

PAPER

Fisher matrix for multiple tracers: the information in the cross-spectra

To cite this article: L. Raul Abramo *et al* JCAP04(2022)013

View the [article online](#) for updates and enhancements.

You may also like

- [DECONSTRUCTING THERMAL SUNYAEV-ZEL'DOVICH—GRAVITATIONAL LENSING CROSS-CORRELATIONS: IMPLICATIONS FOR THE INTRACLUSTER MEDIUM](#)
N. Battaglia, J. C. Hill and N. Murray
- [Forecasts of cosmological constraints from HI intensity mapping with FAST, BINGO and SKA-I](#)
Elimboto Yohana, Yi-Chao Li and Yin-Zhe Ma
- [Improvement of the parameter measurement accuracy by the third-generation gravitational wave detector Einstein Telescope](#)
Hee-Suk Cho



IOP | ebooks™

Bringing together innovative digital publishing with leading authors from the global scientific community.

Start exploring the collection—download the first chapter of every title for free.

Fisher matrix for multiple tracers: the information in the cross-spectra

L. Raul Abramo, João Vitor Dinarte Ferri and Ian Lucas Tashiro

Departamento de Física Matemática, Instituto de Física, Universidade de São Paulo,
R. do Matão 1371, São Paulo 05508-090, SP, Brazil

E-mail: abramo@if.usp.br, joao.vitor.ferri@usp.br, iantashiro@usp.br

Received December 10, 2021

Revised February 21, 2022

Accepted March 16, 2022

Published April 8, 2022

Abstract. We derive general expressions for the multi-tracer Fisher matrix, both assuming that the cross-spectra are constrained by the auto-spectra, and also allowing for independent degrees of freedom in the cross-spectra. We show that, just like the ratios of power spectra, the independent degrees of freedom of the cross-spectra are also not constrained by cosmic variance. Moreover, whereas the uncertainties in the ratios of power spectra decrease with the number density of the tracers as $\sim 1/\sqrt{\bar{n}}$, the uncertainties in the independent degrees of freedom of the cross-spectra decrease even faster, as $\sim 1/\bar{n}$. We also derive simple expressions for the optimal number of tracers in a survey.

Keywords: power spectrum, galaxy clustering, redshift surveys

ArXiv ePrint: [2112.01812](https://arxiv.org/abs/2112.01812)

Contents

1	Introduction	1
2	Fisher matrix for two-point functions	3
2.1	Fisher matrix for the power spectra	4
2.2	Example: two tracers	6
3	Fisher information and the optimal number of tracers	6
3.1	Fisher summary statistics for the auto-spectra	7
3.2	Fisher summary statistics: including the cross-spectra	10
4	The independent degrees of freedom: diagonalizing the Fisher matrix	12
5	Conclusions	16

1 Introduction

The new generation of galaxy surveys are mapping large swaths of the observable Universe using a variety of objects, and with increasing completeness [1–4]. These surveys allow us to infer cosmological information from the positions of galaxies and other tracers of large-scale structure mainly through measurements of their clustering. In particular, when combining two or more tracers, we are able to measure not only their auto-spectra (or, equivalently, their auto-correlations), but also their cross-spectra. And as we increase the diversity of tracers in our surveys, the relative importance of the cross-spectra grows as well: given N tracers there are N auto-spectra, and $N(N - 1)/2$ cross-spectra.

A further advantage of multiple tracers is the fact that some physical parameters can be measured with an accuracy that is not limited by cosmic variance — i.e., we are not necessarily constrained only by the survey volume, and can improve the measurements of some physical parameters by increasing the numbers of tracers in the same volume [5, 6]. As shown by [7], the independent degrees of freedom measured by galaxy surveys are split into two branches: on one hand, the total clustering of the survey (a single degree of freedom), which includes observables such as the shape of the matter power spectrum, is severely constrained by cosmic variance: no matter how many tracers we observe, we are limited by the volume of the survey, which imposes a lower bound to the uncertainties in that total clustering. On the other hand, the *ratios* of power spectra of the tracers (or *relative* clusterings) are independent from the total clustering, as well as from each other, and their covariance is not limited by cosmic variance: it’s always possible to beat down the noise in those variables by detecting larger numbers of tracers in any given survey volume.

However, in this argument it is often assumed that the auto-power spectra and the cross-spectra are manifestations of the same basic degrees of freedom: the biases, the matter growth rate, the amplitude and shape of the matter power spectrum, etc. In practice, this is equivalent to assuming that $P_{ij} = v_i v_j$, and taking these N parameters v_i as the fundamental degrees of freedom. In fact, in linear theory the information in the cross-spectra is degenerate with that already contained in the auto-spectra — for a review see, e.g., [8]. A linear biasing

model implies that the *observed* power spectra (the data covariance in Fourier space), in redshift space, is given by:

$$P_{ij}(\vec{k}) = b_i b_j P_m(\vec{k}) + \frac{\delta_{ij}}{\bar{n}_i} \rightarrow (b_i + f\mu^2)(b_j + f\mu^2)P_m(k) + \frac{\delta_{ij}}{\bar{n}_i}, \quad (1.1)$$

where b_i are the biases of the tracers $i = 1, 2, \dots, N$, f is the matter growth rate, $\mu = \hat{r} \cdot \hat{k}$ is the angle of the Fourier mode with the line of sight, and \bar{n}_i are the number densities of the tracers. In the second part of the expression above we also used the flat-sky approximation to indicate explicitly the bias model — but this assumption is irrelevant for the arguments presented in this paper.

Although the assumption that the degrees of freedom in the cross-spectra are contained in the auto-spectra is approximately correct in the linear regime, and assuming a deterministic biasing model, it misses many important features such as the one-halo term, exclusion effects, stochasticity, as well as deviations from the ideal Poisson shot noise model [8–12].

Extensions of the linear and deterministic biasing model motivated by perturbation theory introduce additional dependencies in the tracer density contrasts, typically of the form:

$$\delta_i \rightarrow B_i \delta_m + \epsilon_i^P + \epsilon_i^S, \quad (1.2)$$

where $B_i = b_i + f\mu^2$, and ϵ_i^P is a shot noise stochastic term that, under the assumption of Poisson statistics, obeys $\langle \epsilon_i^P \epsilon_j^P \rangle = \delta_{ij}/\bar{n}_i$. The last term in eq. (1.2), ϵ_i^S , collects all additional stochastic terms and non-linearities (e.g., the dependencies on δ_m^2). A more general relation between the matter density field and the tracers, similar to eq. (1.2), was in fact used by Gil-Marín et al. [13] to model how the multi-tracer approach could improve measurements of the matter growth through RSDs. In some cases, discarding these stochastic terms, especially in the cross-spectra, can lead to systematics in the measurements of parameters such as f_{NL} [14]. Recently, Mergulhão et al. [15] showed that by splitting a halo population in two according to their mass allows us to measure these bias parameters with higher precision, which implies that, in practice, the advantages of the multi-tracer approach can leverage the larger number of parameters that comes together with considering more tracers.

Let's assume, for simplicity, that shot noise is exactly Poissonian and is uncorrelated with either δ_m or ϵ_i^S . The power spectrum corresponding to the pair $\langle \delta_i \delta_j \rangle$, after shot noise subtraction, takes the form:

$$\hat{P}_{ij} \rightarrow B_i B_j P_m + \langle \epsilon_i^S \epsilon_j^S \rangle, \quad (1.3)$$

where we assumed that the stochastic terms are uncorrelated with the density contrast. Even if we assume the simplest stochastic model, $\langle \epsilon_i^S \epsilon_j^S \rangle = \delta_{ij} S_i$, we still find that the cross-spectra carry additional information with respect with the auto-spectra, in the sense:

$$\hat{P}_{ii} \hat{P}_{jj} - \hat{P}_{ij}^2 \rightarrow \hat{P}_{ii} S_j + S_i \hat{P}_{jj} + S_i S_j. \quad (1.4)$$

In this paper we show that, with multiple tracers, the accuracy with which we can measure the *irreducible* information in the cross-spectra, in the sense of eq. (1.4), is not constrained by cosmic variance. Moreover, the uncertainties in those irreducible degrees of freedom of the cross-spectra fall even *faster* with the number density of tracers, when compared with the ratios of auto-spectra. Therefore, cosmic variance cancellation is not only a feature that allows us to measure bias, redshift-space distortion (RSD) parameters [6] or primordial non-Gaussianities [5] with increased accuracy, but it also opens the way to measure some of the

parameters in the perturbative bias expansion to much higher accuracy by using multiple tracers and their cross-correlations.

