

# Simulation-based inference for random fields in cosmology

Jed Homer

Chair of Astrophysics, Cosmology and Artificial Intelligence  
(ACAI)

## Contents

<b>Synopsis</b>	<b>2</b>
<b>Learning outcomes</b>	<b>2</b>
<b>1 Introduction</b>	<b>3</b>
<b>2 Bayesian inference</b>	<b>3</b>
<b>3 A simple model for the matter density field</b>	<b>5</b>
3.1 Gaussian probability distribution function . . . . .	6
3.2 Gaussian random fields . . . . .	6
3.3 Fourier space . . . . .	11
3.4 How to generate a Gaussian random field . . . . .	13
3.5 Fisher information matrix . . . . .	14
3.6 The Fisher information for the GRF likelihood . . . . .	17
<b>4 Simulation-based inference</b>	<b>18</b>
4.1 Density estimation for SBI . . . . .	18
4.2 Normalizing flows . . . . .	18
4.2.1 A simple change-of-variables . . . . .	21
4.2.2 Optimising a normalizing flow . . . . .	21
4.2.3 NVP normalizing flows . . . . .	22
4.2.4 More layers . . . . .	23
<b>5 Summary</b>	<b>23</b>
<b>6 Testing SBI</b>	<b>24</b>
6.1 Working with a normalizing flow . . . . .	24
6.2 Generate a GRF dataset . . . . .	26
6.3 Fit a normalizing flow to a GRF dataset . . . . .	26
6.4 Two ways of inference . . . . .	27

## Synopsis

**Gaussian random fields** (GRFs) are a very simple statistical model for matter density fluctuations in the Universe evolved to linear order in the fluctuations. The initial conditions of the Universe are thought to be drawn from a GRF with a nearly scale-invariant **power spectrum**. The Cosmic Microwave Background maps temperature fluctuations in the early Universe and has been measured to be a nearly perfect GRF. However, the late-time density field departs from Gaussianity due to non-linear gravitational evolution of the field. This induces higher-order moments above the variance that are not characterised by two-point statistical measures of the field and equivalently not by the power spectrum.

The **likelihood function** of the GRF is of a simple multivariate Gaussian form, for a zero-mean density field, it is entirely defined by the 2-point correlation function of perturbations as a function of physical scale, or equivalently, the power spectrum (in Fourier space). Typically models for the matter density field parameterise the power spectrum as a function of fundamental physical properties of the Universe such as the energy content in terms of matter and radiation, and the index in a power law model for the primordial power spectrum.

These considerations for this simplified model of the large-scale linear perturbations of matter density provide an insightful problem to study **simulation-based inference**(SBI). The likelihood function for the late-time density field does not have an analytical form but we can map the field (or tracers of the field) with modern cosmological survey missions. To use the information in this data to constrain parameters of cosmological models, we need a likelihood function, and SBI provides a way to do exactly this using simulated model predictions. The Gaussian form of the likelihood provides an analytic prediction for the Bayesian posterior on the model parameters given a measurement. The Fisher information matrix provides the uncertainty on the model parameters given an estimate. This means the SBI analysis of the high-dimensional field measurement may be compared to the exact solution.

## Learning outcomes

- Understand Bayesian inference for cosmological inverse problems where the parameters of a physical model are inferred from data.
- Understand problems in inference when the likelihood function is not known.
- Understand an algorithm for fitting a likelihood function using simulations when an analytic likelihood function is not known.

- Understand how to use the algorithm, a normalizing flow, to fit a conditional likelihood distribution.
- Understand what a Gaussian random field is, why it is relevant to cosmology, and how to write down a statistical model for such fields.
- Understand the Fisher information a likelihood contains on the parameters of a model.
- Directly compare two approaches to inferring a posterior on a set of physical parameters from a measurement of a field.

## 1 Introduction

*“Observational cosmology consists of making inferences about the universe by comparing data and theoretical models.”*

Will J. Percival et al. (see [9])

In this lab, we will demonstrate an example of inferring the unknown parameters of a toy model for the matter density field of the Universe from a simulated measurement of the same field. We will fit a model to a noisy measurement of the matter density field and derive a posterior distribution function for the parameters of the model given this measurement. We will see that our simple model for the Universe has an analytic likelihood and therefore an information content that can be calculated. We will compare the analytic information content of the data upon our model to the posterior we derive using machine learning based methods.

Structure in the Universe forms on large-scales from the action of gravitation on small perturbations in the initially smooth Universe. The initial perturbations were likely formed by quantum effects during an inflationary epoch of expansion in the early Universe. Because of this ultimately probabilistic origin, the fields that are observed in cosmology are random - a statistical description is required. This is akin to the descriptions provided by statistical mechanics for molecules in a gas, the positions and velocities of every molecule are not particularly useful, but rather a statistical description of a random process of which the observed system could have been drawn from.

