j(s) = 1, \zeta_k(s) = 1, \dots) = P(\zeta_i(s) = 1) \delta_{ij} \delta_{jk} \dots \end{aligned} \tag{A18}$$

because the sum  $\sum_{i=1}^{n-1} \zeta_i(s) \in \{0, 1\}$ , that is, different allele frequencies for the same site are mutually exclusive. Therefore all these moments are always linear in  $\theta$  and are nonzero only when all indices are equal:

$$E(\zeta_i(s)\zeta_j(s)\zeta_k(s) \dots) = \theta \zeta_i^0 \delta_{ij} \delta_{jk} \dots = \frac{\theta}{i} \delta_{ij} \delta_{jk} \dots \tag{A19}$$

Then the second moment can be evaluated as:

$$\begin{aligned} E(\zeta_i \zeta_j) &= \sum_{s,s'} E(\zeta_i(s)\zeta_j(s')) = \sum_s E(\zeta_i(s)\zeta_j(s)) + \sum_{s \neq s'} E(\zeta_i(s))E(\zeta_j(s')) = \\ &= L \delta_{ij} \frac{\theta}{i} + L(L-1) \frac{\theta^2}{ij} = \delta_{ij} \mu_i + \mu_i \mu_j - \frac{\mu_i \mu_j}{L} \end{aligned} \tag{A20}$$

and neglecting subleading orders in  $\theta$  or  $L^{-1}$  like the last term above, we can calculate the third and fourth moments that are needed for the calculation of  $C^{-1}$ .

The final result for the moments of the spectrum  $\zeta_i$  is:

$$E(\zeta_i \zeta_j) = \mu_{ij} = \delta_{ij} \mu_i + \mu_i \mu_j \tag{A21}$$

$$E(\zeta_i \zeta_j \zeta_k) = \mu_{ijk} = \delta_{ij} \delta_{jk} \mu_i + (\delta_{ik} \mu_i \mu_j + \delta_{jk} \mu_i \mu_j + \delta_{ij} \mu_i \mu_k) + \mu_i \mu_j \mu_k \tag{A22}$$

$$E(\zeta_i \zeta_j \zeta_k \zeta_l) = \mu_{ijkl} = \delta_{ij} \delta_{jk} \delta_{kl} \mu_i + (\delta_{ik} \delta_{jl} \mu_i \mu_j + \delta_{il} \delta_{jk} \mu_i \mu_j + \delta_{ij} \delta_{kl} \mu_i \mu_k) + (\delta_{ij} \delta_{jk} \mu_i \mu_l + \delta_{ij} \delta_{jl} \mu_i \mu_k + \delta_{ik} \delta_{kl} \mu_i \mu_j + \delta_{jk} \delta_{kl} \mu_i \mu_j) + (\delta_{il} \mu_i \mu_j \mu_k + \delta_{jl} \mu_i \mu_j \mu_k + \delta_{ik} \mu_i \mu_j \mu_l + \delta_{jk} \mu_i \mu_j \mu_l) + \delta_{ij} \mu_i \mu_k \mu_l + \delta_{kl} \mu_i \mu_j \mu_k + \mu_i \mu_j \mu_k \mu_l \tag{A23}$$

All the results above can be applied to a general model simply by substituting  $\mu_i$  with  $\bar{\mu}_i$ . Moreover, they can be applied to the folded spectrum by taking  $\mu_i = \theta L n / i(n - i)(1 + \delta_{n,2i})$  for the standard neutral model or  $\bar{\mu}_i = \theta L (\bar{\zeta}_i + \bar{\zeta}_{n-i}) / (1 + \delta_{n,2i})$  for general models.

We define some quantities in order to simplify the expressions for the weights:

$$\Sigma_\mu = \sum_{i=1}^{n-1} \mu_i \quad , \quad \Sigma_{\bar{\mu}} = \sum_{i=1}^{n-1} \bar{\mu}_i \quad , \quad \Sigma_q = \sum_{i=1}^{n-1} \frac{\bar{\mu}_i^2}{\mu_i} \tag{A24}$$

If the spectrum is folded, all the sums in the above expressions run from 1 to  $\lfloor n/2 \rfloor$ .