We start, in section 2, by deriving several useful results related to the multi-tracer Fisher matrix, with and without cross-spectra as independent degrees of freedom. We also present a general expression for the Fisher matrix with any data covariance, and show that its inverse is exactly the covariance matrix that we expect under the Gaussian approximation. Then, in section 3 we show how to maximize the information of a galaxy survey, in terms of the optimal number tracers, by using two different summary statistics for the total amount of Fisher information in that survey. Finally, in section 4 we show that the extra degrees of freedom that arise from the cross-spectra are not constrained by cosmic variance, and can be measured with arbitrary accuracy (at least in principle). We conclude in section 5.

2 Fisher matrix for two-point functions

Let's say that we have many samples of some measurements f_i , from which we wish to estimate physical parameters through the quadratic form (the ‘‘correlations’’) $q_{ij} \rightarrow \langle f_i f_j \rangle$. We define the data covariance as:

$$C_{ij} = \langle f_i f_j \rangle = q_{ij} + s_{ij}, \quad (2.1)$$

where s_{ij} is the noise (assumed symmetric under $i \leftrightarrow j$). Under the hypothesis of Gaussianity for the data f_i , one can easily show by using Wick's theorem that the parameter covariance is given by:

$$\text{Cov}(q_{ij}, q_{i'j'}) = C_{ii'} C_{jj'} + C_{ij'} C_{ji'}. \quad (2.2)$$

Since by construction $q_{ij} = q_{ji}$, this set of parameters will count twice the correlations. For this reason, we introduce the following notation for all the non-equivalent pairs $\{i, j\}$:

$$q_{[ij]} = q_{ij}, \text{ for } i \leq j. \quad (2.3)$$

Clearly, if $i = 1, 2, \dots, N$, then the number of non-equivalent pairs is $N_p = N(N+1)/2$. Notice that with this notation the parameter covariance is still given by the same expression as above, i.e.:

$$\text{Cov}(q_{[ij]}, q_{[i'j']}) = C_{ii'} C_{jj'} + C_{ij'} C_{ji'}. \quad (2.4)$$

However, as opposed to the $((N, N) \times (N, N))$ array of eq. (2.2), in terms of the individual pairs $[ij]$ the expression above is an $N_p \times N_p$ matrix.

With this notation it is straightforward to show that the inverse of the parameter covariance, also known as the Fisher matrix, is given by:

$$F[q_{[ij]}, q_{[i'j']}] = F_{[ij],[i'j']} = \left(1 - \frac{1}{2}\delta^{ij}\right) \left(1 - \frac{1}{2}\delta^{i'j'}\right) \left(C_{ii'}^{-1} C_{jj'}^{-1} + C_{ij'}^{-1} C_{ji'}^{-1}\right). \quad (2.5)$$

where C_{ij}^{-1} is the inverse of the data covariance, i.e., $\sum_j C_{ij}^{-1} C_{ji} = \delta_{ii}$.

In order to show that the covariance of eq. (2.2) is indeed the inverse of the Fisher matrix of eq. (2.5) we need the identity:

$$\begin{aligned} 2 \sum_{[mn]} C_{im}^{-1} C_{mi'} C_{jn}^{-1} C_{nj'} &= \sum_{mn} C_{im}^{-1} C_{mi'} C_{jn}^{-1} C_{nj'} + \sum_m C_{im}^{-1} C_{mi'} C_{jm}^{-1} C_{mj'} \\ &= \delta_{ii'} \delta_{jj'} + \sum_m C_{im}^{-1} C_{mi'} C_{jm}^{-1} C_{mj'}, \end{aligned}$$

where, following the notation introduced in eq. (2.3), the sum $\sum_{[mn]}$ is limited to the indices $m \leq n$. With this result it is then trivial to show that, as expected:

$$\sum_{[mn]} F_{[ij],[mn]} \text{Cov}_{[mn],[i'j']} = \delta_{[ij],[i'j']}. \quad (2.6)$$

We now present two key results for the covariance and Fisher matrices of the 2-point functions of Gaussian variables, which we will employ later. The first identity concerns the determinants of the Fisher matrix and of the parameter covariance (which are, of course, the inverse of each other). It is possible to show that:

$$\det(\text{Cov}_{[ij],[i'j']}) = 2^N (\det C)^{N+1}, \quad (2.7)$$

and therefore that:

$$\det(F_{[ij],[i'j']}) = 2^{-N} (\det C)^{-N-1} = 2^{-N} (\det C^{-1})^{N+1}. \quad (2.8)$$

We stress the fact that $F_{[ij],[i'j']}$ is an $N_p \times N_p$ matrix, while C is an $N \times N$ matrix.

The second identity is the fact that the “grand sum” of the Fisher matrix is proportional to the square of the grand sum of the inverse data covariance, namely:

$$\sum_{[ij]} \sum_{[i'j']} F_{[ij],[i'j']} = \frac{1}{2} \left(\sum_{ij} C_{ij}^{-1} \right)^2. \quad (2.9)$$

We will use eqs. (2.8) and (2.9) in section 3, when we optimize the number of tracers in a survey.

2.1 Fisher matrix for the power spectra

The fundamental degrees of freedom in a survey are the positions of the tracers (galaxies, halos or other point-like objects that follow the underlying matter distribution). When we measure the number densities of tracer species i , $n_i(\vec{x})$, over some volume around the position \vec{x} , that number reflects some mean density of those tracers, $\bar{n}_i(\vec{x})$, as well as the fluctuations $\delta n_i = n_i - \bar{n}_i$. From these observables we compute the main object that carries information about cosmology, the data (or “pixel”) covariance:

$$C_{ij}(\vec{x}, \vec{y}) = \langle \delta n_i(\vec{x}) \delta n_j(\vec{y}) \rangle = \bar{n}_i(\vec{x}) \bar{n}_j(\vec{y}) \xi_{ij}(\vec{x}, \vec{y}) + \bar{n}_i(\vec{x}) \delta_{ij} \delta_D(\vec{x} - \vec{y}), \quad (2.10)$$

where $\xi_{ij}(\vec{x}, \vec{y})$ is the 2-point correlation function, and the last term is shot noise, which we assume here to follow Poisson statistics. The multi-tracer 2-point correlation function is generally assumed to be related to the matter correlation function, $\xi^{(m)}(\vec{x}, \vec{y})$, through some knowable relations such as tracer bias, redshift-space distortions, etc. In real space (i.e., excluding redshift-space distortion), the matter two-point correlation function can be written in terms of the matter power spectrum as:

$$\xi^{(m)}(\vec{x}, \vec{y}) = \xi^{(m)}(|\vec{x} - \vec{y}|) = \int \frac{d^3k}{(2\pi)^3} e^{-i\vec{k} \cdot (\vec{x} - \vec{y})} P^{(m)}(k). \quad (2.11)$$

We can also work directly in Fourier space, and derive the Fisher and covariance matrices for the power spectra. In that case it is more convenient to work with the density contrasts for the tracers, $\delta_i = (n_i - \bar{n}_i)/\bar{n}_i$. The Fourier mode \vec{k} of the density contrast can be expressed as:

$$d_i^a(\vec{k}) = \{\tilde{\delta}_i(\vec{k}), \tilde{\delta}_i^*(\vec{k})\}, \quad (2.12)$$

where $i = 1, 2, \dots, N$ denotes the tracer, and $a = 1, 2$ stand for the Fourier mode and its complex conjugate, respectively. The data covariance is then:

$$\langle d_i^a(\vec{k}) d_j^b(\vec{k}') \rangle = D^{ab} C_{ij}(\vec{k}, \vec{k}') = C_{ij}^{ab}(\vec{k}, \vec{k}'), \quad (2.13)$$

where $D^{ab} = 1 - \delta^{ab}$. The data covariance in Fourier space is then simply the observed power spectrum for those tracers, including shot noise if it is an auto-spectrum:

$$C_{ij}(\vec{k}, \vec{k}') = \delta_{\vec{k}\vec{k}'} \left(P_{ij} + \frac{\delta_{ij}}{\bar{n}_i} \right), \quad (2.14)$$

where in the continuum limit we have $\delta_{\vec{k}\vec{k}'} \rightarrow (2\pi)^3 \delta_D(\vec{k} - \vec{k}')$, but for simplicity here we can consider this to be a Kronecker delta-function, up to a constant.

One of the ways in which we can derive the Fisher matrix is through the Hessian of the log-likelihood. Given a set of parameters θ^μ , the Fisher matrix is given by the generalized trace [16]:

$$F_{\mu\nu} = \frac{1}{4} \sum_k V \tilde{V}_k \sum_{ij'j'} \sum_{aba'b'} \frac{\partial C_{ij}^{ab}}{\partial \theta^\mu} [C_{j'i'}^{ba'}]^{-1} \frac{\partial C_{i'j'}^{a'b'}}{\partial \theta^\nu} [C_{j'i}^{b'a}]^{-1}, \quad (2.15)$$

where V is the survey volume, \tilde{V}_k is the volume in Fourier space of the bandpowers (Fourier bins) k , and the additional factor of 1/2 in eq. (2.15) is due to the fact that our degrees of freedom include the Fourier modes twice. Notice that this Fisher matrix is diagonal in \vec{k} due to the diagonal nature of the data covariance, eq. (2.14), and for simplicity for the remainder of this section we will omit the Fourier space indices.

