ABSTRACT — We discuss the statistical properties of the DTFE and show how the significance of structures in DTFE reconstructed homogeneous, inhomogeneous and filtered density fields can be determined. We describe how different types of sampling errors affect DTFE reconstructed fields. We describe measurement errors in the location or weight of the sampling points as well as systematic errors due to inhomogeneous sampling or sampling distortions. Finally, we summarize and discuss our results.
8.1 Introduction

In this thesis we have described the Delaunay Tessellation Field Estimator (DTFE), a method for reconstructing volume-covering and continuous density fields which are sampled by a discrete set of points. The DTFE is based on a basic concept of stochastic and computational geometry, the Delaunay tessellation of the point sample (Delone 1934, for further references see also Okabe et al. 2000, Møller 1994 and van de Weygaert 1991). This spatial volume-covering division of space into mutually disjoint triangular (in two dimensions) or tetrahedral (in three dimensions) cells adapts to the local density and geometry of the point distribution. The DTFE exploits this virtue and adapts automatically to changes in the density or the geometry of the distribution of sampling points. The Delaunay tetrahedra are used to obtain local estimates of the spatial density and as multi-dimensional intervals for linear interpolation of the field values estimated at the location of the sampling points (this interpolation procedure was introduced by Bernardeau & van de Weygaert 1996).

The DTFE has a self-adaptive spatial resolution. In Chapters 3 and 4 we have shown that its effective smoothing kernel is more localized than that of other reconstruction methods. It is therefore less forgiving with respect to sampling noise and errors. If present these will have a direct impact on the reconstructed field. This makes it necessary to understand how reconstructed structures are affected and how the significance of DTFE reconstructed density fields can be determined.

In the first half of this chapter we discuss the significance of DTFE reconstructed density fields. We restrict ourselves to Poisson sampling noise, which one for example encounters in N-body simulations and galaxy redshift surveys. Other types of sampling noise usually depend on a specific application and fall beyond the scope of this work. We discuss a number of frequently occurring situations, such as the determination of the significance of a signal in a uniform background field and the fitting of a model to an observed data set. We also discuss the determination of the significance of filtered fields. Finally, we discuss how the significance of DTFE reconstructed fields is related to the sampling density.

In the second half of this chapter we discuss the effects of sampling errors on DTFE reconstructed density fields. First we look at measurement errors, which may be present for the locations and the weights of the sampling points. Subsequently we consider systematic sampling errors, such as inhomogeneous sampling procedures and sampling distortions.

8.2 Sampling noise and significance

In this section we work out the significance of DTFE reconstructed density fields. In order to understand how sampling noise affects reconstructed density fields it is important to realize that even uniform fields are affected by sampling noise, because due to the Poisson nature of the sampling procedure different regions will contain more sampling points than others. This is illustrated in Fig. 8.1, in which several uniform Poisson point samplings are shown together with the corresponding Voronoi and Delaunay tessellations as well as the zeroth-order VTFE and first-order DTFE reconstructed density fields. One may observe density fluctuations in both the point samples, the tessellations and the VTFE and DTFE density fields.

Clearly the magnitude and frequency of these fluctuations determine how significant a particular reconstructed density value is. Very high and very low density values occur very rarely and are therefore rather significant. We will base our measure of significance on a
Figure 8.1 — Poisson sampling noise in uniform fields. The rows illustrate three Poisson point samplings of a uniform field with increasing sampling density (from top to bottom consisting of 100, 250 and 1000 points). From left to right the point distribution, the corresponding Voronoi tessellation, the zeroth-order VTFE reconstructed density field, the first-order DTFE reconstructed density field and the corresponding Delaunay tessellation are shown.

Consideration of the one-point distribution function of the DTFE density field reconstruction of a Poisson sampling of a uniform field, which describes the probability that a particular density value occurs due to Poisson fluctuations.

8.2.1 Statistical properties of Poisson sampling noise

Consider a Poisson point sampling of a uniform field. Kiang (1966) has shown that the distribution of sizes of Voronoi cells of such a point sampling may be approximated by

\[ dp(\tilde{a}) = \frac{c}{\Gamma(c)} (c\tilde{a})^{c-1} e^{-c\tilde{a}} d\tilde{a}. \]  

(8.1)

Here \( \tilde{a} = a/\langle a \rangle \) is the size of the Voronoi cell in units of the average cell size. The value of \( c \), a numerical constant, depends on the dimension of space. In two dimensions \( c = 4 \), in three dimensions \( c = 6 \). Note that there is some discussion on the precise functional form of \( dp(\tilde{a}) \) and the value of \( c \) (e.g. Tanemura 1988, Járai-Szabó & Néda 2004, see also Okabe et al. 2000 for the properties of Voronoi tessellations corresponding to Poisson point samplings), but these differences are not relevant for the range of interest of \( \tilde{a} \) for our purposes.

The probability for a random point to lie inside a cell of size \( \tilde{a} \) is equal to \( p(\tilde{a}) \times \tilde{a} \). In the zeroth-order Voronoi scheme this is the likelihood for a random point to have density \( \tilde{\lambda} = 1/\tilde{a} \). The one-point distribution function of the corresponding zeroth-order VTFE density
Figure 8.2 — One-point distribution functions of the DTFE reconstructed field and an analytic approximation (smooth curve) corresponding to a Poisson point process of 10 000 points in two (left-hand frame) and 100 000 in three (right-hand frame) dimensions.