The covariance matrix  $C_{\bar{\mathbf{I}}, \bar{\mathbf{J}}}$  is:

$$C_{i,j} = \mu_{ij} - \mu_i \mu_j = \delta_{ij} \mu_i \tag{A25}$$

$$C_{ij,k} = \mu_{ijk} - \mu_{ij} \mu_k = \delta_{ij} \delta_{jk} \mu_i + (\delta_{ik} \mu_i \mu_j + \delta_{jk} \mu_i \mu_j) \tag{A26}$$

$$C_{ij,kl} = \mu_{ijkl} - \mu_{ij} \mu_{kl} = \delta_{ij} \delta_{jk} \delta_{kl} \mu_i + (\delta_{ik} \delta_{jl} \mu_i \mu_j + \delta_{il} \delta_{jk} \mu_i \mu_j) + (\delta_{ij} \delta_{jk} \mu_i \mu_l + \delta_{ij} \delta_{jl} \mu_i \mu_k + \delta_{ik} \delta_{kl} \mu_i \mu_j + \delta_{jk} \delta_{kl} \mu_i \mu_j) + (\delta_{il} \mu_i \mu_j \mu_k + \delta_{jl} \mu_i \mu_j \mu_k + \delta_{ik} \mu_i \mu_j \mu_l + \delta_{jk} \mu_i \mu_j \mu_l) \tag{A27}$$

with the elements  $C_{ij,k}$  and  $C_{ij,kl}$  that should be considered only for  $i \leq j, k \leq l$ . It can be verified that the inverse matrix  $C_{\bar{\mathbf{I}}, \bar{\mathbf{J}}}^{-1}$  has the form:

$$C_{i,j}^{-1} = 1 + \delta_{ij} \frac{2\mu_i (\Sigma_\mu + 3) + 1}{2\mu_i^2} \tag{A28}$$

$$C_{ij,k}^{-1} = \delta_{ij} \delta_{jk} \frac{2\mu_i - 1}{2\mu_i^2} - (\delta_{ik} + \delta_{jk}) \frac{1}{\mu_k} \tag{A29}$$

$$C_{ij,kl}^{-1} = (\delta_{ik} \delta_{jl} + \delta_{il} \delta_{jk}) \frac{1}{\mu_i \mu_j} - \frac{3}{2} \delta_{ij} \delta_{ik} \delta_{jl} \frac{1}{\mu_i^2} \tag{A30}$$

The formulae for the weakly centred quadratic test (66)–(69) can be obtained from these formulae and the definition (63). The corresponding variance in the denominator (45) is:

$$\text{Var} \left( \sum_{\mathbf{I}} \Gamma_{\mathbf{I}}^{(n_{\mathbf{I}})} (\zeta \dots \zeta)_{\mathbf{I}} \right) = 2(\Sigma_{\bar{\mu}} - \Sigma_q / 2)^2 + \Sigma_\mu (\Sigma_\mu / 2 + 1 - 2\Sigma_{\bar{\mu}} + \Sigma_q) + \Sigma_q - 2\Sigma_{\bar{\mu}} \tag{A31}$$

The matrix  $\mathcal{M}$  is the inverse of the matrix  $\mathcal{M}_{rl}^{-1} = \sum_{\bar{\mathbf{I}}, \bar{\mathbf{L}}} \mu_{\bar{\mathbf{I}}}^{(r)} C_{\bar{\mathbf{I}}\bar{\mathbf{L}}}^{-1} \mu_{\bar{\mathbf{L}}}^{(l)}$ , which can be easily calculated from the above equations as:

$$\mathcal{M}_{11}^{-1} = 2\Sigma_{\mu}^2 + \Sigma_{\mu} \tag{A32}$$

$$\mathcal{M}_{12}^{-1} = -\Sigma_{\mu}^2 \tag{A33}$$

$$\mathcal{M}_{22}^{-1} = \Sigma_{\mu}^2 / 2 \tag{A34}$$

and the matrix  $\mathcal{M}$  is

$$\mathcal{M} = \begin{pmatrix} 1/\Sigma_{\mu} & 2/\Sigma_{\mu} \\ 2/\Sigma_{\mu} & 4/\Sigma_{\mu} + 2/\Sigma_{\mu}^2 \end{pmatrix} \tag{A35}$$