Using the fact that the data covariance is separable, $[C_{ij}^{ab}]^{-1} = [D^{ab}]^{-1} C_{ij}^{-1}$, and using that $[D^{ab}]^{-1} = D^{ab}$, we obtain $\sum_{abcd} D^{ab} D^{bc} D^{cd} D^{da} = \sum_{ac} \delta_{ac} \delta_{ca} = 2$. Hence:

$$F_{\mu\nu} = \frac{1}{2} \sum_k V \tilde{V}_k \sum_{ij'j'} \frac{\partial C_{ij}}{\partial \theta^\mu} C_{j'i}^{-1} \frac{\partial C_{i'j'}}{\partial \theta^\nu} C_{j'i}^{-1}, \quad (2.16)$$

In Fourier space the inverse of the data covariance has a trivial expression, namely:

$$C_{ij}^{-1} = \bar{n}_i \delta_{ij} - \bar{n}_i \frac{P_{ij}}{1 + \mathcal{P}} \bar{n}_j, \quad (2.17)$$

with $\mathcal{P} = \sum_i \bar{n}_i P_{ii}$ — in fact, the denominator in the second term is precisely $\det C = 1 + \mathcal{P}$.

We now set the parameters θ^μ to be the auto- and cross-spectra of the tracers evaluated at some bandpower, $P_{ij}(k)$. Just as was the case in our basic example of the previous section, these spectra are symmetric, $P_{ij} = P_{ji}$. In order to avoid double-counting these degrees of freedom we define the non-degenerate auto- and cross-spectra as:

$$P_{[ij]} = P_{ij}, \text{ for } i \leq j. \quad (2.18)$$

Therefore, the spectra with any index can be expressed as:

$$P_{ij} = P_{[ij]} + P_{[ji]} - \delta_{ij} P_{ii}, \quad (2.19)$$

and a similar expression for the data covariance. We can now evaluate the partial derivatives assuming that the parameters are the non-degenerate spectra:

$$\frac{\partial C_{ij}(k)}{\partial P_{[i'j']}(k')} = \delta_{k,k'} [\delta_{ii'} \delta_{jj'} + \delta_{ij'} \delta_{ji'} - \delta_{ij} \delta_{j'i'} \delta_{i'j'} \delta_{j'i}] \equiv \delta_{k,k'} \delta_{[ij],[i'j']}. \quad (2.20)$$

Substituting this identity into eq. (2.16) results in a Fisher matrix which is diagonal in the bandpowers, and which can be expressed for each mode k as:

$$F[P^{[ij]}, P^{[i'j']}] = F^{[ij],[i'j']} = V\tilde{V}_k \left(1 - \frac{1}{2}\delta^{ij}\right) \left(1 - \frac{1}{2}\delta^{i'j'}\right) \left(C_{ii'}^{-1}C_{jj'}^{-1} + C_{ij'}^{-1}C_{ji'}^{-1}\right). \quad (2.21)$$

It is immediately obvious that the results of the previous section apply here, so the inverse of this Fisher matrix is the usual expression for the covariance of the spectra:

$$\text{Cov}[P_{[ij]}, P_{[i'j']}] = \text{Cov}_{[ij],[i'j']} = \frac{1}{V\tilde{V}_k} (C_{ii'}C_{jj'} + C_{ij'}C_{ji'}). \quad (2.22)$$

2.2 Example: two tracers

As an example, we write explicit expressions for the case when we have two tracers. The covariance matrix becomes:

$$\text{Cov}[P_{[ij]}, P_{[i'j']}] = \frac{1}{V\tilde{V}_k} \begin{pmatrix} 2C_{11}^2 & 2C_{11}C_{12} & 2C_{12}^2 \\ 2C_{11}C_{12} & C_{11}C_{22} + C_{12}^2 & 2C_{12}C_{22} \\ 2C_{12}^2 & 2C_{12}C_{22} & 2C_{22}^2 \end{pmatrix}, \quad (2.23)$$

where we have ordered the degrees of freedom as $\{P_{11}, P_{12}, P_{22}\}$. The Fisher matrix in that case, derived directly from eq. (2.21), is given by:

$$F[P_{[ij]}, P_{[i'j']}] = V\tilde{V}_k \begin{pmatrix} \frac{1}{2}C_{11}^{-2} & C_{11}^{-1}C_{12}^{-1} & \frac{1}{2}C_{12}^{-2} \\ C_{11}^{-1}C_{12}^{-1} & C_{11}^{-1}C_{22}^{-1} + C_{12}^{-2} & C_{12}^{-1}C_{22}^{-1} \\ \frac{1}{2}C_{12}^{-2} & C_{12}^{-1}C_{22}^{-1} & \frac{1}{2}C_{22}^{-2} \end{pmatrix}. \quad (2.24)$$

In the particular case of the power spectra, the inverse of the data covariance is given by eq. (2.17).

3 Fisher information and the optimal number of tracers

The results above can serve as a guide to a first attempt at organizing the data of galaxy surveys: is it worth splitting a galaxy population into sub-types with different properties, or upon doing so we risk degrading the discriminating power of our survey? Although the precise answer depends on the nature of the tracers, as well as the kinds of parameters we are trying to constrain (in that respect see also [13]), there are some general trends that can be inferred in terms of the summary statistics of the Fisher matrix.

Before we proceed any further, it is useful to express the spectra in terms of signal-to-noise ratios (SNR), where the noise is given by shot noise (assumed Poissonian):

$$\mathcal{P}_{ij} = \sqrt{\bar{n}_i \bar{n}_j} P_{ij}. \quad (3.1)$$

We define the total clustering (or total SNR) as $\mathcal{P} = \sum_i \mathcal{P}_{ii} = \sum_i \bar{n}_i b_i^2 P^{(m)}$, and this quantity can be regarded as being approximately constant for a given survey. When two tracer species i and j are joined to form a composite tracer, their numbers are combined, $n_{i+j} = n_i + n_j$, but the number densities and linear biases are constrained by the relation $\bar{n}_{i+j}b_{i+j} = \bar{n}_i b_i + \bar{n}_j b_j$.

Since $\mathcal{P} \sim \sum_i \bar{n}_i b_i^2$, by joining or splitting tracers we can slightly increase or decrease the total SNR.

The Fisher and covariance matrices can also be expressed in terms of signal and noise. For generic degrees of freedom X_μ ($\mu = 1, 2, \dots, n$) we have:

$$\mathcal{F}[X_\mu, X_\nu] = X_\mu F[X_\mu, X_\nu] X_\nu = F[\log X_\mu, \log X_\nu] \quad (3.2)$$

$$\mathcal{C}[X_\mu, X_\nu] = \frac{\text{Cov}[X_\mu, X_\nu]}{X_\mu X_\nu} = \text{Cov}[\log X_\mu, \log X_\nu]. \quad (3.3)$$

Clearly, these Fisher and covariance matrices refer to the relative uncertainties in the parameters X_μ . All our summary statistics will be derived in terms of these relative uncertainties — or, equivalently, in terms of SNR.

There are arbitrarily many summary statistics that one can build from the Fisher matrix, and there is no single expression that can claim to capture the total information [17]. Although in principle we should be guided by invariants such as the trace or the determinant of the matrix, depending on the application one summary statistics may be more suitable than the other.

The determinant of the Fisher matrix (whose square root is known as the Jeffreys prior [18]) is evidently a convenient summary statistics, and in the context of cosmology, when the Fisher matrix is projected into some sub-space, it is also called “figure of merit” (FoM) [19]. In our context we define the FoM as:

$$\Delta = \det \mathcal{F}[X_\mu, X_\nu]. \quad (3.4)$$

The FoM corresponds to the inverse volume of an ellipsoid in n dimensions, defined by the 68% confidence limit of the corresponding multivariate Gaussian distribution. The smaller the volume in parameter space, the greater is the discriminating power.

