A Computationally Efficient Approach for Calculating Galaxy Two-Point Correlations

Regina Demina,1∗, Sanha Cheong,1†, Segev BenZvi,1‡ Otto Hindrichs,1§

1Department of Physics and Astronomy, University of Rochester, 500 Joseph C. Wilson Boulevard, Rochester, NY 14627, USA

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

We develop a modification to the calculation of the two-point correlation function commonly used in the analysis of large scale structure in cosmology. An estimator of the two-point correlation function is constructed by contrasting the observed distribution of galaxies with that of a uniformly populated random catalog. Using the assumption that the distribution of random galaxies in redshift is independent of angular position allows us to replace pairwise combinatorics with fast integration over probability maps. The new method significantly reduces the computation time while simultaneously increasing the precision of the calculation.

Key words: large-scale structure of Universe – distance scale – dark energy – surveys – galaxies: statistics – methods: data analysis

1 INTRODUCTION

The two-point spatial correlation function (2pcf) is a key tool used in astrophysics for the study of large scale structure. Given an object such as a galaxy, the probability of finding another galaxy within a volume element $\delta V$ at a distance $s$ from the first galaxy is

$$\delta P = \pi(1 + \xi(s))\delta V,$$

where $\pi$ is the mean number density and $\xi(s)$ is the two-point correlation function characterizing the deviation from the uniform distribution in separation between galaxies (Peebles 1973).

The correlation function $\xi(s)$ and its Fourier transform, the galaxy power spectrum $P(k)$, have been used to describe the distribution of matter in galaxy surveys (Yu & Peebles 1969; Peebles & Hauser 1974; Groth & Peebles 1977; Feldman et al. 1994). During the past decade $\xi(s)$ has become a popular tool for the reconstruction of the clustering signal known as baryon acoustic oscillations, or BAO (Eisenstein et al. 2005). The BAO signal is the signature of the density differences that arose in the early universe before the thermal decoupling of photons and baryons (Sunyaev & Zeldovich 1970; Peebles & Yu 1970). It is detectable today as a characteristic peak in the galaxy spatial correlation function at roughly $s = 110h^{-1}$ Mpc, where $h$ defines the Hubble parameter $H_0 = 100h$ km s$^{-1}$ Mpc$^{-1}$.

Given a spectroscopic survey containing the 3D coordinates of each galaxy, there exist several possible ways to estimate $\xi(s)$. The most popular estimator is due to Landy & Szalay (1993), which is constructed by combining pairs of galaxies from a catalog of observed objects $D$ (“data”) and a randomly generated catalog $R$ with galaxies distributed uniformly over the fiducial volume of the survey but using the same selection function as the data. The Landy-Szalay (LS) estimator is

$$\hat{\xi}(s) = \frac{DD(s) - 2DR(s) + RR(s)}{RR(s)},$$

where $DD$, $RR$, and $DR$ are the normalized distributions of the pairwise combinations of galaxies from the data and random catalogs (plus cross terms) at a given distance $s$ from each other.

There exist many 2pcf estimators other than the LS estimator, and they have different advantages and limitations. However, all commonly known estimators are functions of $DR$ and/or $RR$ distributions (see Hamilton (1993); Kerscher et al. (2000); Vargas-Magana et al. (2013)) and therefore depend on the random catalog $R$. To limit statistical fluctuations in $\hat{\xi}(s)$, it is typical to generate random catalogs with one to two orders of magnitude more galaxies than the survey under investigation. Unfortunately, the “brute-force” computation of $\hat{\xi}(s)$, in which all possible pair combinations are counted, is an $O(N^2)$ calculation due to $RR$, where $N$ is the number of galaxies in the random catalog $R$; this results in a trade-off between statistical uncertainties and computation time. This trade-off has consequences beyond reducing uncertainties. For example, researchers often simulate a va-
riety of cosmological parameters when studying the BAO, but the computational overhead required to calculate \( \xi(s) \) may limit the number and type of possible analyses that can be carried out. The overhead also increases the effort needed to compute the covariance of the 2pcf.