## 2 Bayesian inference

A Bayesian analysis compares a measurement  $\hat{\Delta}$  with a model  $\Delta[\pi]$  that is parameterised by some model parameters  $\pi$ . For each value of the parameters  $\pi$ , the measurement  $\hat{\Delta}$  is affected by statistical uncertainties. These statistical uncertainties are characterised by

$$\mathcal{L}(\pi) = p(\Delta = \hat{\Delta} | \pi) \tag{1}$$

which is known as the *likelihood function*. This quantity evaluates the likelihood of the data given a fixed value of the model parameters (assuming that the model is true) for a measurement  $\hat{\Delta}$ .

Typically we are interested in fitting a model to measured data, which means we would like the probability of each parameterisation of the model  $\boldsymbol{\pi}$  given the measurement we observed  $\hat{\Delta}$ ; this is known as the *posterior* probability distribution for the parameters and is written

$$p(\boldsymbol{\pi}|\hat{\Delta}) = \frac{1}{\mathcal{N}} \mathcal{L}(\boldsymbol{\pi}) p(\boldsymbol{\pi}) \quad (2)$$

where  $p(\boldsymbol{\pi})$  is the prior density which encodes our prior beliefs on the model before measuring any data. Here,  $\mathcal{N}$  is a parameter-independent normalisation factor which we ignore for now.

A scientist may fit the model, by some procedure, to obtain the posterior probability density function (PDF). This posterior summarises the state of knowledge of the scientist upon the model parameter values in posterity of the data. A more typical result involves a quote of the parameter values that maximise the posterior (the so-called maximum a posteriori or MAP value), with a number that measures the size of the region in parameter space that contains some fraction of the total probability integrated over all of the parameter space. This is the Bayesian's *credible region*.

So our requirements for fitting a model to data are:

- a likelihood  $p(\Delta|\boldsymbol{\pi})$ ,
- a prior density  $p(\boldsymbol{\pi})$ ,
- and a theoretical model for our statistic  $\Delta[\boldsymbol{\pi}]$ .

Only the second of these objects can be defined easily in our initial state of ignorance. Without any prior knowledge upon our model we may choose a uniform prior that is uniform over some region of our parameter space (there is a lot to say about priors, which we leave out for now).

The theoretical model and likelihood are more complex. A model must be ‘accurate’ in that it accurately represents the system being measured (e.g. the motion of a pendulum mass is expressed in a simple form only when the amplitude of the oscillation is very small).

The likelihood  $p(\Delta|\boldsymbol{\pi})$  is something altogether more delicate. Typically in cosmology, a Gaussian likelihood is assumed for a measurement, which amounts to assuming that the value measured scatters around the true value of the random variable being measured. This assumption must be validated, but may be statistically valid if a set of measurements in a survey region add up to a Gaussian distribution via the central limit theorem.

An example of where it would be incorrect to assume Gaussian uncertainties on a random variable would be for low numbers of photon counts. These are described instead by the Poisson distribution characterised by the rate parameter  $\lambda$ . However, at a large number of photons the Poisson distribution tends towards a Gaussian with mean  $\lambda$  and variance  $\lambda$ .

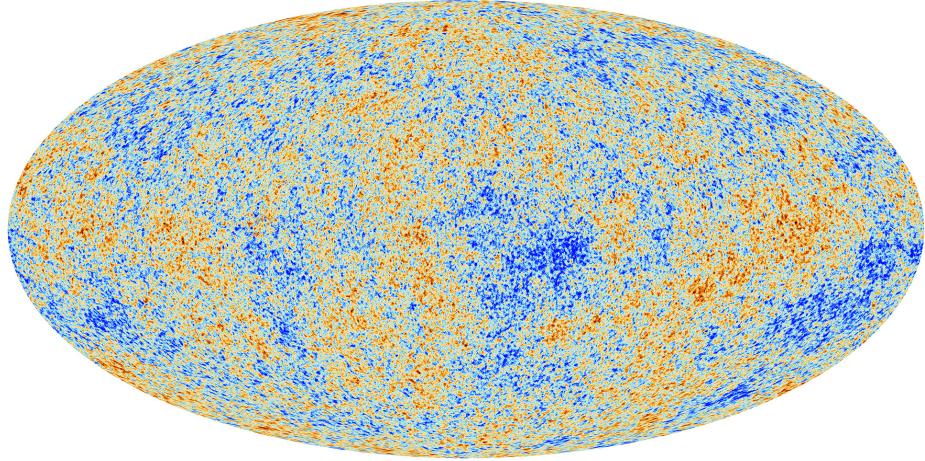


Figure 1: The cosmic microwave background as observed by Planck. This is a map of the oldest light in Universe, at a time when the Universe was around 380,000 years old. This is an example of a *realisation* of a random field. The amplitude of the fluctuations shown is of order  $10^{-5}$ .

So, with a model  $\Delta[\boldsymbol{\pi}]$  for a measurement  $\hat{\Delta}$ , we can find the parameters  $\boldsymbol{\pi}$  that fit to our data the best (via the minimisation of some objective function) by finding the location in parameter space that maximises our posterior probability  $p(\boldsymbol{\pi}|\hat{\Delta})$ . The posterior itself, as a PDF for the parameters given the data, provides the state of uncertainty upon the model parameters after measuring the data.