The formulae (54)–(56) can be obtained from the formula (A35) and from the following results:

$$\sum_{\bar{\mathbf{I}}} C_{i\bar{\mathbf{I}}}^{-1} \bar{\mu}_{\bar{\mathbf{I}}} = \Sigma_{\bar{\mu}} + (\Sigma_{\mu} + 2 - \Sigma_{\bar{\mu}}) \frac{\bar{\mu}_i}{\mu_i} - \frac{1}{2} \left( \frac{\bar{\mu}_i}{\mu_i} \right)^2 \tag{A36}$$

$$\sum_{\bar{\mathbf{I}}} C_{ii\bar{\mathbf{I}}}^{-1} \bar{\mu}_{\bar{\mathbf{I}}} = -\frac{\bar{\mu}_i}{\mu_i} + \frac{1}{2} \left( \frac{\bar{\mu}_i}{\mu_i} \right)^2 \tag{A37}$$

$$\sum_{\bar{\mathbf{I}}} C_{ij\bar{\mathbf{I}}}^{-1} \bar{\mu}_{\bar{\mathbf{I}}} = -\frac{\bar{\mu}_i}{\mu_i} - \frac{\bar{\mu}_j}{\mu_j} + \frac{\bar{\mu}_i \bar{\mu}_j}{\mu_i \mu_j} \tag{A38}$$

$$\sum_{\bar{\mathbf{I}}, \bar{\mathbf{J}}} \mu_{\bar{\mathbf{I}}}^{(1)} C_{\bar{\mathbf{I}}\bar{\mathbf{J}}}^{-1} \bar{\mu}_{\bar{\mathbf{J}}} = \Sigma_{\bar{\mu}} (2\Sigma_{\mu} - \Sigma_{\bar{\mu}} + 1) \qquad \sum_{\bar{\mathbf{I}}, \bar{\mathbf{J}}} \mu_{\bar{\mathbf{I}}}^{(2)} C_{\bar{\mathbf{I}}\bar{\mathbf{J}}}^{-1} \bar{\mu}_{\bar{\mathbf{J}}} = \Sigma_{\bar{\mu}}^2 / 2 - \Sigma_{\mu} \Sigma_{\bar{\mu}} \tag{A39}$$

$$\sum_{\bar{\mathbf{I}}} C_{i\bar{\mathbf{I}}}^{-1} \mu_{\bar{\mathbf{I}}}^{(1)} = 2\Sigma_{\mu} + 2 \qquad \sum_{\bar{\mathbf{I}}} C_{i\bar{\mathbf{I}}}^{-1} \mu_{\bar{\mathbf{I}}}^{(2)} = -\Sigma_{\mu} - 1/2 \tag{A40}$$

$$\sum_{\bar{\mathbf{I}}} C_{ii\bar{\mathbf{I}}}^{-1} \mu_{\bar{\mathbf{I}}}^{(1)} = -1 \qquad \sum_{\bar{\mathbf{I}}} C_{ii\bar{\mathbf{I}}}^{-1} \mu_{\bar{\mathbf{I}}}^{(2)} = 1/2 \tag{A41}$$

$$\sum_{\bar{\mathbf{I}}} C_{ij\bar{\mathbf{I}}}^{-1} \mu_{\bar{\mathbf{I}}}^{(1)} = -2 \qquad \sum_{\bar{\mathbf{I}}} C_{ij\bar{\mathbf{I}}}^{-1} \mu_{\bar{\mathbf{I}}}^{(2)} = 1 \tag{A42}$$

### Appendix B. Proofs

#### Derivation of General Optimisation for Tunable Linear Tests

We consider the maximisation of  $M(\mathcal{E}, \mathcal{V})$  with the conditions  $\mathcal{E}_0 = 0, \mathcal{V}_0 = 1$ . We assume that the function  $M(\mathcal{E}, \mathcal{V})$  is monotonic in both variables.