In this paper, we suggest a method which substantially reduces the time needed to compute \( RR \) and \( DR \) and hence is applicable to any estimator \( \xi(s) \). The method is based on the assumption that the probability for a galaxy to be observed at a particular location in the random catalog can be factorized into separate angular and redshift components (Ross et al. 2012; Anderson et al. 2013). This assumption allows us to replace pairwise combinatorics in the calculations of \( RR \) and \( DR \) with fast integration over probability maps. We find that the estimation of \( \xi(s) \) speeds up by factors > 100 with respect to the brute-force calculation. In practical terms, this means an estimate of \( \xi(s) \) which is typically carried out on large computing clusters can be performed on a modern notebook computer.

The paper is structured as follows. In Section 2 we describe the mathematical justification of the calculation for random-random (\( RR \)) and data-random (\( DR \)) galaxy pairs. In Section 3 we describe the resulting algorithm used to compute \( \xi(s) \) in detail. The performance of the algorithm is studied in Section 4 using mock catalogs and spectroscopic data from the CMASS portion of the Sloan Digital Sky Survey (SDSS) DR9 catalog (Anderson et al. 2013; Ross et al. 2012; Sanchez et al. 2012; Percival et al. 2014). We then conclude in Section 5.

## 2 MATHEMATICAL PROOF

### 2.1 Random-random distribution

The Hubble distance is defined as \( D_H = c/H_0 \), where \( c \) is the speed of light. Given a set of cosmological parameters, the comoving distance of a given galaxy from the Earth, \( D_C \), can be calculated from its observed redshift \( z \) as

\[
D_C(z) = D_H \int_0^z dz' \left( \Omega_M (1 + z')^3 + \Omega_k (1 + z')^2 + \Omega_\Lambda \right)^{-1/2},
\]

where \( \Omega_M \) is the present day relative matter density of the universe, \( \Omega_k \) measures the curvature of space, and \( \Omega_\Lambda \) is the relative density due to the cosmological constant.

The transverse comoving distance \( D_M \) is defined as:

\[
D_M(z) = \begin{cases} 
D_M/H_{\Omega_k} \sinh(\sqrt{\Omega_\Lambda} D_C(z)/D_H), & \text{for } \Omega_k > 0 \\
D_C(z), & \text{for } \Omega_k = 0 \\
D_M/H_{\Omega_k} \sin(\sqrt{\Omega_\Lambda} D_C(z)/D_H), & \text{for } \Omega_k < 0
\end{cases}
\]

As suggested by current cosmological constraints (see e.g. Aubourg et al. (2015); Ade et al. (2016)), \( \Omega_k \) is small and thus eq. 4 can be approximated by

\[
D_M(z) = D_C(z) (1 + \kappa), \quad \kappa = \Omega_k (D_C/D_H)^2.
\]

The comoving coordinates \( \vec{r} \) of any galaxy can be described by its right ascension \( \alpha \), declination \( \delta \), and comoving radial distance \( r = D_C(z) \). In the case of small but non-zero \( \Omega_k \), the distance \( s_{12} \) between two points with coordinates \( r_1 \) and \( r_2 \) (illustrated in Fig. 1) can be approximated by:

\[
s_{12} = \sqrt{\sigma_{12}^2 + \pi_{12}^2},
\]

where \( \sigma_{12} \) and \( \pi_{12} \), the distances transverse and parallel to the line of sight (LOS) respectively, are defined as:

\[
\sigma_{12} = (1 + \kappa)(r_1 + r_2) \sin \frac{\theta_{12}}{2},
\]

\[
\pi_{12} = |r_1 - r_2| \cos \frac{\theta_{12}}{2}.
\]

The angular separation \( \theta \) between the two galaxies is given by

\[
\theta_{12} = \cos^{-1} \left( \cos \delta_1 \cos \delta_2 \cos (\alpha_1 - \alpha_2) + \sin \delta_1 \sin \delta_2 \right).
\]

Since random catalogs are typically generated by uniformly populating the fiducial volume of the survey, the galaxy distribution over \( z \), and thus over \( r \), is factorizable from the angular distribution. In other words, any angular region of the sky has the same distribution of galaxies in \( r \) (see Ross et al. (2012) and Fig. 2). This means that the expected count of random galaxies, \( R(\vec{r}) \), can be factorized into the product of the expected count \( R_{\text{ang}}(\alpha, \delta) \) at a given angular position and a radial probability density function (PDF) \( P_r(r) \):

\[
R(\vec{r}) = R_{\text{ang}}(\alpha, \delta) P_r(r).
\]