It turns out that these equations also form an excellent approximation of the one-point distribution function of the first-order DTFE reconstructed field in two and in three dimensions. This can be seen in Fig. 8.2. It shows the one-point distribution function of the DTFE reconstructed density field corresponding to a homogeneous binomial random field of 100 000 points, with the analytic approximations superimposed. In this figure the cumulative distribution functions are also displayed.

Evidently, the two- and three-dimensional one-point distribution functions are non-Gaussian, with a tail extending toward high densities. Note that the distribution function falls off much more rapidly in three than in two dimensions. Using Eqns. 8.2 and 8.3 one may calculate various statistics corresponding to DTFE reconstructed homogeneous binomial random fields in two and three dimensions, some of which are listed in Table 8.1. The more rapid decline of the three-dimensional one-point distribution function is reflected by the lower value of its variance. The reason for the fact that extreme densities are less likely in three than in two dimensions.

field may therefore be approximated by

\[ 2D : \quad dp(\lambda) = \frac{128}{3} \lambda^{-6} e^{-4\lambda} d\lambda; \]
\[ 3D : \quad dp(\lambda) = \frac{1944}{5} \lambda^{-8} e^{-6\lambda} d\lambda. \]
Table 8.1 — Statistical properties of DTFE reconstructed two- and three-dimensional homogeneous binomial random fields.

<table>
<thead>
<tr>
<th></th>
<th>2D</th>
<th>3D</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>median</td>
<td>0.856</td>
<td>0.900</td>
</tr>
<tr>
<td>variance</td>
<td>1/3</td>
<td>1/5</td>
</tr>
<tr>
<td>skewness</td>
<td>2√3</td>
<td>√5</td>
</tr>
<tr>
<td>kurtosis</td>
<td>42</td>
<td>12</td>
</tr>
</tbody>
</table>

dimensions is that in three dimensions one effectively puts a constraint on three coordinates for a high density to occur and in two dimensions on only two. The positive value of the skewness for both distributions indicates the presence of the tail extending toward high densities. The larger value for the two-dimensional distribution is again indicative of the slower decline compared to the three-dimensional distribution. Finally, the kurtosis is (strongly) positive for both distributions, showing that they are more strongly peaked than normal distributions. The larger value for the two-dimensional distribution indicates that it is more strongly peaked than the three-dimensional one, presumably due to its relatively pronounced tail.

8.2.2 Significance of uniform fields

The one-point distribution function has a local interpretation as well. Given an ensemble of Poisson samplings of a uniform density field, at any location the distribution of DTFE reconstructed density values is equal to the one-point distribution function. This function may thus be used to assign the probability that a particular reconstructed density value is in accordance with an underlying uniform density field of a particular density.

Here we consider the practical application where one observes a particular point distribution and assumes that this is a realization of a certain uniform field with some signal field superposed. In such a case one is interested in the significance of the signal field. To be able to assign a measure of significance to local estimates of the recovered signal field we proceed as follows. Given a reconstructed density field \( \lambda_R(x) \) which is supposed to be the sum of an underlying uniform field \( \lambda \) and a signal field \( \sigma(x) \) we evaluate at each location \( x \) the two probabilities

\[
\begin{align*}
P_1(x) & \equiv P(\tilde{\lambda} \leq \lambda_R(x)) = \int_0^{\lambda_R(x)/\lambda} dp(\tilde{\lambda}) ; \\
\int_0^{\infty} \int_0^{\lambda_R(x)/\lambda} dp(\tilde{\lambda}) .
\end{align*}
\]

\( dp(\tilde{\lambda}) \) is the one-point distribution function of a DTFE reconstructed uniform field (Eqns. 8.2 and 8.3). At each location \( P_1 \) and \( P_2 \) represent the probability to find a density value \( \lambda \) lower \( (P_1) \) or higher \( (P_2) \) than \( \lambda_R \) in a DTFE reconstruction of a Poisson sampling of a uniform field with density \( \lambda \). \( P_1 \) and \( P_2 \) are therefore local measures of the significance of the signal field.
Equations 8.4 and 8.5 can be solved analytically, yielding

\[ P_1 = 1 - f\left(\frac{\lambda_R}{\lambda}\right) \]  
\[ P_2 = 1 - P(\lambda \leq \lambda_R) = f\left(\frac{\lambda_R}{\lambda}\right), \]  

in which the function \( f \) is defined as

\[ f(x) = \int_x^{\infty} dp(x') = \begin{cases} 2D : & e^{-4/x} \left[ \frac{32}{3x^3} + \frac{32}{3x^2} + \frac{8}{x} + \frac{4}{x} + 1 \right] \\ 3D : & e^{-6/x} \left[ \frac{324}{5x^3} + \frac{324}{5x^2} + \frac{54}{x^2} + \frac{36}{x^2} + \frac{18}{x^2} + \frac{6}{x} + 1 \right]. \]  

In Fig. 8.3 the probabilities \( P_1 \) and \( P_2 \) are plotted as a function of density for two and three dimensions. It is interesting to note that the probability to find a density higher than the average density is smaller than 0.5 in both two and three dimensions, or, equivalently, that the median density is smaller than the average density. This reflects the fact that the probability distribution is asymmetric: low density regions are more extended than high density regions. Also visible is that extreme densities are more likely to occur in two than in three dimensions. Finally, from the figure it is clear that low densities are much more unlikely to occur than high densities, both in two and in three dimensions.