The existence of an absolute maximum (and minimum) is ensured by Weierstrass theorem applied to the closed manifold determined by the conditions  $\mathcal{E}_0 = 0, \mathcal{V}_0 = 1$  [10].

Local maxima can be found by the Lagrange multipliers method, i.e., by the extremum conditions for  $M(\mathcal{E}, \mathcal{V}) - \lambda_1 \mathcal{E}_0 - \lambda_2 (\mathcal{V}_0 - 1)$  with respect to  $\Omega_i, \lambda_1$  and  $\lambda_2$ . We obtain the extremum equation:

$$\Omega_i = -\frac{\theta L}{2} \frac{\partial M(\mathcal{E}, \mathcal{V}) / \partial \mathcal{E}}{\partial M(\mathcal{E}, \mathcal{V}) / \partial \mathcal{V}} \cdot \zeta_i^0 \sum_{j=1}^{n-1} (\bar{c} + \tilde{\lambda}_2 c)_{ij}^{-1} (\bar{\xi}_j - \tilde{\lambda}_1 \zeta_j^0) \tag{A43}$$

with  $\tilde{\lambda}_1, \tilde{\lambda}_2$  determined by the constraint equations. Solving for  $\tilde{\lambda}_1$ , we obtain:

$$\Omega_i = -\frac{\theta L}{2} \frac{\partial M(\mathcal{E}, \mathcal{V}) / \partial \mathcal{E}}{\partial M(\mathcal{E}, \mathcal{V}) / \partial \mathcal{V}} \cdot \zeta_i^0 \sum_{j=1}^{n-1} (\bar{c} + \tilde{\lambda}_2 c)_{ij}^{-1} \left[ \bar{\xi}_j - \zeta_j^0 \frac{\sum_{l,m=1}^{n-1} \zeta_l^0 (\bar{c} + \tilde{\lambda}_2 c)_{lm}^{-1} \bar{\xi}_m}{\sum_{l,m=1}^{n-1} \zeta_l^0 (\bar{c} + \tilde{\lambda}_2 c)_{lm}^{-1} \zeta_m^0} \right] \tag{A44}$$

with  $\tilde{\lambda}_2$  determined by the condition  $\sum_{i,j=1}^{n-1} \Omega_i \Omega_j c_{ij} / \zeta_i^0 \zeta_j^0 = 1$ .

We approximate the variances with their values for unlinked sites:

$$c_{ij} \simeq \delta_{ij}\theta L\bar{\zeta}_i^0 \quad , \quad \bar{c}_{ij} \simeq \delta_{ij}\theta L\bar{\zeta}_i \tag{A45}$$

and we obtain the weights:

$$\Omega_i = \bar{\zeta}_i^0 \left( 1 - \frac{\sum_{j=1}^{n-1} \bar{\zeta}_j^0}{(\lambda \bar{\zeta}_i / \bar{\zeta}_i^0 + 1) \sum_{j=1}^{n-1} \frac{\bar{\zeta}_j^0}{\lambda \bar{\zeta}_j / \bar{\zeta}_j^0 + 1}} \right) \tag{A46}$$

where the remaining parameter  $\lambda$  is fixed by the condition:

$$\frac{\theta^3 L^3}{4} \sum_{i=1}^{n-1} \Omega_i^2 / \bar{\zeta}_i^0 = \left( \frac{\partial M(\mathcal{E}, \mathcal{V}) / \partial \mathcal{V}}{\partial M(\mathcal{E}, \mathcal{V}) / \partial \mathcal{E}} \right)^2 \tag{A47}$$

This equation fixes all values of  $\lambda$  corresponding to the local maxima. To obtain the absolute maximum, one has to choose the solution  $\lambda$  that actually maximises  $M(\mathcal{E}, \mathcal{V})$ .