\( R(\vec{r}) \) evaluated this way has smaller statistical uncertainty since the precision of \( R_{\text{ang}}(\alpha, \delta) \) depends on the number of points in the 2D angular cell, and the precision of \( P_r(r) \) depends on the number of points in the 1D range in \( r \). In
contrast, the statistical precision without the factorizability assumption is determined by the number of points in a 3D cell \((r, \alpha, \delta)\). We will show in Section 4 that the uncertainties in \(\xi(s)\) evaluated using the suggested method are indeed smaller. The PDF of random-random galaxy pairs separated by a distance \(s\) can be expressed as:

\[
RR(s) = \frac{1}{2N_{RR}} \int d\Omega_1 d\Omega_2 d\theta d\phi d\delta \frac{R_{\text{ang}}(\alpha_1, \delta_1) P_r(r_1)}{2\pi R^3} \times R_{\text{ang}}(\alpha_2, \delta_2) P_r(r_2) \delta(s - s_{12}),
\]

where \(\Omega\) represents a solid angle (and \(d\Omega = \cos \delta \, d\phi \, d\alpha\)) and \(N_{RR}\) is used to normalize the random-random distribution. For the unweighted calculation (all galaxies have weight 1), the normalization constant is simply \(N_{RR} = N_R (N_R - 1)/2\) where \(N_R\) is the total number of galaxies in the random catalog. However, for the weighted calculation, the normalization constant \(N_{RR}\) is:

\[
N_{RR} = \sum_{i=1}^{N_R} \sum_{j=1}^{N_R} w_i^R w_j^R
\]

where \(w_i^R\) is the weight of the \(i^{th}\) galaxy in the random catalog. In this specific form, the calculation is \(O(N_R^2)\) and requires a double loop over indices. However, it can be rewritten as:

\[
N_{RR} = \frac{1}{2} \left[ \left( \sum_{i=1}^{N_R} w_i^R \right)^2 - \sum_{i=1}^{N_R} \sum_{j=1}^{N_R} w_i^R w_j^R \right]
\]

which is now an \(O(N_R)\)-calculation. The same approach can be applied when normalizing the \(DD\) histogram.

In eq. 11, the integral is taken over the entire fiducial volume of the random catalog \(R\), and the Dirac-\(\delta\) function is introduced to ensure that the distance between two galaxies \(s_{12}\) is equal to the distance of interest \(s\). To isolate the angular variables, we can write the \(\delta\) function in eq. 11 as an integral of the product of two \(\delta\) functions:

\[
\delta(s - s_{12}) = \int du \delta(s - s_{12}) \delta(\theta - \theta_{12}).
\]

Note that the second \(\delta\) function is independent of the radial positions \(r_1, r_2\) of the galaxies. Thus, \(RR\) can be rewritten as

\[
RR(s) = \frac{1}{N_{RR}} \int dr_1 dr_2 d\theta P_r(r_1) P_r(r_2) f(\theta) \times \delta(s - s_{12}),
\]

where

\[
f(\theta) = \frac{1}{2} \int d\Omega_1 d\Omega_2 \frac{R_{\text{ang}}(\alpha_1, \delta_1) R_{\text{ang}}(\alpha_2, \delta_2)}{2\pi R^3} \times \delta(\theta - \theta_{12})
\]

is the count of \(RR\) galaxy pairs whose angular separation is \(\theta\).

The probability distribution of random galaxy pairs is constructed in two steps: first, histogramming, where we construct the distribution \(f(\theta)\) over angular separation using the count of random galaxies \(R_{\text{ang}}(\alpha, \delta)\) according to eq. 16, and then integration, where we convolve the angular distribution with the radial PDF \(P_r(r)\) to obtain the distribution of random pairs over \(s\).