Another useful summary statistics is a generalization of the χ^2 , given by the grand sum of the Fisher matrix. Let’s take the usual χ^2 :

$$\chi^2 = \sum_{\mu\nu} (X_\mu - \bar{X}_\mu) F[X_\mu, X_\nu] (X_\nu - \bar{X}_\nu) = \sum_{\mu\nu} \left(1 - \frac{\bar{X}_\mu}{X_\mu}\right) \mathcal{F}[X_\mu, X_\nu] \left(1 - \frac{\bar{X}_\nu}{X_\nu}\right), \quad (3.5)$$

where \bar{X}_μ are the fiducial values of the parameters X_μ . Now, make all X_μ equal to a certain fraction of the \bar{X}_μ , such that the terms multiplying \mathcal{F} in the sum above reduce to a constant. We then define the grand sum of the Fisher matrix as:

$$\Xi = \sum_{\mu\nu} \mathcal{F}[X_\mu, X_\nu] = \sum_{\mu\nu} X_\mu \mathcal{F}[X_\mu, X_\nu] X_\nu. \quad (3.6)$$

3.1 Fisher summary statistics for the auto-spectra

Before we examine the Fisher matrix in full generality, it is instructive to consider the case when the cross-spectra are *not* independent degrees of freedom, but are in fact constrained in terms of the auto-spectra, $P_{ij}^2 = P_{ii}P_{jj}$. This corresponds to assuming that the non-linear and stochastic terms $\epsilon_i^S \rightarrow 0$ in eq. (1.2), and that shot noise can be perfectly subtracted. In that case the Fisher matrix is given by [7, 20]:

$$\mathcal{F}[P_{ii}, P_{jj}] = \frac{V\tilde{V}_k \delta_{ij} P_{ii} \mathcal{P}(1 + \mathcal{P}) + P_{ii} P_{jj} (1 - \mathcal{P})}{4(1 + \mathcal{P})^2}. \quad (3.7)$$

It is worth stressing here that by including only the auto-spectra in the Fisher matrix does not mean that we are discarding measurements of the cross-spectra: it simply means that the degrees of freedom in the cross-spectra are assumed to be redundant with those already found in the cross-spectra.

It is straightforward to show that the FoM associated with the Fisher matrix of eq. (3.7) is given by:

$$\Delta = \left(\frac{V\tilde{V}_k}{4} \right)^N \frac{2\mathcal{P}^N}{(1+\mathcal{P})^{N+1}} \prod_i \mathcal{P}_{ii}. \quad (3.8)$$

We can now ask what happens as we change the number of tracers — either by splitting one tracer into two or more sub-types, or by combining many tracers into a single one. However, notice that the determinant of eq. (3.8) carries a phase-space volume factor for each degree of freedom, $(V\tilde{V}_k)^N$. Moreover, when projecting into a final set of parameters, those phase space volumes will tend to be compensated by the larger number of parameters from each tracer species. This discussion indicates that what we ought to use as a summary statistics is the Fisher information per unit phase space volume, which in this case (where we ignore the information from the cross-spectra) results in the renormalized FoM:

$$\Delta_{\text{Ph}} = \frac{\Delta_N}{(V\tilde{V}_k)^N} = \frac{1}{4^N} \frac{2\mathcal{P}^N}{(1+\mathcal{P})^{N+1}} \prod_i \mathcal{P}_{ii}. \quad (3.9)$$

In order to maximize the Fisher information it is clear that we must first maximize the clustering SNR \mathcal{P} , given the constraint that the total number of tracers is kept fixed — this is indeed what we expect from a survey where we observe some total number of objects, that we can subdivide into one or more species of tracers. But for a fixed number of tracers (N), what is the optimal way to draw the lines between the tracers?

The answer clearly depends on the way that bias varies with the number density. Let's assume that the bias is given by a power-law in terms of the number density: $b_i^2 \sim \bar{n}_i^{-\gamma}$. The total SNR is therefore given by:

$$\mathcal{P} = \sum_i \mathcal{P}_{ii} = \sum_i \bar{n}_i b_i^2 P^{(m)} = \bar{n}_T b_T^2 P^{(m)} \sum_i w_i^{1-\gamma}, \quad (3.10)$$

where $\bar{n}_T = \sum_i \bar{n}_i$, $b_i^2 = b_T^2 (\bar{n}_i/\bar{n}_T)^{-\gamma}$, and we defined the weights:

$$w_i = \frac{\bar{n}_i}{\bar{n}_T}. \quad (3.11)$$

Extremizing the total clustering SNR subject to the constraint that the total number of objects is fixed ($\sum_i w_i = 1$) leads to an “equipartition” between all the tracers, i.e., $w_i = 1/N$. This extremal value is a maximum only if $\gamma(1-\gamma) < 0$, or $0 < \gamma < 1$ — otherwise equipartition yields a minimum. Substituting $w_i = 1/N$ back into eq. (3.10) we obtain that $\mathcal{P} \rightarrow \bar{n}_T b_T^2 P^{(m)} N^\gamma$, which grows with the number of tracers if the power-law γ is in the range where equipartition yields a maximum of the clustering SNR.

Interestingly, the halo mass function and halo bias fits found by Tinker et al. [21] indicate that, for the halos masses of cosmological interest ($10^{10} \lesssim M[h^{-1}M_\odot] \lesssim 10^{15}$), this index is $0.1 \lesssim \gamma \lesssim 0.5$. Therefore, for tracers which behave similarly to dark matter halos, the configuration which maximizes the total SNR \mathcal{P} has the populations divided in equal numbers, with as many sub-species of tracers as practically possible. If, on the other

hand, the tracers behave in a completely different way, such that $\gamma \leq 0$ or $\gamma \geq 1$, then the maximum of \mathcal{P} is obtained by taking one of the $w_i \rightarrow 1$, and all others to zero — i.e., in such a situation it would be optimal to join all the tracers into a single population. In particular, if all tracers have the same bias (i.e., $\gamma = 0$), then it is sub-optimal to divide them into sub-species. For mixed tracers, such as galaxies, separated by properties such as stellar mass, luminosity, morphology, etc., as long as the parameter used to separate the populations generates a dependence of the bias with number density which follows the same trends as the dark matter halos, then the main conclusion remains the same.

Notice that equipartition in terms of the number densities, which maximizes \mathcal{P} , also implies equipartition of that total clustering SNR amongst all tracers: since $w_i = 1/N$, we have that $w_i^{1-\gamma} = N^{\gamma-1}$, and therefore $\mathcal{P}_{ii} = \mathcal{P}/N$. Moreover, equipartition of the clustering also happens to maximize the product $\prod_i \mathcal{P}_{ii}$, which implies that the FoM of eq. (3.9) is maximized when $w_i = 1/N$, and $\mathcal{P}_{ii} = \mathcal{P}/N$.

It is more enlightening to express the FoM at its maximum value in terms of the total SNR \mathcal{P} . Substituting the optimal configuration (i.e., equipartition) into eq. (3.8) we obtain:

$$\Delta_{\text{Ph}}^{\max} = \frac{1}{4^N} \frac{2 \mathcal{P}^N}{(1 + \mathcal{P})^{N+1}} \left(\frac{\mathcal{P}}{N} \right)^N. \quad (3.12)$$

While in this expression the total SNR \mathcal{P} still depends on the number of tracers through a relation such as $\mathcal{P} \sim N^\gamma$, we can find an approximate expression for the number of tracers that maximize the FoM, assuming that \mathcal{P} varies slowly with N — i.e., in the limit that γ is small. We have, in that limit:

$$\left. \frac{d\Delta_{\text{Ph}}^{\max}}{dN} \right|_{\mathcal{P}} = 0 \quad \Rightarrow \quad N_0^{\max} = \frac{1}{4e} \frac{\mathcal{P}^2}{1 + \mathcal{P}}, \quad (3.13)$$

where e is Euler's number. Clearly, then, we obtain that, as the total clustering strength \mathcal{P} increases, so grows the optimal number of tracers that we should use in order to maximize the information from our survey.

In a more realistic situation, we cannot hold the total clustering SNR \mathcal{P} constant as we change the number of species of tracers. In figure 1 (left panel) we show the FoM as a function of N , assuming that $\mathcal{P} = P_0 N^\gamma$ — here $P_0 = \bar{n}_T b_T^2 P^{(m)}$ is the baseline clustering SNR.¹ The solid, dashed and thin-dashed lines denote the power-law values $\gamma = 0, 0.05$ and 0.1 , while the colors refer to the values of P_0 which make $N_{\max}^0 = 2$ (red, $P_0 = 22.2$), 4 (orange, $P_0 = 44$), 6 (green, $P_0 = 65.7$) and 8 (blue, $P_0 = 87.5$), according to eq. (3.13) — i.e., assuming that $\gamma \rightarrow 0$. This plot shows that the higher the clustering SNR \mathcal{P} is, the higher is also the optimal value for the number of tracers, N^{\max} . The plot also shows that, for a fixed value of the baseline clustering P_0 , the optimal number of tracers grows slightly for higher values of γ (as denoted by the shift to higher values of N of the dashed lines with respect to the solid lines).