In practice, for the standard neutral model, it is sufficient to use the usual form of the test:

$$T = \frac{\sum_{i=1}^{n-1} i \Omega_i \bar{\zeta}_i}{\sqrt{\text{Var}(\sum_{i=1}^{n-1} i \Omega_i \bar{\zeta}_i)}} \tag{A48}$$

with the (unnormalised) weights:

$$\Omega_i = \frac{1}{i} \left( 1 - \frac{a_n}{(\lambda i \bar{\zeta}_i + 1) \sum_{j=1}^{n-1} \frac{1}{j(\lambda \bar{\zeta}_j + 1)}} \right) \tag{A49}$$

with  $\lambda$  determined numerically either by the condition:

$$\frac{\theta^3 L^3}{4} \sum_{i=1}^{n-1} i \Omega_i^2 = \left( \frac{\partial M(\mathcal{E}, \mathcal{V}) / \partial \mathcal{V}}{\partial M(\mathcal{E}, \mathcal{V}) / \partial \mathcal{E}} \right)^2 \tag{A50}$$

or more easily, looking directly for the value of  $\lambda$  that maximises numerically  $M(\mathcal{E}, \mathcal{V})$ .

We discuss explicitly two important limit cases:

- $\partial M / \partial \mathcal{V} = 0$ : this case reduces to the usual optimal tests. It corresponds to the limit  $\lambda \rightarrow 0^+$  in the above equation, which gives a unique solution:

$$\Omega_i = \lambda \left( \bar{\zeta}_i - \frac{1}{i} \frac{\sum_{j=1}^{n-1} \bar{\zeta}_j}{a_n} \right) \tag{A51}$$

where the  $\lambda$  factor is irrelevant.

- $\partial M / \partial \mathcal{E} = 0$ : maximisation or minimisation of variance. In this case the values of  $\lambda$  are determined by the condition  $\sum_{j=1}^{n-1} \frac{1}{j(\lambda \bar{\zeta}_j + 1)} = 0$ , which is equivalent to  $a_n = \sum_{j=1}^{n-1} \frac{1}{j+1/\lambda \bar{\zeta}_j}$ . The r.h.s. is monotonically increasing and diverges for the  $n - 1$  values  $\lambda = -1/i\bar{\zeta}_i, i = 1 \dots n - 1$ . Monotonicity and the Weierstrass theorem ensure that there are precisely  $n - 2$  solutions for this equation. If we order the values of  $i\bar{\zeta}_i$  from the lowest  $i_1\bar{\zeta}_{i_1}$  to the highest  $i_{n-1}\bar{\zeta}_{i_{n-1}}$ , the value of  $\lambda$  that maximises  $\mathcal{V}$  is the only solution that lies in the interval  $-1/i_1\bar{\zeta}_{i_1} < \lambda < -1/i_2\bar{\zeta}_{i_2}$  and the value that minimises  $\mathcal{V}$  is the only solution in  $-1/i_{n-2}\bar{\zeta}_{i_{n-2}} < \lambda < -1/i_{n-1}\bar{\zeta}_{i_{n-1}}$ .

**Proof of Theorem 1.** Choose a vector  $\Delta_i$  in  $\mathbb{R}^{n-1}$  that is orthogonal both to  $i\Omega_i$  and to  $i$ , that is, such that  $\sum_i i\Delta_i\Omega_i = 0$  and  $\sum_i i\Omega_i = 0$ . Since  $\alpha/ia_n + (1 - \alpha)\Delta_i$  is a set of continuous functions of  $\alpha$ , its minimum is also continuous in  $\alpha$ . Moreover, the minimum is clearly positive if  $\alpha = 1$ , while it is negative by construction if  $\alpha = 0$ . The theorem follows from the intermediate value theorem.  $\square$