**Question:** *What is an example of an ‘objective function’ that quantifies the fit of a model (for some value of its parameters) to data? There are multiple potential answers.*

### 3 A simple model for the matter density field

Random fields are common in cosmology. The quantum mechanical origin of the density fluctuations in our Universe leads us to interpret observed fields in cosmology as realisations of random fields. Inflation predicts fluctuations in the primordial matter density field that are very nearly Gaussian, with a nearly scale-invariant power spectrum  $P(k) \approx Ak$  as a function of physical scale  $k$ . The property of Gaussianity in the fluctuations is preserved by linear evolution of the density field across cosmic time. Linear evolution of the density field is described by the growth equation for perturbations  $\delta(\mathbf{x})$  that are small in amplitude. The cosmic microwave background probes fluctuations mostly in the linear regime, and so the fluctuations look very Gaussian (see Figure 1).

However, at late times structure formation becomes non-linear<sup>1</sup> which destroys Gaussianity and gives the filamentary cosmic web.

So what are we doing in cosmology? Put simply, we want to understand the random process that gave rise to the Universe as we see it today. This description, due to the stochastic origins described above, is ultimately statistical in nature because we live in one realisation of the random process unfolding. Ultimately, we must use tracers of the underlying matter density field, which is not directly observable in its three-dimensional form, but any tool for statistical analysis must at least prove its effectiveness on a simplified case where we ignore the tracers of matter (such as galaxies) before a more realistic and complex application.

Here we discuss a simple model that is an analogy to the matter density of the Universe.

### 3.1 Gaussian probability distribution function

Multivariate Gaussian random variables have a PDF given by

$$\mathcal{G}[\mathbf{x}|\boldsymbol{\mu}, \mathbf{C}] = \frac{1}{\sqrt{|2\pi\mathbf{C}|}} \exp \left[ -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{C}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right], \quad (3)$$

where  $\boldsymbol{\mu} = \langle \mathbf{x} \rangle$  is the  $N$ -dimensional expectation value of the random variable  $\mathbf{x}$  and  $\mathbf{C}$  is the covariance matrix defined as

$$\mathbf{C} = \langle (\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T \rangle. \quad (4)$$

As you can see, the Gaussian PDF is described entirely by its expectation value  $\boldsymbol{\mu}$  and  $N \times N$  covariance matrix  $\mathbf{C}$ .

### 3.2 Gaussian random fields

Cosmological observations (from the Planck mission) and theoretical considerations indicate that the initial density contrast of our Universe is well described by a *Gaussian random field*<sup>2</sup> [5, 6]. A random field  $\delta(\mathbf{x})$  is a Gaussian random field if for any  $N$  points in real space the vector

$$\delta(\mathbf{x}) = (\delta(\mathbf{x}_1), \dots, \delta(\mathbf{x}_N)), \quad (5)$$

follows a Gaussian distribution as defined above<sup>3</sup>.

The random field is completely defined by its expectation value  $\mu(\mathbf{x})$  and by its covariance at every two locations  $\mathbf{x}_1$  and  $\mathbf{x}_2$

---

<sup>1</sup>Overdensities can grow indefinitely and will thereby become more and more attractive, while underdensities are bounded by zero (the density *contrast* is bounded by minus one).

<sup>2</sup>GRFs are also called Gaussian processes in one dimensional scenarios.

<sup>3</sup>More abstractly a random field is a random variable at each point in space, meaning our field is an infinitely long collection of random variables enumerated by each point in space. For GRFs the field at a discrete set of points in the field, are also multivariate Gaussian distributed.