As an alternative to the determinant of the Fisher matrix, we could also consider the grand sum defined in eq. (3.6) as a summary statistics for the total Fisher information. In the case where the auto-spectra contain all the information, that grand sum is obtained directly from eq. (3.7):

$$\Xi = \frac{V \tilde{V}_k}{2} \frac{\mathcal{P}^2}{(1 + \mathcal{P})^2}. \quad (3.14)$$

¹E.g., a low-redshift survey with $\bar{n}_T = 2. \times 10^{-3} h^3 \text{Mpc}^{-3}$ and a mean bias $b_T = 1.5$, at a reference scale of $k_0 = 0.1 h \text{Mpc}^{-1}$, and considering a spectrum $P^{(m)}(k_0) = 10^4 h^{-3} \text{Mpc}^3$, yields $P_0 = 45$.

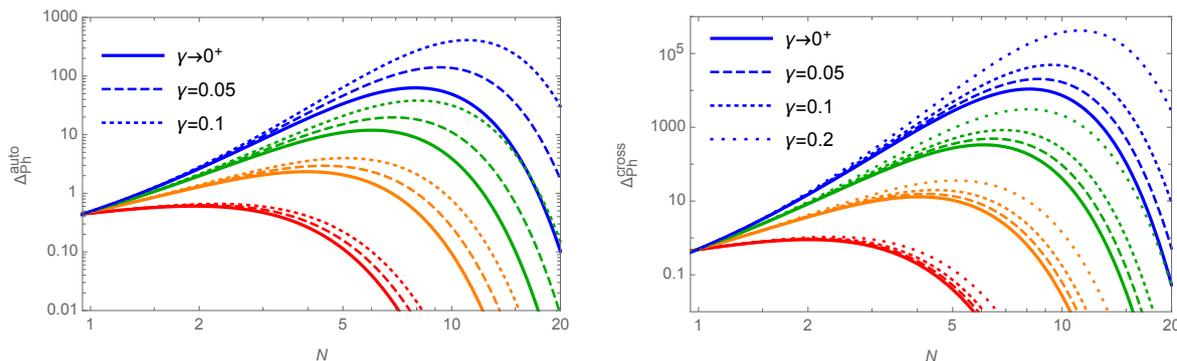


Figure 1. Figure of Merit (FoM) as a function of the number of tracers, N , assuming that $\mathcal{P} = P_0 N^\gamma$. Left panel: FoM considering only the auto-spectra, eq. (3.9), for the power law indices $\gamma \rightarrow 0^+$ (solid), 0.05 (dashed), and 0.1 (thin-dashed lines). The different colors refer to different values of P_0 that make $N_0^{\max} = 2$ (red), 4 (orange), 6 (green) and 8 (blue lines) — see eq. (3.13). Right panel: FoM considering the auto- and cross-spectra, eq. (3.15), for the power law indices $\gamma = 0, 0.05, 0.1$ and 0.2 , which are respectively denoted by the solid, dashed, thin-dashed and dotted lines. As in the left panel, the red, orange, green and blue lines refer to the values of P_0 that make $N_0^{\max} = 2, 4, 6$ and 8 — in this instance, see eq. (3.18). Notice that for $\gamma \leq 0$ the equipartition solution used to derive eqs. (3.9) and (3.15) is invalid.

Interestingly, this summary statistics does not carry an explicit dependency on the number of tracers: for a fixed volume and bandwidth, the total information depends only on the total clustering SNR \mathcal{P} . Since maximizing that SNR also maximizes the grand sum, this leads us back to the same conclusion that equipartition is the optimal solution.

It is rather interesting that figure 1 follows the intuitive notion that, given a fixed data set (which is our underlying assumption), there is an optimal number of parameters that we can try to measure. On one hand, too few parameters yield a small amount of information, but if on the other hand we try to measure too many parameters, then all those parameters become undetermined and the total information falls.

3.2 Fisher summary statistics: including the cross-spectra

We now include the independent information from the cross-spectra in our analysis. In that case, the determinant of the Fisher matrix given by eq. (2.5) can be derived directly from eq. (2.8).

Here we assume, *a posteriori*, that for the fiducial model we have $\mathcal{P}_{12}^2 \rightarrow \mathcal{P}_{11}\mathcal{P}_{22}$. This means that in our expressions we assume that the non-linear and stochastic terms ϵ_i are relatively small — a good approximation on large scales. The resulting FoM is then given by:

$$\Delta = \det(\mathcal{F}_{[ij],[i'j']}) = 2^{-N} \left(V \tilde{V}_k \right)^{N(N+1)/2} \frac{\prod_i \mathcal{P}_{ii}^2}{(1 + \mathcal{P})^{N+1}}. \quad (3.15)$$

This expression should be compared with eq. (3.8), in the case when the cross-spectra are not independent degrees of freedom. Once again, there are multiplicative factors of the phase space volume for each one of the degrees of freedom — in this case, the $N(N+1)/2$ auto- and cross-spectra. The renormalized FoM then becomes:

$$\Delta_{\text{Ph}} = 2^{-N} \frac{\prod_i \mathcal{P}_{ii}^2}{(1 + \mathcal{P})^{N+1}}. \quad (3.16)$$

Just as it happened in the previous section, maximizing the FoM leads to the equipartition of power between the tracers, $\mathcal{P}_{ii} \rightarrow \mathcal{P}/N$ — and the same conditions still apply regarding the dependence of the biases of the tracers with the number densities, which we assume to be $b_i^2 \sim \bar{n}_i^{-\gamma}$, with $0 < \gamma < 1$ so that the extremum of the FoM is in fact a maximum.

Finally, knowing that the equipartition of power maximizes the FoM $\Delta_{N,\text{Ph}}$, we can ask whether or not it is beneficial to split a sample into sub-types of tracers. Let's again consider the FoM at its maximum, i.e., substituting $P_{ii} = \mathcal{P}/N$ into eq. (3.16):

$$\Delta_{\text{Ph}}^{\text{max}} = \frac{1}{2^N N^{2N}} \frac{\mathcal{P}^{2N}}{(1 + \mathcal{P})^{N+1}}. \quad (3.17)$$

In the right panel of figure 1 we show the FoM of eq. (3.17) as a function of N , again assuming that $\mathcal{P} = P_0 N^\gamma$. In this case we have plotted the FoM for the values $\gamma = 0$ (solid), 0.05 (dashed), 0.1 (thin-dashed), and 0.2 (dotted lines), and the colors refer to the values of P_0 which make $N_0^{\text{max}} = 2$ (red, $P_0 = 60$), 4 (orange, $P_0 = 237$), 6 (green, $P_0 = 533$) and 8 (blue, $P_0 = 947$), according to eq. (3.18). As was the case when we eliminated the information in the cross-spectra (left panel of figure 1) we see that higher values of the clustering SNR \mathcal{P} lead to higher values for the optimal number of tracer species.

However, there are some important differences when we include the cross-spectra. First, for similar values of the baseline SNR P_0 , the optimal number of tracers when we disregard the degrees of freedom of the cross-spectra is higher than the optimal number of tracers when we take into account those degrees of freedom: e.g, compare the green lines in the left plot of figure 1 (for $P_0 = 65.7$) with the red lines in the plot on the right (for $P_0 = 65.7$). Discarding the cross-spectra, we would arrive at an optimal number of $N \sim 8$ –10 tracers if we neglected the cross-spectra, and an optimal number of $N \sim 2$ –3 tracers if we fully incorporate those degrees of freedom. Another way of putting this is to notice that, when including the cross-spectra, the baseline clustering SNR needs to be significantly higher for the same optimal number of tracers compared with the case where the cross-spectra are not independent degrees of freedom: e.g., for $N_0^{\text{max}} = 4$ we only need $P_0 = 44$ when the cross-spectra are discarded, while that value grows to $P_0 = 237$ when they are included.

The second difference is that, when we include the cross-spectra, the FoM becomes less sensitive to the power-law index γ , as can be seen by the difference between the solid, dashed and dotted lines of the plot in the right panel of figure 1, which is much less pronounced than in the case when we discarded the degrees of freedom in the cross-spectra (left panel of the same figure).

Just as we did in the previous section, an approximate expression for the optimal number of tracers can be found by taking the derivative of the FoM while assuming that the clustering SNR is kept fixed:

$$\left. \frac{d \Delta_{\text{Ph}}^{\text{max}}}{dN} \right|_{\mathcal{P}} = 0 \quad \Rightarrow \quad N_0^{\text{max}} = \frac{1}{\sqrt{2e}} \frac{\mathcal{P}}{\sqrt{1 + \mathcal{P}}}. \quad (3.18)$$

Since in the case where the cross-spectra are included the FoM is not very sensitive to γ (see figure 1), we can take the limit $\gamma \rightarrow 0$ and substitute the maximal number found in eq. (3.18) into eq. (3.17), to express the FoM as a function of the total SNR \mathcal{P} . The result is shown in figure 2, for $N = 1, 2, 3$, and 4.