**Proof of Theorem 2.** The proof is similar to the previous one. Choose a function  $\Delta(f)$  in  $C^\infty(0, 1)$  to satisfy both  $\int_{1/N}^1 df f\Omega(f)\Delta(f) = 0$  and  $\int_{1/N}^1 df f\Delta(f) = 0$ . (Since these conditions correspond just to two independent functionals of  $\Delta(f)$  and  $C^\infty(0, 1)$  is an infinite-dimensional linear space, the existence of such a function is guaranteed). Since  $\alpha/f \log(N) + (1 - \alpha)\Delta(f)$  is a continuous functions of  $\alpha$  and its infimum is not  $\pm\infty$ , its infimum is also continuous in  $\alpha$ . Moreover, the infimum is clearly positive if  $\alpha = 1$ , while it is negative by construction if  $\alpha = 0$ . The theorem follows from the intermediate value theorem.  $\square$

**Proof of Theorem 3.** The vectors  $\Omega_i$  are a basis of the subspace  $\mathbb{R}^{n-2} \subset \mathbb{R}^{n-1}$  defined by the condition  $\sum_i \Omega_i = 0$ , that is, the space of vectors orthogonal to the vector whose components are  $v_i = 1$ . Therefore the only vectors  $i\Delta_i$  that are orthogonal to all the vectors in this basis are precisely of the form  $i\Delta_i \propto v_i$ , that is,  $\Delta_i = const/i$ .  $\square$

The theorems on the form of optimal tests can be easily proved from a general lemma.

**Lemma A1.** Consider a function  $f : \mathbb{R}^M \setminus \{0\} \rightarrow \mathbb{R}$  of the form:

$$f(\vec{v}) = \frac{\vec{v} \cdot \vec{w}}{\sqrt{\vec{v} \cdot Q\vec{v}}} \tag{A52}$$

where  $\vec{w} \in \mathbb{R}^M$  and  $Q$  is a  $M \times M$  symmetric positive matrix, and a  $K \times M$  matrix  $R$  with  $K < M$  and maximum rank. The extrema of the function  $f$  restricted to the subspace  $R\vec{v} = 0$  are given by:

$$\vec{v}_\alpha = \alpha \left( Q^{-1}\vec{w} - Q^{-1}R^t \left( RQ^{-1}R^t \right)^{-1} RQ^{-1}\vec{w} \right) \tag{A53}$$

The extrema with  $\alpha > 0$  are maxima and the extrema with  $\alpha < 0$  are minima of the function  $f$ . These extrema satisfy the identity:

$$\vec{v}_\alpha \cdot Q\vec{v}_\alpha = \alpha \vec{v}_\alpha \cdot \vec{w} \tag{A54}$$

**Proof.** The existence of maxima and minima can be proved by the Weierstrass extreme value theorem. In fact  $f$  is continuous and invariant under a homothety with centre in the origin of  $\mathbb{R}^M$  and positive scale factor, therefore the codomain of the function on the linear subspace defined by  $R\vec{v} = 0$  is the same as the codomain of its restriction to the submanifold of unit vectors  $|\vec{v}| = 1$ , which is a compact space. The restriction of  $f$  is also continuous and the conclusion follows. To determine the extrema, the method of Lagrange multipliers states that it is sufficient to extremise the function:

$$F(\vec{v}, \vec{\lambda}) = \frac{\vec{v} \cdot \vec{w}}{\sqrt{\vec{v} \cdot Q\vec{v}}} + \vec{\lambda} \cdot R\vec{v} \tag{A55}$$

and since there are no boundaries, this is equivalent to the solution of the equations:

$$0 = \vec{\nabla}_v f = \frac{\vec{w}}{\sqrt{\vec{v} \cdot Q\vec{v}}} - \frac{\vec{v} \cdot \vec{w}}{(\vec{v} \cdot Q\vec{v})^{3/2}} Q\vec{v} + R^t \vec{\lambda} \tag{A56}$$

$$0 = \vec{\nabla}_\lambda f = R\vec{v} \tag{A57}$$

The solution satisfies:

$$\frac{\vec{v} \cdot \vec{w}}{\vec{v} \cdot Q \vec{v}} \vec{v} = Q^{-1} \vec{w} + Q^{-1} R^t \vec{\lambda} \quad (\text{A58})$$