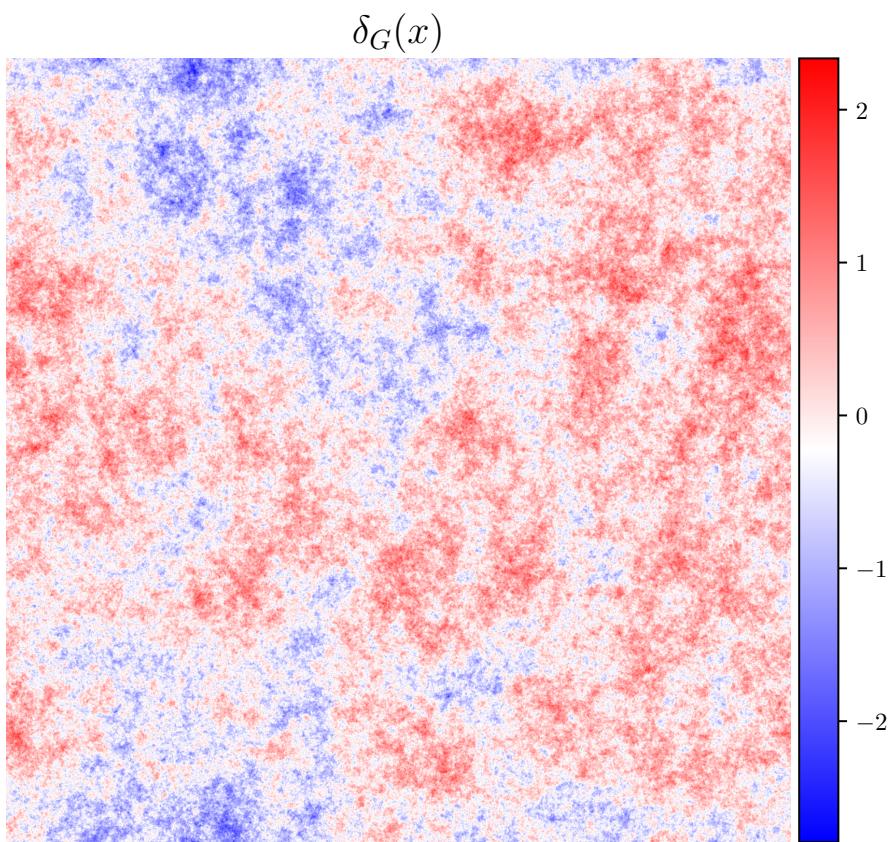


Figure 2: A Gaussian random field  $\delta(\boldsymbol{x})$  sampled with the power spectrum  $P(k) = 1.0 * k^{-3.0}$ .

$$C(\mathbf{x}_1, \mathbf{x}_2) = \langle (\delta(\mathbf{x}_1) - \mu(\mathbf{x}_1))(\delta(\mathbf{x}_2) - \mu(\mathbf{x}_2)) \rangle, \quad (6)$$

which is commonly known as the 2-point correlation function of the field which describes how density field values at different locations vary with each other - how they are correlated.

The density contrast is defined to have a mean of zero at all points ( $\mu(\mathbf{x}_i) = \langle \delta(\mathbf{x}_i) \rangle = 0$ ) as well as being statistically homogeneous and isotropic so that the covariance of the contrasts at two points is only a function of their absolute distance apart

$$\begin{aligned} C(\mathbf{x}_1, \mathbf{x}_2) &= C(\mathbf{x}_1 - \mathbf{x}_2), \text{ (translation invariance)} \\ &= C(|\mathbf{x}_1 - \mathbf{x}_2|). \text{ (rotation invariance)} \end{aligned} \quad (7)$$

But what exactly does ‘statistically isotropic and homogeneous’ mean? This definition concerns the  $N$ -point correlation functions  $\xi_\delta$  of the field  $\delta(\mathbf{x})$ , expressed as  $\langle \delta(\mathbf{x}_1) \cdots \delta(\mathbf{x}_N) \rangle$ . The field is statistically homogeneous and isotropic if

$$\begin{aligned} \xi_\delta(\mathbf{x}_1, \dots, \mathbf{x}_N) &\equiv \langle \delta(\mathbf{x}_1) \cdots \delta(\mathbf{x}_N) \rangle, \\ &= \langle \delta(\mathbf{y}_1) \cdots \delta(\mathbf{y}_N) \rangle. \end{aligned} \quad (8)$$

where  $\mathbf{y}_i = \mathbf{R} \cdot \mathbf{x}_i + \mathbf{a}$  for some rotation matrix  $\mathbf{R} \in \mathbb{R}^{N \times N}$  and translation vector  $\mathbf{a} \in \mathbb{R}^N$ . This means that the statistics of the field are *invariant* to translations and rotations.

For the GRFs we consider, the higher order correlation functions  $N > 2$  are functions of the variance of the PDF of the density fluctuations at each point in the field. The PDF of density fluctuations at a single point  $\mathbf{x}_i$  is known as the *one-point* PDF, calculated by marginalising the  $N$ -dimensional multivariate PDF of  $\delta(\mathbf{x})$  over all but one of the  $N$  points the field is sampled at. This is written as

$$p(\delta(\mathbf{x}_i)) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp \left[ -\frac{1}{2} \frac{(\delta(\mathbf{x}_i))^2}{\sigma_i^2} \right] \quad (9)$$

For this Gaussian PDF, all higher order moments only depend on the variance  $\sigma_i^2$ . This is known as *Isserlis’ Theorem*. The preceding definitions tell us that the physical process that created these fluctuations in the primordial density field of our Universe does not distinguish one point from another (homogeneity), nor does it prefer one direction over another (isotropy). In reality, the late-time density field has a one-point PDF that depends on higher-order moments of the field in a complex and unknown hierarchy.