We now proceed to define the significance of the signal field. From Eqns. 8.6 and 8.7 one may observe that \( P_1 \) and \( P_2 \) are equal when \( P_1 = P_2 = 0.5 \). Given an ensemble of DTFE reconstructions of Poisson point samplings of a given uniform density field, at any particular location half the reconstructions will have a smaller density and half will have a higher density than the value of the reconstructed density field for which \( P_1 \) and \( P_2 \) are equal. This value is the median density \( \lambda_{\text{median}} \) of a DTFE reconstruction of a Poisson sampling of a uniform field with density \( \lambda \). We define the significance \( S \) of the signal corresponding to the median density to be zero, while smaller and larger signals have a positive significance. The median density corresponds to a signal \( \sigma = \lambda_{\text{median}} - \lambda \). Since \( \lambda_{\text{median}} < \lambda \) (see Table 8.1) a signal \( \sigma = 0 \) (for which the reconstructed field has the same value as the underlying uniform field) has a positive significance. This counterintuitive result is the consequence of the asymmetric form of the one-point distribution function (Eqns. 8.2 and 8.3): low density regions are more extended than high density regions.

Next consider reconstructed density values higher than the median density. We define the significance of such densities (or equivalently, of the corresponding signals) to be equal to \( P_2 \), the probability to find in a DTFE reconstructed uniform Poisson field densities which are higher than \( \lambda_R \). Similarly we define the significance of reconstructed density values (or equivalently, of the corresponding signals) smaller than the median to be equal to \( P_1 \), the probability to find in a DTFE reconstructed uniform Poisson field densities which are smaller than \( \lambda_R \).

It is convenient to express the probabilities \( P_1 \) and \( P_2 \) in terms of Gaussian standard deviations (\( \sigma \)'s). We convert the significance \( S \) of a particular reconstructed density field value to the distance in units of \( \sigma \) from the median density to the density for which the probability to find a point in a Gaussian distribution with a value equal to or more extreme than this density
Figure 8.3 — The probability $P$ (dashed lines) and significance $S$ (solid lines, see description in text) to find a more extreme density than a particular density value $\lambda_R/\lambda$ in a DTFE reconstruction of a homogeneous binomial random field. Results are shown for both two-dimensional (thin lines) and three-dimensional (thick lines) fields.

is equal to the probability calculated using Eqns. 8.6 and 8.7. Mathematically,

$$\int_{S, \sigma} \infty \, \mathrm{d}t \, G(t) = \begin{cases} P(\lambda \leq \lambda_R), & \text{if } \lambda_R < \lambda_{\text{median}} \\ 0, & \text{if } \lambda_R = 0 \\ P(\lambda \geq \lambda_R), & \text{if } \lambda_R > \lambda_{\text{median}} \end{cases} ; \quad G(t) = \frac{1}{\sigma \sqrt{2\pi}} \, e^{-t^2/2\sigma^2}. \quad (8.9)$$

Solving this equation yields the following expressions for the significance $S$:

$$S(\lambda_R) = \begin{cases} \sqrt{2} \, \text{erfc}^{-1} \left[ 2f \left( \frac{\lambda_R}{\sigma} \right) - 1 \right], & \text{if } \lambda_R < \lambda_{\text{median}} \\ 0, & \text{if } \lambda_R = \lambda_{\text{median}} \\ \sqrt{2} \, \text{erfc}^{-1} \left[ 1 - 2f \left( \frac{\lambda_R}{\sigma} \right) \right], & \text{if } \lambda_R > \lambda_{\text{median}}. \end{cases} \quad (8.10)$$

For general values of $\lambda_R$ this equation has to be solved numerically. In Fig. 8.3 the significance is plotted as a function of density for two and three dimensions.

Interestingly, Eqn. 8.10 does not contain any dependence on the sampling density. Naively one might think that the influence of sampling noise diminishes for a larger number of sampling points. However, since the DTFE is a local procedure, the magnitude of the density fluctuations due to sampling noise is independent of the sampling density, only their spatial scale changes. The effects of sampling noise may be suppressed by filtering the reconstructed field. The effects of filtering on the significance of reconstructed features is discussed in section 8.2.4.
8.2.2.1 Example: significance of a detected signal in a uniform background

As an example of how the above formalism works in practice we have simulated the observation of a number of Gaussian and ellipsoidal objects in a Poisson background, with varying ratios between the density of the background and the peak density of the object. The simulated observed point distributions are shown in Fig. 8.4. In the left-hand panel of Fig. 8.4 a Gaussian object is shown with, going from top to bottom, a peak-to-background ratio (P/B) of 4, 2, 1 and 0.5. In the right-hand panel the same is done for an ellipsoidal object with an axis ratio of 10 : 1 and the same peak-to-background ratios as for the Gaussian object. In Table 8.2 we have listed a number of parameters for these simulations, including the relative amount of points belonging to the Gaussian and ellipsoidal structures with respect to the background, for the frames shown in Fig. 8.4. In the central columns the DTFE reconstructed density fields are displayed. In the right-hand columns the significance of the reconstructed field values is plotted. As we are interested in a signal on top of the background, we have only displayed the significance values for positive deviations from the background.