### 2.2 Data-random and data-data correlations

Unlike in the random catalog, in the data catalog the distribution of galaxies over \(r\) cannot be factorized from their angular distribution. Thus, the probability distribution over the opening angle between galaxies in the data and random catalogs is also a function of \(r_1\), the position of the data galaxy:

\[
g(\theta, r_1) = \int d\Omega_1 d\Omega_2 D(r_1, \alpha_1, \delta_1) R_{\text{ang}}(\alpha_2, \delta_2) \times \delta(\theta - \theta_{12}).
\]

The \(DR\) distribution can then be calculated as

\[
DR(s) = \frac{1}{N_{DR}} \int dr_1 dr_2 d\theta P_r(r_2) g(\theta, r_1) \times \delta(s - s_{12}),
\]

where \(N_{DR}\) is used to normalize the data-random distribution. For the unweighted calculation, the normalization constant is simply \(N_{DR} = N_D \times N_R\) where \(N_D\) is the total number of galaxies in the data catalog. For the weighted calculation, the normalization constant \(N_{DR}\) is:

\[
N_{DR} = \sum_{i=1}^{N_D} \sum_{j=1}^{N_R} w_i^D w_j^R
\]

where \(w_i^D\) is the weight of the \(i^{th}\) galaxy in the data catalog. The normalization constant \(N_{DR}\) only requires an \(O(N_R)\) computation and hence can be computed quickly even in the weighted calculation.

Finally, the data-data count is estimated using a brute-force iteration over all possible data galaxy pairs.

### 2.3 Generalization of the 2pcf to anisotropic case

The discussion up to this point has dealt with only a spherically symmetric 2pcf. However, the algorithm is easily generalized to study anisotropy in the 2pcf (Davis & Peebles 1983).

The histogramming stage proceeds in the same fashion as for the isotropic case. At the integration stage, we compute the distances transverse and parallel to LOS according to eqs. 7 and 8 and fill in \(RR, DR\) and \(DD\) histograms in two dimensions (\(\sigma, \pi\)). The BAO signal is expected to manifest itself as an ellipse in this 2D histogram. In the isotropic case the ellipse is reduced to a circle.

### 3 Description of the Algorithm

We compute \(DD\) by summing over all pairs of galaxies \(i\) and \(j\) separated by the distance \(s\), with each pair weighted by the product of the weights of individual galaxies \(w_i^D \times w_j^D\) in the survey. To data we apply the weights according to the prescription of Ross et al. (2012) to deal with the issues of close-pair corrections (\(w_{\text{cp}}\)), redshift-failure corrections (\(w_{\text{rf}}\)), systematic targeting effects (\(w_{\text{sys}}\)) and shot noise and cosmic variance (\(w_{\text{PKP}}\)) (Feldman et al. 1994) such that the total weight of a given data galaxy is:

\[
w^D = w_{\text{PKP}} \cdot w_{\text{sys}} \cdot (w_{\text{rf}} + w_{\text{cp}} - 1).
\]
Figure 3. *Top:* Distribution of the DR9 random catalog in $(\alpha, \delta)$ in the northern sky. *Middle:* the DR9 random catalog in the southern sky. *Bottom:* a zoomed-in view of northern sky. The positions of galaxies observed in the DR9 survey are plotted as white stars.

**RR** and **DR** are calculated using finely binned probability densities in $(\alpha, \delta)$ and $r$ defined from the random catalog. The choice of the bin sizes is important and is determined by the final bin size, $\Delta s$, desired in $\xi(s)$. In practice, this means that the bin sizes in $\alpha$ and $\delta$ must be smaller than the angle subtended by $\Delta s$ at the outermost radius of the data set by a factor greater than 4. The bin size in $r$ should also be smaller than $\Delta s$ by the same factor. Examples of binned angular histograms are shown in Fig. 3. It is clear from the zoomed up view that the statistical fluctuations, determined by the finite size of the random catalog are significant. However, it is important to note that these fluctuations are much smaller than the fluctuations in 3D cells with full coordinates ($r, \alpha, \delta$), which is effectively what is used in the brute-force approach, or in any other approach that relies on 3D number density distribution. In this paper, we generate the 2D angular and 1D radial probability maps based on an existing random catalog (published along with the observational data), but ideally these maps should be generated directly based on the observational conditions. This approach will create smoother probability maps and minimize the shot noise associated with the use of random catalogs.