The other summary statistics that we can employ is the grand sum of the relative Fisher matrix, eq. (3.6). We have, for the relative Fisher matrix of the auto- and cross-spectra:

$$\Xi = \sum_{[ij]} \sum_{[i'j']} \mathcal{F}_{[ij],[i'j']} = \frac{1}{2} V \tilde{V}_k \left(N - \frac{1}{1 + \mathcal{P}} \sum_{ij} \mathcal{P}_{ij} \right)^2. \quad (3.19)$$

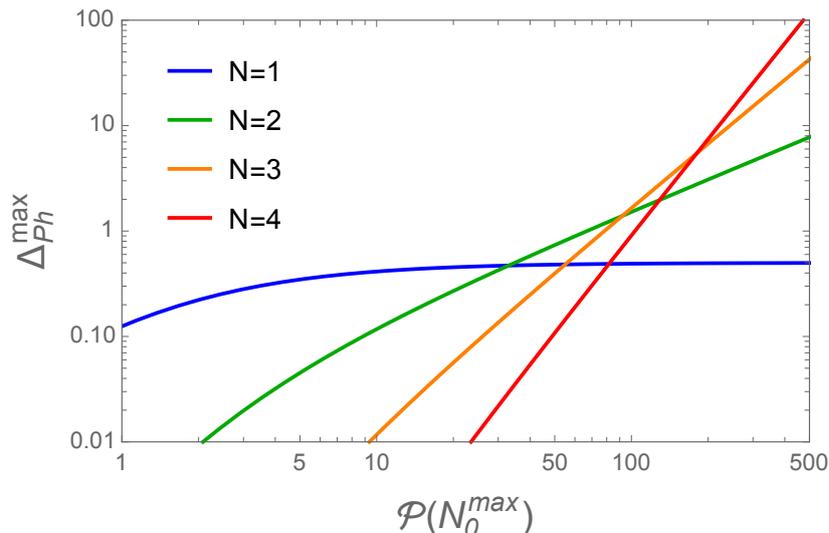


Figure 2. Figure of Merit (FoM) for the Fisher information at its maximum, Δ_{N, P_h}^{\max} , as a function of the total power \mathcal{P} , for 1 (blue), 2 (green), 3 (orange), and 4 (red) tracers. Here we assumed that \mathcal{P} does not vary significantly with N — see the text.

We can again extremize this functional assuming that $b_i^2 \sim \bar{n}_i^{-\gamma}$, subject to the constraint $\sum_i w_i = \sum_i \bar{n}_i / \bar{n}_T = 1$. The result is the same as before: the weights are all the same, meaning equipartition of the number densities and SNRs, $w_i = 1/N$ and $\mathcal{P}_{ij} = \mathcal{P}_{ii} = \mathcal{P}/N$. Substituting this extremal solution back into eq. (3.19) we obtain:

$$\begin{aligned} \Xi^{\max} &= \frac{1}{2} V \tilde{V}_k \left(N - \frac{1}{1 + \mathcal{P}} N^2 \frac{\mathcal{P}}{N} \right)^2 \\ &= \frac{1}{2} V \tilde{V}_k \frac{N^2}{(1 + \mathcal{P})^2}. \end{aligned} \quad (3.20)$$

Now, recall that $\mathcal{P} \sim N^\gamma$, with $0.1 \lesssim \gamma \lesssim 0.5$, and then it becomes clear that the grand sum Fisher information always grows when we split the tracers into more sub-types.

A caveat is in order with regards to the conclusions that were drawn above. The precise behaviours of the FoM and the grand sum Fisher information depend on the nature of the relationship between the number densities of the tracers, \bar{n}_i , and their biases, b_i . In the derivation above we assumed that $b_i^2 \sim \bar{n}_i^{-\gamma}$ for all tracers, with $0 < \gamma < 1$, regardless of how the tracers are split. But this is not strictly valid even for halos: at $z = 0$ the fit by Tinker for halo bias at low ($M_h \lesssim 10^{12} h^{-1} M_\odot$) and high ($M_h \gtrsim 10^{15} h^{-1} M_\odot$) masses yields $\gamma \sim 0.1$ – 0.3 , while in the intermediate mass range $\gamma \sim 0.4$ – 0.5 [21]. For galaxies, quasars and Ly- α systems the way in which we can break up the tracers into sub-types, according to luminosity, stellar mass or other properties, could be even less consistent with the hypothesis of a simple power-law relationship between number densities and biases. If that turns out to be the case in a given galaxy survey, then maximizing the information may be significantly more complex than in this simplified model.

4 The independent degrees of freedom: diagonalizing the Fisher matrix

If the cross-correlations do not carry any additional information, then, as shown by [7], the independent degrees of freedom which diagonalize the Fisher matrix of eq. (3.7) are given by

the total clustering power, \mathcal{P} , and ratios of their clusterings. In the case of two tracers, the diagonalized Fisher matrix for the degrees of freedom $\{\mathcal{P}, \log(\mathcal{P}_{11}/\mathcal{P}_{22})\}$ is given by:

$$F[\mathcal{P}, \log(\mathcal{P}_{11}/\mathcal{P}_{22})] = V\tilde{V}_k \begin{pmatrix} \frac{1}{2} \frac{1}{(1+\mathcal{P})^2} & 0 \\ 0 & \frac{1}{4} \frac{\mathcal{P}_{11}\mathcal{P}_{22}}{1+\mathcal{P}} \end{pmatrix}. \quad (4.1)$$

This expression makes it clear that, at least in the Gaussian approximation, measurements of the total clustering strength \mathcal{P} are limited by cosmic variance: even if the total clustering becomes arbitrarily large, $\mathcal{P} \rightarrow \infty$, its relative uncertainty is still limited by the volume of the survey and the Fourier-space volume of the bandwidth, $\sigma^2(\mathcal{P})/\mathcal{P}^2 \rightarrow 2/(V\tilde{V}_k)$. The ratios of spectra, on the other hand, can be measured with arbitrarily large accuracy (at least in principle), as long as we keep increasing the number densities of the two tracers: since the information regarding $\log(\mathcal{P}_{11}/\mathcal{P}_{22})$ depends not only on the phase space volume $V\tilde{V}_k$, but also on $\sim \bar{n}_1 \bar{n}_2 / \bar{n}_T$, the uncertainty in that parameter (the inverse of the Fisher matrix element) can be lowered by increasing the detection of the tracers, even for a fixed volume.

Here we stress the fact that it is unnecessary (and utterly unpractical) to try to measure directly the total clustering and ratios such as $\mathcal{P}_{11}/\mathcal{P}_{22}$ in order to rip the benefits implicit in the Fisher matrix of eq. (4.1): all that is needed is to measure the usual observables (e.g., P_{11} and P_{22}) and then project them into some chosen set of parameters — in this case, $\{\mathcal{P}, \log(\mathcal{P}_{11}/\mathcal{P}_{22})\}$. The fact that these degrees of freedom diagonalize the Fisher matrix are just the simplest manifestation of the fact that constraints on physical parameters that only depend on \mathcal{P} are subject to cosmic variance limitations which do not apply to physical parameters that depend only on the ratio ($\mathcal{P}_{11}/\mathcal{P}_{22}$).

Notice that it is only on very large scales that one can realistically expect that the auto- and cross-spectra are degenerate, $P_{ij}^2 = P_{ii}P_{jj}$. Indeed, as argued in the Introduction, on small scales the cross-spectra may carry information about additional physical dependencies that are not directly available through the auto-correlations. But more importantly, cross-correlations and cross-spectra constitute different observables, which can be estimated from the data in different ways, in order to optimize the amount of information that is extracted from the survey.