In Fig. 8.4 the objects with P/B = 4 and P/B = 2 are conspicuously visible, both in the point distributions as well as in the field reconstructions and even more so in the significance maps. The same is true to a lesser extent for the object with P/B = 1. In this case the significance maps are the most clear indicator of the presence of these objects. The objects
Table 8.2 — Parameters for the simulations of Gaussian and ellipsoidal objects in a Poisson background. Listed is the number of points belonging to these objects as well as their relative amount with respect to the total number of points.

<table>
<thead>
<tr>
<th>peak-to-background</th>
<th>Gauss</th>
<th>Percentage</th>
<th>Ellipsoid</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.0</td>
<td>6283</td>
<td>20.4%</td>
<td>4734</td>
<td>20.8%</td>
</tr>
<tr>
<td>2.0</td>
<td>3141</td>
<td>11.4%</td>
<td>2370</td>
<td>11.6%</td>
</tr>
<tr>
<td>1.0</td>
<td>1570</td>
<td>6.0%</td>
<td>1186</td>
<td>6.2%</td>
</tr>
<tr>
<td>0.5</td>
<td>785</td>
<td>3.1%</td>
<td>603</td>
<td>3.2%</td>
</tr>
</tbody>
</table>

with $P/B = 0.5$ are not visible in the point distribution and only a hint of them in the field reconstructions. However, the significance maps reveal a stronger indication of the presence of an object. The last example shows that the DTFE manages to pick up objects below the level of the background. Nevertheless, in all cases it is clear that the reconstructed fields are strongly influenced by the background noise, which is the result of the high spatial resolution of the DTFE procedure. In practical applications one would often proceed by fitting a model to the observed signal (see section 8.2.3) or by filtering the reconstructed density field (see section 8.2.4).

### 8.2.3 Significance of inhomogeneous fields

In the previous section we have analyzed the significance of a detected signal on top of a uniform background field. Another frequently occurring practical situation is the fitting of a model to a particular point sampling. In such a case one determines the significance of the deviations of the model and evaluates whether these are likely to be caused by sampling noise or whether these are statistically significant. It is important to realize that it is not possible to directly determine the significance of the model itself, one can only rule out models to a particular degree by evaluating the significance of the deviations. The previously discussed example of a signal in a uniform background is in fact a special case in which the model is a uniform field and one evaluates the significance of the deviation of the signal from the uniform background (see section 8.2.2).

For a proper analytical description of inhomogeneous fields one should take into account the multi-point distribution functions or moments. However, for practical purposes a proper assessment on the basis of a multi-point distribution is not feasible within the scope of this work. To take into account spatial correlations within the field would quickly lead to a cumbersome mathematical treatment, not necessarily going along with a clearly defined significance (see e.g. the discussion in Gaztañaga 1989, 1992, 1994, Lemson 1995).

To be able to assign a measure of significance to local estimates of a recovered density field, we simply proceed by assuming that the field is locally uniform. In such a case one may determine at each location the significance using the formalism described in section 8.2.2 after replacing the uniform underlying density field by an inhomogeneous density field.

The assumption that the field is locally uniform clearly is valid when the sampling density is high enough for the density field not to fluctuate significantly over the extent of a Delaunay
Figure 8.5 — Gaussian model fitting. Frame 1: Poisson sampling of a Gaussian density profile by 500 points. The circle denotes the outermost radius $R_0$ at which the Gaussian is sampled relatively well. Frame 2: DTFE reconstruction of the region inside $R_0$. Frames 3 to 6: significance maps of the deviations of the DTFE reconstruction with four different models. Frame 3: model equals the underlying profile. Frame 4: model with a too large mass. Frame 5: Model with a too small spatial extent. Frame 6: model with an offset central position. The amplitude of the significance is indicated by the color bar.

triangle or tetrahedron:

$$\frac{|\nabla \lambda| \cdot \lambda^{-D}}{\lambda} \ll 1,$$

in which $D$ is the dimension of space. If this condition is not valid then the calculated significance will be an upper limit and Monte Carlo simulations may be used to get an estimate of the true significance. An example of how Monte Carlo simulations can be used for determining the significance of a DTFE reconstructed density field is discussed in section 8.3.1.

### 8.2.3.1 Example: Gaussian model fitting

As an example we have fitted a number of models to the DTFE reconstruction of a point sampling following a Gaussian profile:

$$\lambda(x) = \frac{M}{\pi\sigma^2} \exp \left\{ -\left( \frac{x-x_0}{\sigma} \right)^2 \right\}.$$
The realization consists of 500 points and is shown in frame 1 of Fig. 8.5. The circle corresponds to the outermost ring which is still properly sampled. It can be shown that this radius is in fact somewhat larger than appropriate for DTFE reconstructions, but it will serve to illustrate the deviations between the reconstruction and the underlying model which arise due to the finite number of sampling points. The DTFE reconstruction is shown in frame 2.

We have fitted four models to this reconstruction. All consist of Gaussian profiles, given by Eqn. 8.12. The models are graphically illustrated in Fig. 8.6, while their parameters are listed in Table 8.3. Model 1 corresponds to the underlying profile, model 2 has a larger mass, model 3 has smaller spatial extent and in model 4 the position of the center is offset. For each of these models we have calculated the significance of the deviations of the DTFE reconstruction with respect to the model. The resulting significance maps are shown in Fig. 8.5, frames 3 to 6.