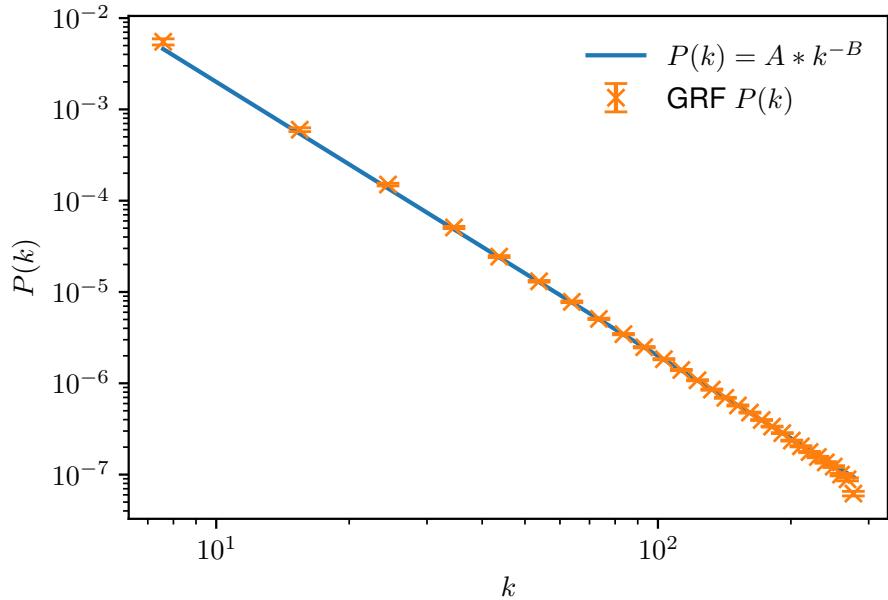


Figure 3: The power spectra calculated for an ensemble of fields generated at the true  $(A, B)$  parameters. Error bars are calculated over the independently generated power spectra of the fields. The theoretical power spectrum used to generate the fields is also shown. The power spectrum describes the variance of fluctuations in the field as a function of physical scale  $k$  (high  $k$  implies small scales).

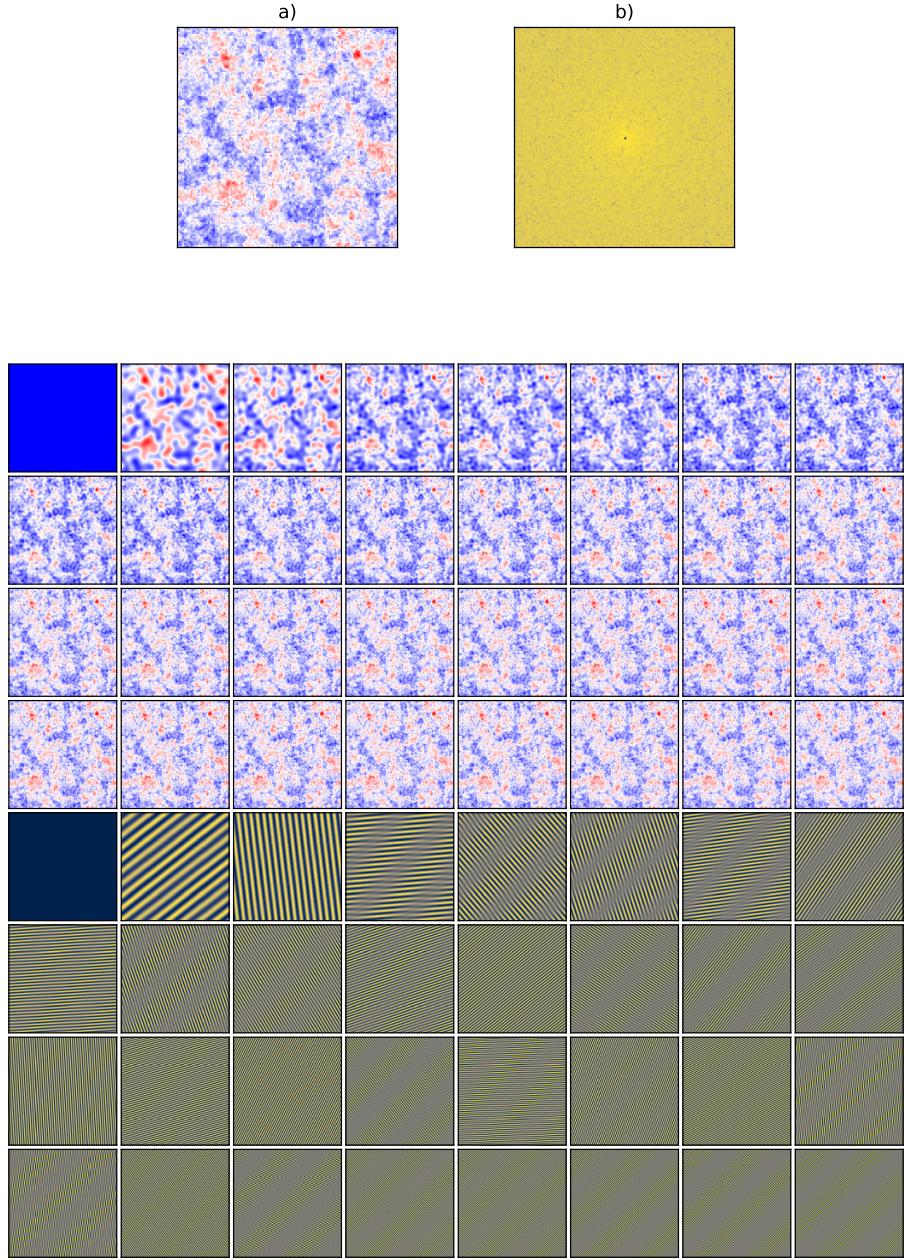


Figure 4: A GRF decomposed into a set of Fourier modes. a) the original GRF being decomposed b) the Fourier transform showing each of the Fourier modes (note that this is the real part of the complex set of modes). The top half of the grid shows the accumulating sum of the Fourier modes derived from the signal. The lower half of the grid shows the real part of each of the Fourier mode coefficients in increasing spatial frequency, where each of these plots represents two points in the space plotted in panel b) mirrored due to the reality condition  $\delta_{-\mathbf{k}} = \delta_{\mathbf{k}}^*$ .