Once the bin sizes are chosen, the algorithm proceeds as follows. First, using the random catalog, we produce the histograms $R_{\text{ang}}(\alpha, \delta)$ and $P_{\text{r}}(r)$. We note that each galaxy in the random catalog has an $r$-dependent weight $w^{R}(r)$, and these weights are used to fill the histograms: for each galaxy, $P_{\text{r}}(r)$ is incremented by $w^{R}(r)/N_{\text{R}}$, and $R_{\text{ang}}(\alpha, \delta)$ is incremented by 1. This can be generalized to the case where the weights also depend on angular position and are factorizable into angular and radial weights: $w^{R}(\alpha, \delta)$. Then $R_{\text{ang}}(\alpha, \delta)$ must be incremented by $w^{R}(\alpha, \delta)$.

At the next step, we compute the distribution $f(\theta)$ using $R_{\text{ang}}(\alpha, \delta)$. To compute $f(\theta)$, we consider all non-repeating pairs of angular cells ($\alpha_{1}, \delta_{1}$) and ($\alpha_{2}, \delta_{2}$), and calculate $\theta_{12}$ between them using the angular positions of the barycenters of the two cells using eq. 9. According to eq. 16, $f(\theta)$ is incremented by $R_{\text{ang}}(\alpha_{1}, \delta_{1})R_{\text{ang}}(\alpha_{2}, \delta_{2})$ when the two cells are different and by $R_{\text{ang}}(\alpha_{1}, \delta_{1})R_{\text{ang}}(\alpha_{2}, \delta_{2})/2$ when the two cells are identical.

Then, we calculate the distribution $g(r_{1}, \theta)$ using galaxies from the data catalog and $R_{\text{ang}}(\alpha, \delta)$. Here, we consider every pair of a data galaxy and an angular cell of $R_{\text{ang}}$. The angular separation $\theta_{12}$ is calculated using the position ($\alpha_{1}, \delta_{1}$) of the data galaxy and the angular position ($\alpha_{2}, \delta_{2}$) of the barycenter of the $R_{\text{ang}}$ cell. For each such pair, the distribution $g(r_{1}, \theta)$ is incremented by $w^{D}R_{\text{ang}}(\alpha_{2}, \delta_{2})$, where $w^{D}$ is the weight of the data galaxy.

Next, we perform an integration over distances $r_{1}$ and $r_{2}$ and produce the distributions **RR** and **DR**. This is achieved in three nested loops iterating over $\theta$, $r_{1}$ and $r_{2}$. Given these three variables, the distance separation $s_{12}$ is calculated using eq. 6. The distribution **RR** is incremented by $f(\theta)P_{\text{r}}(r_{1})P_{\text{r}}(r_{2})$ according to eq. 15, while **DR** is incremented by $g(r_{1}, \theta)P_{\text{r}}(r_{2})$ according to eq. 18. Then, the two histograms are normalized according to eqs. 13 and 19 respectively.

For the computation of the anisotropic 2pcf distances $\sigma_{12}$ and $\pi_{12}$ are computed using $\theta$, $r_{1}$ and $r_{2}$ according to eq. 7 and eq. 8 respectively. The distributions **RR**($\sigma, \pi$) and **DR**($\sigma, \pi$) are incremented by $f(\theta)P_{\text{r}}(r_{1})P_{\text{r}}(r_{2})$ and $g(r_{1}, \theta)P_{\text{r}}(r_{2})$ respectively. These histograms are normalized similarly.

In the calculation presented in this paper, we only consider galaxy pairs separated by a distance less than a certain maximum distance scale of interest $l_{\text{max}}$. For the estimate of **DD**, the whole space is divided into cubes with an edge length $l_{\text{max}}$. Each data galaxy is only paired with other galaxies within the same cube or the neighboring cubes. For **RR** and **DR**, the ($\alpha, \delta$) plane is divided into rectangular regions with angular size subtended by $l_{\text{max}}$ at the smallest value of $\rho$. The algorithm proceeds to calculate $f(\theta)$ and $g(r_{1}, \theta)$ only within one angular region and its neighbors. Finally, the LS estimator is calculated according to eq. 2. All the other estimators can be calculated just as easily, based on **RR** and **DR** distributions computed above.

MNRAS 000, 1–6 (2016)
4 PERFORMANCE OF THE ALGORITHM

The performance of the new algorithm is evaluated by calculating the runtime and the uncertainty of the LS estimator of the 2pcf.