The first thing to note is that there are strong differences between the four models. Naturally we have adopted model parameters which result in clearly distinct profiles, so that the differences between the models stand out. Model 1, the underlying profile, fits best. Most deviations have a small significance (< 2) and only few locations have a large significance (> 3). These correspond to small areas, usually confined to a single triangle. An exception to this are the outer edges, which deviate from the model with a large significance. This is due to the poor sampling in this area. Also note that the scale of the variations is varying, from relatively small in the central regions to larger variations in the outer regions. This is obviously due to the variation of the sampling density, which decreases as a function of distance to the center, resulting in a lower effective resolution towards the edge.

Table 8.3 — Overview of the properties of the models used for fitting a DTFE reconstruction of an inhomogeneous binomial random point sampling of a Gaussian density profile.

<table>
<thead>
<tr>
<th>Model</th>
<th>$M$</th>
<th>$\sigma$</th>
<th>$x_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>model 1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>model 2</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>model 3</td>
<td>1</td>
<td>0.5</td>
<td>0</td>
</tr>
<tr>
<td>model 4</td>
<td>1</td>
<td>1</td>
<td>$-0.5\sigma$</td>
</tr>
</tbody>
</table>
The significance map of model 2 looks very similar to the one of model 1, except that the significance of the deviations is much larger everywhere. This is because model 2 is a scaled up version of model 1, with only the amplitude different (see Fig. 8.6). The map of model 3 shows a very different pattern, with a systematically very high significance at the center and the outer regions, while the significance is low in a ring around the center. This ring corresponds to regions where the model happens to have an amplitude which is comparable to that of the underlying field (see Fig. 8.6). The map of model 4 shows a different pattern, with a diagonal band with low significance and a systematically very high significance in the other regions. This band corresponds to regions where the model happens to have an amplitude which is comparable to that of the underlying field (see Fig. 8.6).

In conclusion we have seen that it is possible by calculating the significance maps to discern between different models. In this way one could do a parameter fit for a particular generic type of model (such as the Gaussian profile of Eqn. 8.12) or compare completely different types of models.

8.2.4 Significance of filtered fields

So far we have only discussed the uncertainty due to sampling noise in unprocessed DTFE density field reconstructions. However, in practical applications reconstructed fields are often filtered in order to increase the signal-to-noise ratio. In such cases it becomes rather complex to calculate the significance of reconstructed density values. The reasons are similar to those encountered when determining the significance of inhomogeneous fields: one has to take into account the multi-point distribution functions of the reconstructed fields. The results also depend on the particular shape of the smoothing filter, which may depend on the application.

This problem is rather familiar from the analysis of for example radio observations, in which the observed signal is convolved with an observational beam. There it is customary to define the significance of an observed signal in units of the variance of the signal in a homogeneous part of the sky which does not contain distinct radio sources. This variance is a measure of the probability that a certain signal occurs due to background noise. Note that due to the particular parameters specific for the observations, this variance is not usable for other observation runs or other telescopes. Moreover, due to the spatial extent of the beam the variance is not uniquely related to the probability for such a signal to occur due to the background noise. This makes the interpretation of the significance measured in units of the variance somewhat subjective. Nevertheless, the straightforwardness and conceptual simplicity of the variance has made it a reliable and standard discriminator of which regions are statistically significant and which are not.

In order to get an impression of the effect filtering has on the significance of DTFE reconstructed density fields, we have filtered the DTFE reconstruction of a 10 000 point Poisson sampling of a uniform field and measured the variance for a number of different filtering radii. On the left-hand side of Fig. 8.7 the unprocessed DTFE reconstruction is shown (upper left-hand frame) as well as the filtered density fields (subsequent frames). The fields have been filtered by a fixed Gaussian kernel whose FWHM is plotted in the upper left-hand section of each density field. The FWHM is specified in units of the average scale of the DTFE kernel, which has been defined as the square root of the volume of the contiguous Delaunay cell corresponding to a sampling point. On the right-hand side of Fig. 8.7 the measured variance of the filtered fields is plotted as a function of the FWHM of the Gaussian filter.
Figure 8.7 — The effect of filtering on the sampling noise in DTFE reconstructed density fields. In the upper left-hand frame a DTFE reconstructed density field corresponding to a 10 000 point Poisson sampling is shown. In the subsequent frames the same field is shown after filtering with a Gaussian filter whose FWHM is plotted in the upper left-hand section of each density field. The FWHM is specified in units of the scale of the DTFE kernel (see text for a description). On the right-hand side the measured variance in these density fields is plotted as a function of the FWHM of the Gaussian filter.

It can be seen that filtering the DTFE reconstructed field by a Gaussian filter with a FWHM up to twice the average size of the DTFE kernel reduces the variance quite substantially, while for larger filters the variance decreases relatively more slowly. This result suggests that one may suppress the sampling noise of a DTFE reconstructed field substantially by filtering it with a Gaussian with a FWHM up to twice the average size of the DTFE kernel. Filtering with a larger kernel is less profitable in terms of reducing the variance of the sampling noise. The reason for this different behaviour is that only for filters which are larger than the average scale of the DTFE kernel the density field is averaged over different contiguous Voronoi cells, while for smaller filters one merely changes the shape, but not the scale of the DTFE kernel.