### 3.3 Fourier space

As you will see, the representation of a function in Fourier space is of critical importance. Figure 4 shows one such representation in Fourier space for a GRF, where the field is decomposed into a Fourier space basis. We may consider a general field to be built from a superposition of a large number of plane waves, each representing a *mode* with an amplitude and phase depending on a linear transformation of the field. The tool for this job is of course the Fourier transform, assuming the Universe has a flat comoving geometry.

The Fourier transform of the matter density field is

$$\delta(\mathbf{k}) = \int d^N \mathbf{x} \delta(\mathbf{x}) \exp(-i\mathbf{k} \cdot \mathbf{x}) \quad (10)$$

where the covariance is given by

$$\tilde{C}(\mathbf{k}_1, \mathbf{k}_2) = \langle \delta(\mathbf{k}_1) \delta^*(\mathbf{k}_2) \rangle. \quad (11)$$

Since the Fourier transform is a linear operation, the field  $\delta(\mathbf{k})$  is also a Gaussian random field with an expectation value of zero.

$$\begin{aligned} \langle \delta(\mathbf{k}) \rangle &= \left\langle \int d^N \mathbf{x} \delta(\mathbf{x}) \exp(-i\mathbf{k} \cdot \mathbf{x}) \right\rangle \\ &= \int d^N \mathbf{x} \exp(-i\mathbf{k} \cdot \mathbf{x}) \langle \delta(\mathbf{x}) \rangle \\ &= 0 \end{aligned} \quad (12)$$

Since we defined the expectation value of  $\delta(\mathbf{x})$  to be zero. The linearity of the Fourier transform means that all the information in the original field is encoded in the Fourier transformed field. This is because the statistical description, encoded in the distribution function of the field itself, is Gaussian and a linear operation on a Gaussian variate produces another Gaussian variate.

**Question:** *Can you show that  $\mathbf{y} = A\mathbf{x} + \mathbf{b}$  means that  $\mathbf{y}$  is Gaussian distributed if  $\mathbf{x}$  is Gaussian distributed?*

The density field is a real valued field, meaning that we have, for the Fourier coefficients  $\delta(\mathbf{k})$ , that  $\delta(-\mathbf{k}) = \delta^*(\mathbf{k})$ . Therefore  $\delta(\mathbf{k})$  and  $\delta(-\mathbf{k})$  contain the same information since they are just a conjugation apart. This means that the two modes  $\delta(\mathbf{k})$  and  $\delta(-\mathbf{k})$  are the same mode.

It can be shown that the covariance of the Fourier transform of the density field is expressed as

$$\begin{aligned} \tilde{C}(\mathbf{k}_1, \mathbf{k}_2) &= \langle \delta(\mathbf{k}_1) \delta^*(\mathbf{k}_2) \rangle, \\ &= (2\pi)^3 \delta^D(\mathbf{k}_2 - \mathbf{k}_1) P(k_2), \end{aligned} \quad (13)$$

where  $k_i$  is the magnitude of  $\mathbf{k}_i$  which holds in the equation above due to the isotropy of the Gaussian random field, that is, if  $C(\mathbf{x})$  depends only on  $|\mathbf{x}|$  as in Equation 6. The function  $\delta^D$  is the Dirac-delta function. In this equation,  $P(k)$  is the *power spectrum* of the field. This function gives the variance of matter density fluctuations as a function of physical scale  $k$  and is of central importance in modern cosmology.

**Question:** Derive Equation 13 (Hint: this is a standard problem in cosmology, you may need to change variables to a translation of one point around another).

Meditate for a moment on the definition in Equation 13; the power spectrum defined in this equation gives the covariance between two Fourier modes, and is zero if they are not the same mode. Therefore, in Fourier space, the covariance of the different modes is zero, meaning the covariance is a diagonal matrix for the transformed  $N$ -dimensional field  $\delta(\mathbf{x})$ .

Another important point is that, since  $\langle \delta(\mathbf{k}) \rangle = 0$ , the power spectrum  $P(k)$  is all we need to describe the field in real and Fourier space.