and multiplying it by  $R$  and using (A57) we obtain:

$$\vec{\lambda} = -\left(RQ^{-1}R^t\right)^{-1}RQ^{-1}\vec{w} + \vec{l} \quad , \quad R^t\vec{l} = 0 \quad (\text{A59})$$

that can be inserted again in Equation (A58) to eliminate  $\vec{\lambda}$ . The resulting equation in  $\vec{v}$  admits only solutions of the form (A53) and by substituting (A53) into it, it can be checked that all values of  $\alpha \neq 0$  correspond to solutions of (A56) and (A57). The invariance of  $f$  under a central homothety with positive scale factor implies that the value of the function does not depend on  $|\alpha|$ . The function is positive for  $\alpha > 0$  and negative for  $\alpha < 0$ , therefore solutions with  $\alpha > 0$  correspond to maxima and solutions with  $\alpha < 0$  correspond to minima. The identity (A54) can be proved by substituting the solution (A53).

□

**Proof of Theorem 4.** The expected values of the tests (41) have the same functional form as the function  $f$  of Lemma A1. The correspondence is the following:

$$\vec{v} \rightarrow v_{\bar{I}} = \sigma(\bar{\mathbf{I}})\Omega_{\bar{\mathbf{I}}}^{(n_{\bar{I}})} \quad (\text{A60})$$

$$\vec{w} \rightarrow w_{\bar{I}} = \bar{\mu}_{\bar{I}} \quad (\text{A61})$$

$$Q \rightarrow Q_{\bar{I}\bar{J}} = \mu_{\bar{I}\bar{J}} - \mu_{\bar{I}}\mu_{\bar{J}} \quad (\text{A62})$$

$$R \rightarrow R_{k,\bar{I}} = \mu_{\bar{I}}^{(k)} \quad (\text{A63})$$

and the positivity of the matrix  $Q$  is guaranteed by the positivity of the variance for all possible choices of the weights. Application of Lemma A1 with  $\alpha = 1$  gives the result (50). □

**Proof of Theorem 5.** We can immediately solve Equation (58) for  $\gamma$  and substitute it in Equation (57). Then  $\gamma$  is a function of the other weights and the maximisation is unconstrained. It can be seen that also in this case, the expected values of the tests have the same functional form as the function  $f$  of Lemma A1. The correspondence is the following:

$$\vec{v} \rightarrow v_{\bar{I}} = \sigma(\bar{\mathbf{I}})\Gamma_{\bar{\mathbf{I}}}^{(n_{\bar{I}})} \quad (\text{A64})$$

$$\vec{w} \rightarrow w_{\bar{I}} = \bar{\mu}_{\bar{I}} - \mu_{\bar{I}} \quad (\text{A65})$$

$$Q \rightarrow Q_{\bar{I}\bar{J}} = \mu_{\bar{I}\bar{J}} - \mu_{\bar{I}}\mu_{\bar{J}} \quad (\text{A66})$$

$$R \rightarrow \text{empty } 0 \times M \text{ matrix} \quad (\text{A67})$$

and the positivity of the matrix  $Q$  is implied by the positivity of the variance. Then the result (63) follows from Lemma A1 with  $\alpha = 1$ . □

## References

1. Kreitman, M. Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster*. *Nature* **1983**, *304*, 412–417. [[CrossRef](#)] [[PubMed](#)]
2. Hudson, R.; Kreitman, M.; Aguadé, M. A test of neutral molecular evolution based on nucleotide data. *Genetics* **1987**, *116*, 153. [[CrossRef](#)] [[PubMed](#)]
3. Lewontin, R.; Krakauer, J. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics* **1973**, *74*, 175. [[CrossRef](#)] [[PubMed](#)]
4. Sabeti, P.; Reich, D.; Higgins, J.; Levine, H.; Richter, D.; Schaffner, S.; Gabriel, S.; Platko, J.; Patterson, N.; McDonald, G.; et al. Detecting recent positive selection in the human genome from haplotype structure. *Nature* **2002**, *419*, 832–837. [[CrossRef](#)]
5. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **1989**, *123*, 585. [[CrossRef](#)] [[PubMed](#)]