Note that in general fields are inhomogeneous, which introduces a complication. For such fields using a fixed filter will reduce the amount of sampling noise by a factor which depends on the local density. To circumvent this one may use an adaptive filter.

8.2.4.1 Example: Filtering of a signal in a uniform background

As an example we have filtered the DTFE reconstructed density maps corresponding to the Gaussian and ellipsoidal objects in a Poisson background with $P/B = 1$ (see section 8.2.2). The resulting filtered maps are plotted in Fig. 8.8. In the left-hand column the unsmoothed fields are plotted, while the subsequent frames show the same images, but smoothed with a
Gaussian filter of increasing size. The diameter of the shaded circle shown in the lower right-hand corner of these frames indicates the FWHM of the filter. The significance of the signal may be expressed in units of the variance due to the background. This variance could either be measured in the filtered fields or be deduced from Fig 8.7.

This example clearly shows how such filtering succeeds in increasing the signal-to-noise levels, but it also shows that the properties of the objects are affected by the size and shape of the smoothing kernel. In particular, the anistropy of the object seen in the bottom right-hand frame is less pronounced than the true anisotropy of the underlying object. The advantage of using the DTFE as reconstruction tool guarantees that no extra smoothing is introduced by the reconstruction procedure itself and the interpretation of these smoothed maps is therefore relatively straightforward.

8.3 Sampling errors

Apart from being affected by sampling noise, data in practical applications are usually affected by one or more types of sampling errors. An important example concerns measurement errors in the position or the mass of the sampling points. Systematic errors may be introduced by an inhomogenous sampling procedure or by sampling distortions. In this section we work out the effect of these errors on the DTFE reconstruction.

8.3.1 Measurement errors: the location of the sampling points

A common type of error in many types of applications is an uncertainty in the determination of the position of the sampling points. In Fig. 8.9 this is illustrated for the observation of a Voronoi wall model of the large scale galaxy distribution (see Chapter 4 for a description), in which the location of the sampling points has not been measured with absolute certainty,
but with some measurement error. In the figure the error increases from no error (left-hand frame) to a rather large error (right-hand frame). It is obvious that the point distribution does not change significantly when the measurement errors are small, while it appears very differently for large measurement errors.

One might expect that the Delaunay tessellation corresponding to a set of sampling points does not change when measurement errors are small, while it does for large measurement errors. These two cases are illustrated in Fig. 8.10. In the left-hand panel the errors are so small that although the position of each point has changed somewhat, the effects are not large enough for the triangle identities of the tessellation to change. In other words, the vertices of the triangles or tetrahedra in the Delaunay tessellations will still correspond to the same sampling points. If the triangle identities do not change, a formal expression may be written

Figure 8.9 — Measurement errors in the position of the sampling points. Shown is a slice through a Voronoi wall model of the large scale galaxy distribution (see Chapter 4 for a description). The position of the sampling points is measured with an uncertainty whose magnitude increases from no error (left-hand frame) to a significant error (right-hand frame).

Figure 8.10 — Consequences of changes in the position of vertices for the triangle identities in a Delaunay tessellation. Case 1: the uncertainties in the position of the sampling points are not large enough for the triangle identities to change. Case 2: the uncertainties are so large that the tessellation must have changed.
Figure 8.11 — Demonstration that the contiguous Voronoi cell of a point does not change due to a change in its position as long as the triangle identities remain unchanged.

Figure 8.12 — Change of triangle identities due to changes in the position of the sampling points. Shown is a series of three frames depicting a set of four sampling points. The upper right-hand point is moving downwards to the left as denoted by the arrow. In the central frame the tessellation has become degenerate, as the four points have equal distance to each other. In the right-hand frame the triangle identities of the tessellation have changed with respect to the left-hand frame.

down for the uncertainty in the density estimates at the location of the sampling points due to the uncertainty in the position of the sampling points errors. As a side remark, it is interesting to note that for an unchanged tessellation an uncertainty in the position of a point does not lead to an uncertainty in the density estimate at the location of that point, but instead to an uncertainty in the density estimates at the location of its surrounding points. The reason for this is that as a particular point moves around, its contiguous Voronoi cell remains unchanged as long as the triangle identities stay the same, which is illustrated in Fig. 8.11.

In the right-hand panel of Fig. 8.10 the errors are so large that the triangle identities of the tessellation have changed. There it can be seen that the errors lead to a configuration which is not a Delaunay tessellation, so that the triangle identities must have changed. In such cases it is not possible to write down a formal expression for the error for the density estimation at the location of the sampling points. Instead, one should perform Monte Carlo simulations in order to get a handle on the uncertainty in the reconstructed density field values.