Though the method described in this paper can be applied under any cosmology, for certainty we assume a $\Lambda$CDM+GR flat cosmology with parameters consistent with those used in the analysis of the SDSS-III DR9 data set (Anderson et al. 2013), i.e. $H_0 = 70$ km s$^{-1}$ Mpc$^{-1}$, $\Omega_M = 0.274$, $\Omega_L = 0.726$, and $\Omega_k = 0$.

We choose the following algorithm settings. The maximum galaxy pair separation $l_{\text{max}}$ is $400 \, h^{-1} \, \text{Mpc}$. The number of angular bins in the $(\alpha, \delta)$ plane is $N_\alpha \times N_\delta = 3500 \times 1600$, corresponding to an angular size of 1 mrad, or a resolved separation of galaxies of $2.55 \times 10^4 \, r_{\text{outermost}}$, where $r_{\text{outermost}}$ is the outermost radius. The number of bins in $\theta$ is $N_\theta = 3142$, corresponding to an angular size of 1 mrad. The number of bins in $r$ is $N_r = 900$, corresponding to a bin size of $1 \, h^{-1} \, \text{Mpc}$.

The running time scales as $(N_\alpha \times N_\delta)^3$ for the part that fills the histogram in $\theta$, plus $N_\theta \times N_\delta^2$ for the integration. The number of bins scales inversely with the bin size in the final histogram, $\Delta s$, and linearly with the distance scale of interest $l_{\text{max}}$. The time performance of the histogramming scales as the number of 2D cells squared or $O\left((l_{\text{max}}/\Delta s)^3\right)$. The integration time scales as the number of bins in $\theta$ times the number of bins in $r$ squared or $O\left((l_{\text{max}}/\Delta s)^5\right)$. Thus, the histogramming part has the dominant contribution to the computational time. Other fast computational techniques based on the number density rather than the actual permutation counting rely on data distribution in 3D cells (Moore et al. 2001) and thus have stronger scaling with the number of cells.

Usually, the calculations of $RR$ and $DR$ are the most computationally expensive since the size of the random catalog is typically much larger than the size of the data catalog. Therefore, we report the results of $RR$ and $DR$ for different sizes of random catalogs, while we do not change the data catalog.

To generate random catalogs of any desired size, we begin with the northern sky of the existing SDSS-III DR9 random catalog which contains $\approx 3.5 \times 10^5$ galaxies and extract the distributions $R_{\text{ang}}(\alpha, \delta)$ and $P_s(r)$. We then randomly generate new galaxies according to the product of these distributions. While the new catalogs may amplify existing statistical fluctuations in the DR9 random catalog, our interest is in testing the runtime of the algorithms and estimating the variance in $\xi(s)$ as a function of random catalog size. Therefore, biases introduced by the generation procedure are unimportant.

For random catalog sizes of $N_R = 1\text{M}$, $3\text{M}$, and $10\text{M}$ galaxies, we generate 20 realizations each and calculate $RR$ and $DR$ using the brute-force technique and the fast algorithm. We estimate the root mean square (RMS) of $\xi(s)$ using the 20 random catalogs and present it in Fig. 4. Note that though the bin size used in computation is small as specified above, the result is plotted with a much larger bin size, simply for visual purposes. In both the brute-force and fast algorithms, the uncertainties in $\xi(s)$ decrease as the size of the random catalog increases, but the uncertainties in the results based on the fast algorithm are smaller than those in the respective brute-force results, in agreement with the qualitative argument presented in Section 2.

Before discussing the runtime, it is worth noting that the brute-force method allows for parallelization (Alonso 2012), as does the suggested fast method. Hence, the runtime is measured in CPU hours rather than elapsed wall-clock time. As is done in the fast algorithm, the brute-force pair counting was done using a grid scheme as suggested in (Alonso 2012) up to a maximum distance of $l_{\text{max}} = 400 \, h^{-1} \, \text{Mpc}$. Hence, the two sets of results include the same amount of statistics.

In Fig. 4, we show the runtime (in CPU hours) of the fast and brute-force calculations of $\xi(s)$ as a function of the size of the random catalog.

![Figure 4](image-url)  
**Figure 4.** Uncertainties in $\xi(s)$ as a function of the size of the random catalog. The uncertainties are calculated as the RMS of $\xi(s)$ obtained with 20 catalogs generated for each of three catalog sizes, for both methods of computing $\xi(s)$.