6. Fu, Y.; Li, W. Statistical tests of neutrality of mutations. *Genetics* **1993**, *133*, 693. [[CrossRef](#)]
7. Fay, J.; Wu, C. Hitchhiking under positive Darwinian selection. *Genetics* **2000**, *155*, 1405. [[CrossRef](#)]
8. Achaz, G. Frequency Spectrum Neutrality Tests: One for All and All for One. *Genetics* **2009**, *183*, 249. [[CrossRef](#)]
9. Fu, Y. Statistical properties of segregating sites. *Theor. Popul. Biol.* **1995**, *48*, 172–197. [[CrossRef](#)]
10. Ferretti, L.; Perez-Enciso, M.; Ramos-Onsins, S. Optimal Neutrality Tests Based on Frequency Spectrum. *Genetics* **2010**, *186*, 353–365. [[CrossRef](#)]
11. Fu, Y. New statistical tests of neutrality for DNA samples from a population. *Genetics* **1996**, *143*, 557. [[CrossRef](#)] [[PubMed](#)]
12. Ferretti, L.; Klassmann, A.; Raineri, E.; Ramos-Onsins, S.E.; Wiehe, T.; Achaz, G. The neutral frequency spectrum of linked sites. *Theor. Popul. Biol.* **2018**, *123*, 70–79. [[CrossRef](#)] [[PubMed](#)]
13. Klassmann, A.; Ferretti, L. The third moments of the site frequency spectrum. *Theor. Popul. Biol.* **2018**, *120*, 16–28. [[CrossRef](#)] [[PubMed](#)]
14. Giner-Delgado, C.; Villatoro, S.; Lerga-Jaso, J.; Gayà-Vidal, M.; Oliva, M.; Castellano, D.; Pantano, L.; Bitarello, B.D.; Izquierdo, D.; Noguera, I.; et al. Evolutionary and functional impact of common polymorphic inversions in the human genome. *Nat. Commun.* **2019**, *10*, 4222. [[CrossRef](#)] [[PubMed](#)]
15. Fu, Y. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **1997**, *147*, 915. [[CrossRef](#)]
16. Zeng, K.; Fu, Y.; Shi, S.; Wu, C. Statistical tests for detecting positive selection by utilizing high-frequency variants. *Genetics* **2006**, *174*, 1431. [[CrossRef](#)]
17. Tajima, F. Evolutionary relationship of DNA sequences in finite populations. *Genetics* **1983**, *105*, 437. [[CrossRef](#)]
18. Watterson, G. On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* **1975**, *7*, 256. [[CrossRef](#)]
19. Achaz, G. Testing for neutrality in samples with sequencing errors. *Genetics* **2008**, *179*, 1409. [[CrossRef](#)]
20. Ramos-Onsins, S.E.; Mitchell-Olds, T. Mlcoalsim: Multilocus coalescent simulations. *Evol. Bioinform.* **2007**, *3*, 41–44. [[CrossRef](#)]
21. Schaeffer, S. Molecular population genetics of sequence length diversity in the Adh region of *Drosophila pseudoobscura*. *Genet. Res.* **2003**, *80*, 163–175. [[CrossRef](#)] [[PubMed](#)]
22. Schmid, K.; Ramos-Onsins, S.; Ringys-Beckstein, H.; Weisshaar, B.; Mitchell-Olds, T. A multilocus sequence survey in *Arabidopsis thaliana* reveals a genome-wide departure from a neutral model of DNA sequence polymorphism. *Genetics* **2005**, *169*, 1601. [[CrossRef](#)] [[PubMed](#)]
23. Hutter, S.; Li, H.; Beisswanger, S.; De Lorenzo, D.; Stephan, W. Distinctly different sex ratios in African and European populations of *Drosophila melanogaster* inferred from chromosomewide single nucleotide polymorphism data. *Genetics* **2007**, *177*, 469. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.