Even though these two cases may be discerned in principle, in practice it is usually impossible to tell beforehand if uncertainties in the position determination may lead to identity changes. The reason for this is that in certain configurations it is possible that an infinitesimal change in position leads to a change of triangle identities. One may see this from Fig. 8.12, in which the four sampling points in the left-hand and right-hand frame are close to a degenerate configuration, which is shown in the central frame. In practice one should therefore always perform Monte Carlo simulations in order to get an estimate of the uncertainty in reconstructed density values.
8.3.1.1 Example: Monte Carlo simulations of the observation of anisotropic structures

As an example of how this works in practice, we have simulated the observation of an ellipsoidal structure with axis ratio of 10 : 1, in which the positions of the sampling points are measured with a Gaussian error. For this we have simulated an ellipsoidal structure whose density follows the profile

\[
\rho(x, y) = \frac{M}{\pi(\sigma_x^2 + \sigma_y^2)} \exp\left[-\left(\frac{x - x_0}{\sigma_x}\right)^2 + \left(\frac{y - y_0}{\sigma_y}\right)^2\right].
\]  

(8.13)

The resulting point distribution is shown in the top left-hand frame of Fig. 8.13. We have assumed that the positions of these sampling points have been measured with a Gaussian error with a standard deviation \( F \) which is five times larger than the standard deviation \( \sigma_y \) of the ellipsoidal profile along its minor axis. The resulting observed point distribution is shown in the top right-hand frame of the figure. This distribution is equivalent to another ellipsoidal distribution, but with the \( x \) and \( y \) replaced by

\[
\begin{align*}
\sigma_x &\rightarrow \sigma'_x = \sqrt{\sigma_x^2 + \sigma_F^2} \approx 1.1\sigma_x; \\
\sigma_y &\rightarrow \sigma'_y = \sqrt{\sigma_y^2 + \sigma_F^2} \approx 5.1\sigma_y.
\end{align*}
\]  

(8.14, 8.15)

The observed distribution forms an ellipsoidal object with axis ratio of about 2.2 : 1.

We have checked if the DTFE reconstruction follows this theoretically expected behaviour by measuring at four characteristic locations, denoted by the thick dots in the top right-hand frame of Fig. 8.13, the uncertainty in the DTFE reconstructed density field. For this we have constructed 1000 realizations of possible underlying point distributions, based on the measured locations and the Gaussian distribution of measurement errors, and reconstructed the corresponding density field. For each of the four positions the resulting error distributions are shown in the bottom four frames of the figure. In these frames the one-point probability distribution function of the reconstructed density at that position is shown, normalized by the expected observed density at that position (as given by an ellipsoidal with parameters given by Eqns. 8.14 and 8.15). The smooth curve depicts the theoretical probability distribution function given by Eqn. 8.2. The thick vertical line denotes the observed average reconstructed density, which should equal unity, because we have normalized the observed densities by the expected observed density.

In the figure it can be seen that at all four positions the average reconstructed density is close to unity and that at positions 1, 2 and 3 the observed probability distribution function fits the theoretical distribution very well. At position 4 this is not the case: an overabundance of very low densities is present at this location. This happens because the number of sampling points is too small for the density field to be properly sampled at that position. We conclude that the DTFE reconstruction behaves in accordance with the theoretical expectation.

8.3.2 Measurement errors: the weight of the sampling points

The second type of measurement errors we discuss is an uncertainty in the weight of the sampling points. An example of this is the uncertainty in the mass of a galaxy in deriving the density field corresponding to an observed redshift survey, or the uncertainty in the flux
Figure 8.13 — Monte Carlo simulations of the observation of an ellipsoidal structure. In the top row the real underlying point distribution is shown which consists of 500 points (left-hand frame) as well as the observed point distribution (right-hand frame), in which the observed positions have measurement errors. The numbers 1, 2, 3 and 4 denote the locations at which the one-point probability distribution function (PDFs) of the reconstructed density has been determined on the basis of 1000 Monte Carlo simulations. These PDFs are shown in the bottom four frames. In these frames the smooth curve denotes the expected PDF, which corresponds to the ellipsoidal profile smoothed with a Gaussian filter. The thick line denotes the observed average reconstructed density.
Figure 8.14 — Demonstration that the uncertainty due to the weight of a sampling point is confined to the contiguous Voronoi cell of this point.

of a photon when deriving an density field. Since an uncertainty in the mass of a particle \( i \) only leads to an uncertainty in the mass estimate at this location and to an uncertainty in the gradient of the triangles (tetrahedra) constituting its contiguous Voronoi cell, the effects of this uncertainty are confined to this contiguous Voronoi cell, unless other types of errors are present as well. This is illustrated in Fig. 8.14, in which the uncertainty in the effective DTFE kernel corresponding to a particular sampling point is shown in \((r, \rho)\)-space.

It is straightforward to write down the expression for the uncertainty in the density estimate at the location of this point and also, although a bit more cumbersome, for the uncertainty in the density gradient in its surrounding Delaunay triangles. Consider a sampling point \( i \) with mass \( m_i \pm \Delta m_i \) and a contiguous Voronoi cell with volume \( V(W_i) \). The uncertainty \( \Delta \lambda_i \) in the density estimate \( \lambda_i \) at this location due to the uncertainty \( \Delta m_i \) in its mass is then given by

\[
\Delta \lambda_i = \frac{(D + 1)}{V(W_i)} \Delta m_i. \tag{8.16}
\]

Once the densities \( \lambda_i \) at the locations of the sampling points have been estimated, the components of the density gradient in each triangle \( j \) may be found by solving the following set of equations:

\[
\begin{align*}
\lambda(r_1) &= \lambda(r_0) + \nabla \lambda_j \cdot (r_1 - r_0) ; \\
& \vdots \tag{8.17} \\
\lambda(r_D) &= \lambda(r_0) + \nabla \lambda_j \cdot (r_D - r_0).
\end{align*}
\]