The PDF for the Fourier transformed density field  $\delta(\mathbf{k})$  is given by

$$p(\delta(\mathbf{k}_1), \dots, \delta(\mathbf{k}_N)) = \frac{1}{\sqrt{|2\pi\tilde{C}|}} \exp \left[ -\frac{1}{2} \sum_{i,j=1}^N \delta(\mathbf{k}_i) [\tilde{C}(\mathbf{k}_i, \mathbf{k}_j)]^{-1} \delta^*(\mathbf{k}_j) \right], \quad (14)$$

or, using the power spectrum  $P(k)$  (input as the matrix  $P(k_i)\delta_{ij}^D$ )

$$p(\delta(\mathbf{k}_1), \dots, \delta(\mathbf{k}_N)) = \frac{1}{\sqrt{(2\pi)^N \prod_{i=1}^N P(k_i)}} \exp \left[ -\frac{1}{2} \sum_{i=1}^N \frac{|\delta(\mathbf{k}_i)|^2}{P(k_i)} \right]. \quad (15)$$

In this lab, we use a simple power law model for the power spectrum of matter density fluctuations in our Universe is defined by two parameters  $A$  and  $B$ , expressed as

$$P(k) = A * k^{-B}. \quad (16)$$

The parameters  $A$  and  $B$  correspond to the amplitude of the power spectrum and the scale-dependency of the modes  $k$ . The index  $B$  governs the balance between large- and small-scale power. For a fixed pair of parameters  $A$  and  $B$  this form of the power spectrum is known as *scale-invariant* because the covariance of the density field at two points has the same value independently of their separation; that is,  $\langle \delta(\lambda\mathbf{x})\delta(\lambda\mathbf{y}) \rangle = \langle \delta(\mathbf{x})\delta(\mathbf{y}) \rangle$  for all  $\lambda \in \mathbb{R}$ .

The motivation for the power law form, in the absence of a physical theory, means we do not assume that the spectrum contains any preferred length scale. If there was a unique scale we would need to explain the physics of this preference. Consequently, the spectrum must be a featureless power law. Figure 3

shows this form of the power spectrum for a given amplitude and index against a power spectrum measured for a set of fields generated at the same parameters.

As we previously saw, Equation 13 shows us that the form of the PDF of the field in Fourier space has a diagonal covariance matrix. This promotes a simpler form for the likelihood of the field in Fourier space - a simple product of independent Gaussians for each mode. Given the isotropy of the mode covariance, the likelihood of a randomly sampled GRF, denoted as a measurement by  $\hat{\delta}(\mathbf{x})$ , when transformed to Fourier space as  $\hat{\delta}(\mathbf{k})$  is given by

$$p(\delta(\mathbf{k}_1, \dots, \delta(\mathbf{k}_N)) | \boldsymbol{\pi}) = \prod_{i=1}^N \mathcal{G}[\delta(\mathbf{k}_i) | 0, P(k_i; \boldsymbol{\pi})] \quad (17)$$

which conserves the information of the original field since the statistical descriptions denoted by the probability distributions in real and Fourier space are linearly related by the Fourier transform.

Let's take a moment to unpack the likelihood of the GRF in Equation 17. This is a function of the Fourier transform of a set of points in a GRF such as that in Figure 2. Here we have  $k_i = |\mathbf{k}_i|$  from the statistical isotropy and homogeneity of the field which causes the radial symmetry in  $k$ -space shown in Figure 5. The power spectrum evaluated at each  $\mathbf{k}_i$  mode is explicitly written as a function of the model parameters for the power spectrum in Equation 16. The likelihood is a product of individual Gaussians because, given Equation 13, we see that individual modes are independent (e.g.  $P(x, y) = P(x)P(y)$ ). This is our statistical model for the GRF measurement.

### 3.4 How to generate a Gaussian random field

In this lab we can only generate fields of a finite extent. The preceding description of the fields in the continuous limit holds only if the fields are infinitely large in real space. Whilst the field in real space is still a superposition of many modes, we assume periodic boundary conditions <sup>4</sup>.

Here we generate two-dimensional fields of physical length  $L = 1.0$  and ‘volume’  $V = L^2 = 1.0$  in arbitrary units. The idea is to create a field in Fourier space, that when inverse transformed, will have the correct statistics in real space. This unfolds in 4 steps:

- Create a grid of  $k$ -space frequencies dictated by the dimensions and number of pixels in the real space field we desire,
- we then generate an initial white-noise field is generated by sampling independent Gaussian noise with zero mean and unit variance for the real and imaginary components for each Fourier mode  $\delta_{\mathbf{k}} = A_{\mathbf{k}} + iB_{\mathbf{k}}$  of each frequency point in  $k$ -space,

---

<sup>4</sup>This does not mean the field is periodic in reality, we are only interested in the field contained in the volume  $V$ . The finite volume restricts the the Fourier space to a discrete set of points, whilst the field itself still has infinitely many points in real space.

- this must be scaled by the correct covariance which in  $k$ -space is the power spectrum evaluated at each point (since this white-noise field has a power spectrum  $P(k) = 1$ ). This covariance (panel b) in Figure 5) is used to multiply the white-noise field to give a field in Fourier space. This is shown in panel c) in Figure 5,
- finally the complex  $k$ -space field is inverse Fourier transformed to give the GRF in panel d) of Figure 5.