Therefore, the question arises as to what are the independent degrees of freedom when the cross-spectra are regarded as carrying irreducible degrees of freedom, which are not degenerate with the auto-spectra. To be specific, the problem we wish to solve, in the particular case of two tracers, is how to diagonalize the Fisher matrix of eq. (2.24). A straightforward calculation shows that the three degrees of freedom which diagonalize that Fisher matrix are:

$$\mathcal{Q}_1 = \mathcal{P} \left[1 + \frac{1}{2} \log \left(\frac{\mathcal{P}_{11}^2 + \mathcal{P}_{22}^2 + 2\mathcal{P}_{12}^2}{\mathcal{P}^2} \right) \right] \quad (4.2)$$

$$\mathcal{Q}_2 = \log \left(\frac{\mathcal{P}_{11}^2 + \mathcal{P}_{12}^2}{\mathcal{P}_{22}^2 + \mathcal{P}_{12}^2} \right) \quad (4.3)$$

$$\mathcal{Q}_3 = \frac{\mathcal{P}_{12}^2 - \mathcal{P}_{11}\mathcal{P}_{22}}{\mathcal{P}^2}. \quad (4.4)$$

It is clear that, in the limit that $\mathcal{P}_{12}^2 \rightarrow \mathcal{P}_{11}\mathcal{P}_{22}$, the first two degrees of freedom reduce to $\mathcal{Q}_1 \rightarrow \mathcal{P}$ and to $\mathcal{Q}_2 \rightarrow \log \mathcal{P}_{11}/\mathcal{P}_{22}$, while the third one effectively disappears, $\mathcal{Q}_3 \rightarrow 0$. For

that reason, it is useful to express the irreducible degree of freedom in the cross-spectrum in terms of an adimensional quantity ϵ_{12} :

$$\mathcal{P}_{12}^2 = \mathcal{P}_{11}\mathcal{P}_{22}(1 + \epsilon_{12}) \quad \Leftrightarrow \quad \mathcal{Q}_3 = \frac{\mathcal{P}_{11}\mathcal{P}_{22}}{\mathcal{P}^2} \epsilon_{12}. \quad (4.5)$$

Computing the Jacobian for the transformation from $\{\mathcal{P}_{11}, \mathcal{P}_{12}, \mathcal{P}_{22}\} \rightarrow \{\mathcal{Q}_1, \mathcal{Q}_2, \mathcal{Q}_3\}$ and using it to project the Fisher matrix of eq. (2.24) into the new degrees of freedom, yields the result:

$$F[\mathcal{Q}_i, \mathcal{Q}_j] = V\tilde{V}_k \begin{pmatrix} \frac{1}{2} \frac{1}{(1+\mathcal{P})^2} & 0 & 0 \\ 0 & \frac{1}{4} \frac{\mathcal{P}_{11}\mathcal{P}_{22}}{1+\mathcal{P}} & 0 \\ 0 & 0 & \frac{1}{2}(1+\mathcal{P})^2 \end{pmatrix}. \quad (4.6)$$

Notice that our *fiducial model* is such that $\mathcal{P}_{12}^2 \rightarrow \mathcal{P}_{11}\mathcal{P}_{22}$, or $\epsilon_{12} \rightarrow 0$, but we only take this limit *after* transforming to the new degrees of freedom.

Some facts are immediately clear: first, the two generalized degrees of freedom \mathcal{Q}_1 and \mathcal{Q}_2 , have the same Fisher information as the simplified degrees of freedom \mathcal{P} and $\log(\mathcal{P}_{11}/\mathcal{P}_{22})$ — see eq. (4.1). But more importantly, just as it happens with the relative clustering strength \mathcal{Q}_2 , the irreducible information in the cross-spectrum, encapsulated in \mathcal{Q}_3 , also have a signal-to-noise ratio (SNR) that grows arbitrarily with the number density of the tracers, i.e.:

$$\frac{\mathcal{Q}_3^2}{\sigma^2(\mathcal{Q}_3)} = \frac{V\tilde{V}_k}{2} (1+\mathcal{P})^2 \left(\frac{\mathcal{P}_{12}^2 - \mathcal{P}_{11}\mathcal{P}_{22}}{\mathcal{P}^2} \right)^2 = \frac{V\tilde{V}_k}{2} (1+\mathcal{P})^2 \left(\frac{\mathcal{P}_{11}\mathcal{P}_{22}}{\mathcal{P}^2} \right)^2 \epsilon_{12}^2. \quad (4.7)$$

Figure 3 shows the behavior of the SNRs of the three independent degrees of freedom in some typical scenarios.

Taking the high SNR limit, $\mathcal{P}_1 = \bar{n}_1 P_{11} \gg 1$ and $\mathcal{P}_2 = \bar{n}_2 P_{22} \gg 1$, we see that the accuracies in the measurements of the three independent variables scale as:

$$\frac{\mathcal{Q}_1^2}{\sigma^2(\mathcal{Q}_1)} \sim \left(\frac{\bar{n}_1 \bar{n}_2}{\bar{n}_1 + \bar{n}_2} \right)^0 \rightarrow \left(\frac{\bar{n}_T}{4} \right)^0 \quad (4.8)$$

$$\frac{\mathcal{Q}_2^2}{\sigma^2(\mathcal{Q}_2)} \sim \left(\frac{\bar{n}_1 \bar{n}_2}{\bar{n}_1 + \bar{n}_2} \right)^1 \rightarrow \left(\frac{\bar{n}_T}{4} \right)^1 \quad (4.9)$$

$$\frac{\mathcal{Q}_3^2}{\sigma^2(\mathcal{Q}_3)} \sim \left(\frac{\bar{n}_1 \bar{n}_2}{\bar{n}_1 + \bar{n}_2} \right)^2 \rightarrow \left(\frac{\bar{n}_T}{4} \right)^2, \quad (4.10)$$

where, for two tracers, $\bar{n}_T = \bar{n}_1 + \bar{n}_2$, and the expressions on the right-hand sides result from using the optimal “equipartition” configuration that maximizes the SNR, namely $\bar{n}_1 = \bar{n}_2 = \bar{n}_T/2$.

The first limit, eq. (4.8), expresses cosmic variance: the accuracy in measurements of the matter power spectrum is fundamentally limited by the volume where we measure the density perturbations and by the width of the bandpower, even if we can count on arbitrarily large numbers of tracers to determine the density field inside that particular volume. The second limit, eq. (4.9), means that ratios of power spectra of different tracers can be measured with an accuracy that is only limited by the numbers of tracers that we have, and scale with the number density [7]. Finally, the third limit, eq. (4.10), tells us that the irreducible degrees of freedom of the cross-correlations scale even faster with the tracer densities, compared with

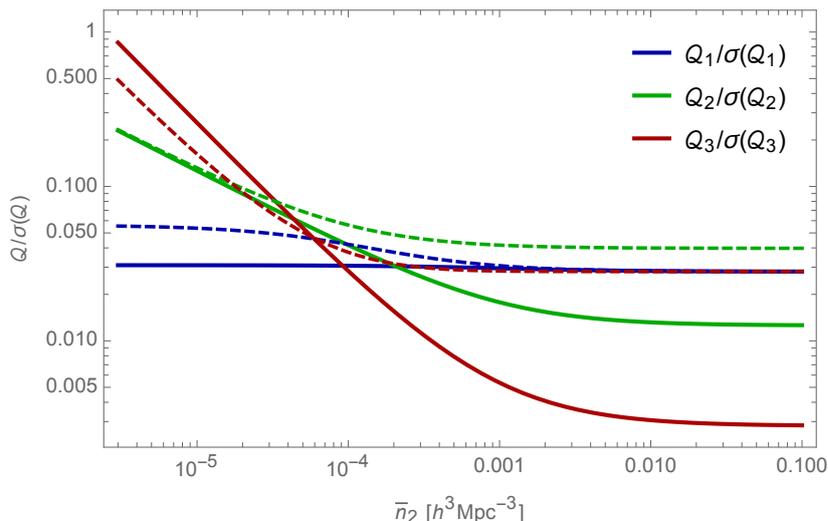


Figure 3. Signal-to-noise ratios $Q_i/\sigma(Q_i)$, for the three independent degrees of freedom in the case of two tracers — see eqs. (4.2)–(4.4). The blue, green and red curves correspond to the SNRs of Q_1 , Q_2 and Q_3 . The thin, dashed lines correspond to the scenario with $\bar{n}_1 = 10^{-4} h^3 \text{Mpc}^{-3}$, while the thick, solid lines correspond to the case where $\bar{n}_1 = 10^{-3} h^3 \text{Mpc}^{-3}$, both as a function of \bar{n}_2 . For these examples we considered a volume $V = 10^9 h^{-3} \text{Mpc}^3$, a spherical bandpower at $k = (0.1 \pm 0.0025) h \text{Mpc}^{-1}$, and a value of the power spectrum of $P^{(m)}(k) = 10^4 h^{-3} \text{Mpc}^3$.

the ratios of tracers. Therefore, in the limit of high number of tracers, these degrees of freedom can be determined with extremely high accuracy, and are even less constrained by cosmic variance.

On the other hand, in the limit of small number densities, the situation is reversed. In that case the Fisher matrix of eq. (4.6) tells us that the SNR of the degrees of freedom scale as $(\bar{n}_1 + \bar{n}_2)^2$ for Q_1 , as $\bar{n}_1 \bar{n}_2$ for Q_2 , and as $\bar{n}_1^2 \bar{n}_2^2 / (\bar{n}_1 + \bar{n}_2)^4$ for Q_3 . Therefore, in the limit of very sparse tracers it becomes increasingly difficult to measure the ratios of spectra, and even harder to determine the irreducible degrees of freedom in the cross-spectra.