Here \( D \) is the dimension of space and \( r_0, \ldots, r_D \) are the \( D \) vertices of triangle \( j \). The uncertainty in the density gradient in its surrounding triangles due to the uncertainty in the mass of particle \( i \) can be found by solving this set of equations and applying standard error analysis. Note that these equations may only be used for an error determination if no other types of errors, such as errors in the position of the sampling points are present. If that is not the case, Monte Carlo simulations can be used to determine the error distribution.
8.3.3 Systematic errors: inhomogeneous sampling

An important issue for many practical applications occurs when the number density of sampling points is not just a function of the underlying field, but also of some other variable(s). In such a case the sampling is inhomogeneous over the region of space one is looking at. An example of an application in which one is dealing with a selection function are magnitude-limited cosmological redshift surveys, in which one is measuring the positions and redshifts of galaxies in some part of the sky. Because the amount of the light observed from these galaxies decreases with their redshift, a smaller fraction of galaxies is observed at larger redshifts. This is clearly visible in Fig. 8.15, in which the galaxy distribution is shown as has been measured by the 2dF Galaxy Redshift Survey (Colless et al. 2001).

The number density of sampling points $n(x)$ may be written as proportional to the product of the underlying field and a sampling function $\phi(x)$, which is often called the selection function,

$$n(x) \propto \lambda(x) \phi(x). \quad (8.18)$$

Here we have defined the selection function as the fraction of sampling points relative to some normalization. The selection function includes the dependency on all variables other than the underlying field.

In case the selection function is known one may correct for the smaller fraction of observed galaxies by simply multiplying the weight of each sampling point by the inverse of the selection function. Mathematically, the corrected density estimate at the location of sampling
point $i$ is given by

$$\tilde{\lambda}_i = \frac{1}{\phi(x_i)} \frac{(D + 1) m_i}{W_i}. \quad (8.19)$$

Clearly an uneven sampling will have repercussions on the resolution of the reconstructed field, which have to be taken into account when interpreting or analyzing a reconstructed field. In particular, the effective resolution of features with the same density is different at different locations.

### 8.3.4 Systematic errors: sampling distortions

A more complicated effect occurs when the sampling of an underlying field is not measured as a function of position, but instead as a function of some other variable(s) $\alpha(x)$. In this case a directly reconstructed density field will yield a distorted view of the proper density field. Such distortions occur for example in cosmological redshift surveys, where the distance to observed galaxies is not measured directly, but via their redshifts. The redshift of a galaxy depends for a large part on the distance through Hubble’s law (Hubble 1929), which contains a dependency on the local dynamical environment, the so-called proper motion of a galaxy. The proper motion of a galaxy induces a non-linear relation between distance and redshift, which makes the transformation between the two very hard. For example, large-scale infalling regions appear compressed in redshift-space, the Kaiser effect (Kaiser 1987). Conversely, large underdense regions which are growing are elongated in redshift space. Perhaps the most striking and readily visible distortions are the well-known fingers of God (see Fig. 8.16). Due to the large virial motions inside a galaxy cluster these do not appear as compact regions in redshift space, but instead as very elongated finger-like structures which appear to be pointing toward us. A distorted point sampling should be transformed to real space before a proper DTFE field reconstruction can be done. This is only straightforward when the mapping between the variables $\alpha$ and the proper positions $x$ is known and invertible. In practical applications this is often not the case and the mapping is strongly dependent on the specific application. In the case of the galaxy redshift surveys the mapping is for example determined by the local mass distribution.
8.4 Summary and discussion

In this chapter we have discussed how the significance of DTFE reconstructed density fields may be determined. For this purpose we have defined the significance as a measure of the probability to find a point in a reconstructed homogeneous Poisson field of a particular intrinsic density $\lambda$ with an density more extreme than the reconstructed value $\hat{\lambda}_R$. This complicates the interpretation of derived significance maps, as nearby points may lie within the same triangle and are therefore not independent from each other. This is similar to the problem one encounters in the analysis of many types of astronomical observations, in which the observing instrument has a particular point spread function which affects the statistics of the observed signal. An extra complication is that the sizes and shapes of the Delaunay triangles are varying over the map. One way to avoid the problem of dependent image pixels would be to define the significance not at each location but for each individual triangle. This calculation is in principle not different from the calculation in section 8.2.2. However, the significance maps calculated along these lines would also be complicated to interpret as in this case the size of the independent regions (the triangles) would vary over the map. In particular, a visual comparison would be strongly biased towards low density regions, where the triangles are much larger than in high density regions.

We have discussed two types of systematic sampling effects, inhomogeneous sampling and sampling distortions. Depending on the application other systematic sampling effects may be present as well. An important example of these which occurs in the analysis of cosmological redshift surveys is the (inhomogeneous) Malmquist bias. Since these effects depend on the application, we will not discuss them any further here.

References

Hubble E., 1929, Proc NAS., 15, 168
Járai-Szabó F., Néda Z., 2004, cond-mat/0406116
Kiang T., 1966, Z. Astrophys., 64, 433
Møller J., 1994, in Lecture Notes in Statistics 87, Springer-Verlag, New York, USA
Tanemura M., 1988, J. Micr., 151, 247