![Figure 5](image-url)  
**Figure 5.** The total runtime (in CPU hours) of the fast and brute-force calculations of $\xi(s)$ as a function of the size of the random catalog.
a constant value because it depends on the fiducial volume and the number of angular and radial bins, rather than the size of the random catalog $N_R$. The maximum runtime of the fast algorithm is reached when each bin in the $R_{ang}(\alpha, \delta)$ distribution is populated and hence is used in the calculation of $\hat{\xi}(s)$.

To check the fast algorithm for bias, we compute $\hat{\xi}(s)$ for 20 mock catalogs representing $D$ with 200k galaxies Poisson-distributed within the survey volume. These mock catalogs are produced using the same procedure used to generate the random catalogs, described previously. We find no bias; that is, the mean of $\hat{\xi}(s)$ is centered at zero as expected for a truly random distribution, and its RMS obeys a Poisson distribution (Fig. 6). Note that the Poisson error is dominated by $DD$, since the $DR$ and $RR$ are much larger and hence contribute smaller relative errors.

Finally, the algorithm is also applied to the SDSS-III DR9 BOSS data catalog. The results obtained from the suggested algorithm and the brute-force algorithm and their difference are presented in Fig. 7. The BAO peak is clearly visible in both, and the two distributions are consistent.

5 CONCLUSION

We have presented a new computational method of the galaxy 2pcf that replaces summation over all possible galaxy pairs with a probability map integration. The method provides a significant reduction in the calculation time and improves the precision of the calculation. In the future, the method could be used for a fast evaluation of the galaxy correlations in large spectroscopic surveys. In this case, the generation of large-size random catalogs can be replaced by weighted probability maps determined by observational conditions. The minimum pair separation for which the method is applicable is set by the fiber size, as it is for the brute-force method. Reduced calculation time may also allow for the evaluation of larger numbers of mock scenarios and, thus, more extensive tests of cosmological parameters. Finally, we point out that the approach can be easily extended to evaluate higher order correlations.

ACKNOWLEDGMENTS

We thank L. Samuchia, E. Blackman and B. Betchart for useful discussions. The authors acknowledge the support from the Department of Energy under the grant DE-SC0008475.

REFERENCES

Adami C., et al., 2014, Astron. Astrophys., 572, A15
Alonso D., 2012, preprint, (arXiv:1210.1833)
Andrade L., et al., 2013, Mon. Not. Roy. Astron. Soc., 427, 3425
Aubourg E., et al., 2015, Phys. Rev. D 92, 123516 (2015), 92, 123516
Davis M., Peebles P. J. E., 1983, Astrophys. J., 267, 465
Eisenstein D. J., et al., 2005, Mon. Not. Roy. Astron. Soc., 362, 507
Feldman H. A., et al., 1994, Astrophys. J., 426, 23
Groth E. J., Peebles P. J. E., 1977, Astrophys. J., 217, 385
Hamilton A. J. S., 1993, Astrophys. J., 417, 19
Kerscher M., Szapudi I., Szalay A. S., 2000, Astrophys. J., 535, L13
Landy S. D., Szalay A. S., 1993, Astrophys. J., 412, 64
Moore A. W., et al., 2001, in Mining the Sky. pp 71–82 (arXiv:astro-ph/0012333)
Peebles P. J. E., 1973, Astrophys. J., 185, 413
Peebles P. J. E., Hauser M. G., 1974, Astrophys. J. Suppl., 28, 19
Peebles P. J. E., Yu J. T., 1970, Astrophys. J., 162, 815
Perbival W. J., et al., 2014, Mon. Not. Roy. Astron. Soc., 439, 2531
Ross A. J., et al., 2012, Mon. Not. Roy. Astron. Soc., 424, 564
Sanchez A. G., et al., 2012, Mon. Not. Roy. Astron. Soc., 425, 415
Sunyaev R. A., Zeldovich Ya. B., 1970, Astrophys. Space Sci., 7, 3
Vargas-Magana M., et al., 2013, Astron. Astrophys., 554, A131
Yu J. T., Peebles P. J. E., 1969, Astrophys. J., 158, 103

This paper has been typeset from a TeX/LaTeX file prepared by the author.

MNRAS 000, 1–6 (2016)