These results put the issue of the cancellation of cosmic variance into a new light: some degrees of freedom profit even more from a denser sampling than others. Ratios of spectra (green lines in figure 3), which allow us to measure, e.g., redshift-space distortions and primordial non-Gaussianities, as well as many parameters in the bias expansion [15], start to become more interesting when the numbers of tracers are larger than $\bar{n} \gtrsim 10^{-4} h^3 \text{Mpc}^{-3}$, but they saturate at a SNR ~ 1 –2% unless we have both tracers with number densities $\bar{n} \gtrsim 10^{-2} h^3 \text{Mpc}^{-3}$, which doesn't seem reasonable. On the other hand, the independent degrees of freedom in the cross-spectrum (red lines in figure 3) start to become detectable for $\bar{n} \gtrsim 10^{-4} h^3 \text{Mpc}^{-3}$, but their accuracy grows even faster with the number density, such that we can reasonably achieve accuracies of $\sim 0.3\%$ when the two tracers have number densities of $\bar{n} \sim 10^{-3} h^3 \text{Mpc}^{-3}$. We again call the attention to the fact that these results apply regardless of whether the observables measured from the data are the degrees of freedom of eqs. (4.2)–(4.4) or the usual auto- and cross-spectra. Indeed, it would be utterly unpractical to measure directly these combinations and ratios, and besides, the auto- and cross-spectra are already imbued with the properties expressed by the Fisher matrix of eq. (4.6).

Finally, we note that we were unable to determine an analytic expression that generalizes to N tracers the independent degrees of freedom Q_i when we include the cross-spectra.

Nevertheless, a numerical study of the eigenvalues of the Fisher matrix seems to indicate that the irreducible degrees of freedom in the cross-spectra indeed scale in the manner shown in eq. (4.10). We will return to this issue in a future paper.

5 Conclusions

This paper conveys two main results. Firstly, we showed how to use two summary statistics of the Fisher matrix in order to optimize the number of tracer species in a survey. We employed the following proxies for the total Fisher information: the determinant (or figure of Merit), and the grand sum of the Fisher matrix. By assuming a simple power-law relation between bias and number density we have shown that using either one of these two summary statistics, with or without the explicit inclusion of the independent degrees of freedom for the cross-spectra, the configuration that optimizes the total Fisher information is that in which the tracers are divided into equal samples: $\bar{n}_i = \bar{n}_T/N$, where $\bar{n}_T = \sum_{i=1}^N \bar{n}_i$ is the total number of tracers in the survey. Moreover, in general, the higher the total clustering SNR $\mathcal{P} = \sum_i \bar{n}_i P_{ii}$, the higher is the optimal number of tracers that one should employ in order to maximize that Fisher information.

We have also shown in this paper that the information in the irreducible degrees of freedom of the cross-spectra, expressed in terms of $\mathcal{P}_{ij}^2 - \mathcal{P}_{ii}\mathcal{P}_{jj}$, can be measured with an accuracy that is not limited by cosmic variance. This result is analogous to the well-known fact that ratios of spectra are partially immune to cosmic variance, however, while the accuracy of spectral ratios increase with $\sim \sqrt{\bar{n}_T}$, the accuracy of the degrees of freedom of the cross-spectra grow with $\sim \bar{n}_T$. This means that physical parameters which are manifested in the combination $\mathcal{P}_{ij}^2 - \mathcal{P}_{ii}\mathcal{P}_{jj}$ can be measured, at least in principle, to exquisite precision. In particular, stochastic and non-linear terms in the perturbative bias expansion, such as those shown in eq. (1.4), can benefit from this windfall of the multi-tracer analysis.

Acknowledgments

We thank Henrique Rubira, Thiago Mergulhão and Rodrigo Voivodic for useful comments. We also acknowledge the financial support of FAPESP (R.A.), CNPq (R.A & I.L.T.) and CAPES (J.V.D.F.).

References

- [1] DES collaboration, *The dark energy survey*, [astro-ph/0510346](#) [INSPIRE].
- [2] EUCLID THEORY Working Group, *Cosmology and fundamental physics with the Euclid satellite*, *Living Rev. Rel.* **16** (2013) 6 [[arXiv:1206.1225](#)] [INSPIRE].
- [3] LSST collaboration, *LSST: from Science Drivers to Reference Design and Anticipated Data Products*, *Astrophys. J.* **873** (2019) 111 [[arXiv:0805.2366](#)] [INSPIRE].
- [4] J-PAS collaboration, *J-PAS: The Javalambre-Physics of the Accelerated Universe Astrophysical Survey*, [arXiv:1403.5237](#) [INSPIRE].
- [5] U. Seljak, *Extracting primordial non-Gaussianity without cosmic variance*, *Phys. Rev. Lett.* **102** (2009) 021302 [[arXiv:0807.1770](#)] [INSPIRE].
- [6] P. McDonald and U. Seljak, *How to measure redshift-space distortions without sample variance*, *JCAP* **10** (2009) 007 [[arXiv:0810.0323](#)] [INSPIRE].

- [7] L.R. Abramo and K.E. Leonard, *Why multi-tracer surveys beat cosmic variance*, *Mon. Not. Roy. Astron. Soc.* **432** (2013) 318 [[arXiv:1302.5444](#)] [[INSPIRE](#)].
- [8] V. Desjacques, D. Jeong and F. Schmidt, *Large-Scale Galaxy Bias*, *Phys. Rept.* **733** (2018) 1 [[arXiv:1611.09787](#)] [[INSPIRE](#)].
- [9] T. Baldauf, U. Seljak, R.E. Smith, N. Hamaus and V. Desjacques, *Halo stochasticity from exclusion and nonlinear clustering*, *Phys. Rev. D* **88** (2013) 083507 [[arXiv:1305.2917](#)] [[INSPIRE](#)].
- [10] V. Assassi, D. Baumann, D. Green and M. Zaldarriaga, *Renormalized Halo Bias*, *JCAP* **08** (2014) 056 [[arXiv:1402.5916](#)] [[INSPIRE](#)].
- [11] M. Mirbabayi, F. Schmidt and M. Zaldarriaga, *Biased Tracers and Time Evolution*, *JCAP* **07** (2015) 030 [[arXiv:1412.5169](#)] [[INSPIRE](#)].
- [12] M. Schmittfull, M. Simonović, V. Assassi and M. Zaldarriaga, *Modeling Biased Tracers at the Field Level*, *Phys. Rev. D* **100** (2019) 043514 [[arXiv:1811.10640](#)] [[INSPIRE](#)].
- [13] H. Gil-Marín, C. Wagner, L. Verde, R. Jimenez and A.F. Heavens, *Reducing sample variance: halo biasing, non-linearity and stochasticity*, *Mon. Not. Roy. Astron. Soc.* **407** (2010) 772 [[arXiv:1003.3238](#)] [[INSPIRE](#)].
- [14] D. Ginzburg and V. Desjacques, *Shot noise in multitracer constraints on f_{NL} and relativistic projections: Power spectrum*, *Mon. Not. Roy. Astron. Soc.* **495** (2020) 932 [[arXiv:1911.11701](#)] [[INSPIRE](#)].
- [15] T. Mergulhão, H. Rubira, R. Voivodic and L.R. Abramo, *The Effective Field Theory of Large-Scale Structure and Multi-tracer*, [arXiv:2108.11363](#) [[INSPIRE](#)].
- [16] M. Tegmark, A. Taylor and A. Heavens, *Karhunen-Loeve eigenvalue problems in cosmology: How should we tackle large data sets?*, *Astrophys. J.* **480** (1997) 22 [[astro-ph/9603021](#)] [[INSPIRE](#)].
- [17] J.M. Bernardo and A.F.M. Smith, *Bayesian Theory*, John Wiley & Sons (1994).
- [18] H. Jeffreys, *An invariant form for the prior probability in estimation problems*, *Proc. Roy. Soc. Lond. A* **186** (1946) 453.
- [19] A. Albrecht et al., *Report of the Dark Energy Task Force*, [astro-ph/0609591](#) [[INSPIRE](#)].
- [20] L.R. Abramo, *The full Fisher matrix for galaxy surveys*, *Mon. Not. Roy. Astron. Soc.* **420** (2012) 2042 [[arXiv:1108.5449](#)] [[INSPIRE](#)].
- [21] J.L. Tinker et al., *Toward a halo mass function for precision cosmology: The Limits of universality*, *Astrophys. J.* **688** (2008) 709 [[arXiv:0803.2706](#)] [[INSPIRE](#)].
- [22] T. Lazeyras, C. Wagner, T. Baldauf and F. Schmidt, *Precision measurement of the local bias of dark matter halos*, *JCAP* **02** (2016) 018 [[arXiv:1511.01096](#)] [[INSPIRE](#)].
- [23] J. Carron, *On the assumption of Gaussianity for cosmological two-point statistics and parameter dependent covariance matrices*, *Astron. Astrophys.* **551** (2013) A88 [[arXiv:1204.4724](#)] [[INSPIRE](#)].