Note that this is actually a *discrete* Fourier transform [10], which is subtly different to what we discussed before but not for the underlying physics. The key is that generating a white noise realization is easy since spatially separated points are uncorrelated ( $P(k) = 1 \forall k$ ), and so the white-noise field is generated by independently drawing each pixel value from a normal distribution.

### 3.5 Fisher information matrix

The Fisher information matrix is defined as

$$\begin{aligned} F_{\alpha\beta}(\boldsymbol{\pi}) &= \left\langle \frac{\partial \log p(\mathbf{x}|\boldsymbol{\pi})}{\partial \pi_\alpha} \frac{\partial \log p(\mathbf{x}|\boldsymbol{\pi})}{\partial \pi_\beta} \right\rangle, \\ &= -\left\langle \frac{\partial^2 \log p(\mathbf{x}|\boldsymbol{\pi})}{\partial \pi_\alpha \partial \pi_\beta} \right\rangle, \end{aligned} \quad (18)$$

and it is the expected variance of the derivative of the log-likelihood of the data  $\mathbf{x}$  with respect to the model parameters at a fixed point in parameter space  $\boldsymbol{\pi}$ <sup>5</sup>. The ‘information’ here refers to the variance of the log-likelihood. The more information the data  $\mathbf{x}$  contains on the data, the more peaked the likelihood is at that point in parameter space - for that particular parameterisation of the model.

**Question:** Over which distribution is the expectation value in the Fisher information calculated over?

Lets look at a simple case with a simple 1-dimensional random variable  $x$  drawn from a Gaussian distribution  $\mathcal{G}[x|\mu(\boldsymbol{\pi}), \sigma^2]$  with a known variance  $\sigma$  and an expectation value  $\mu(\boldsymbol{\pi})$ . The expectation value here depends on the parameters in some way. The Fisher information is expressed as

$$\begin{aligned} F(\boldsymbol{\pi}) &= -\left\langle \frac{\partial^2 \log p(x|\boldsymbol{\pi})}{\partial \pi^2} \right\rangle \\ &= \left( \frac{\partial \mu(\boldsymbol{\pi})}{\partial \boldsymbol{\pi}} \right)^2 \frac{1}{\sigma^2} \end{aligned} \quad (19)$$

---

<sup>5</sup>Note that the step between the equations in Equation 18 depends on certain regularity conditions.

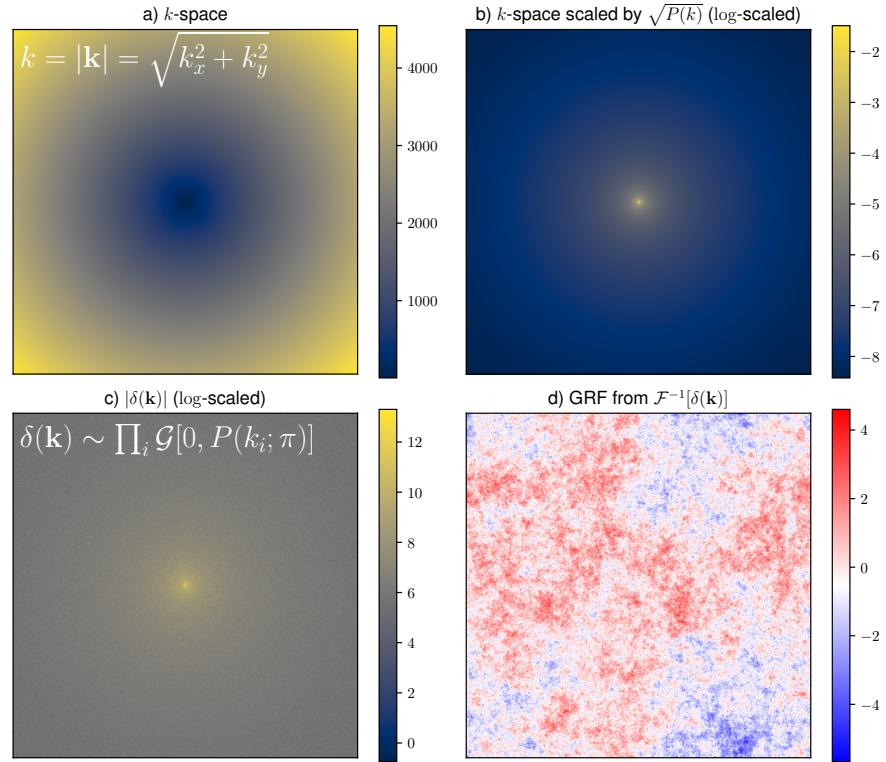


Figure 5: A figure showing the steps to generate a GRF. a) Calculating the  $k$ -mode magnitudes for the homogeneous and isotropic field. b) The power spectrum evaluated over the  $k$ -space grid, which will scale a white noise field in Fourier space to have the correct covariance. c) The real part of the  $k$ -space complex white noise field scaled by the covariance. d) the GRF resulting from taking the real part of the Fourier transform of the complex  $k$ -space field in c).

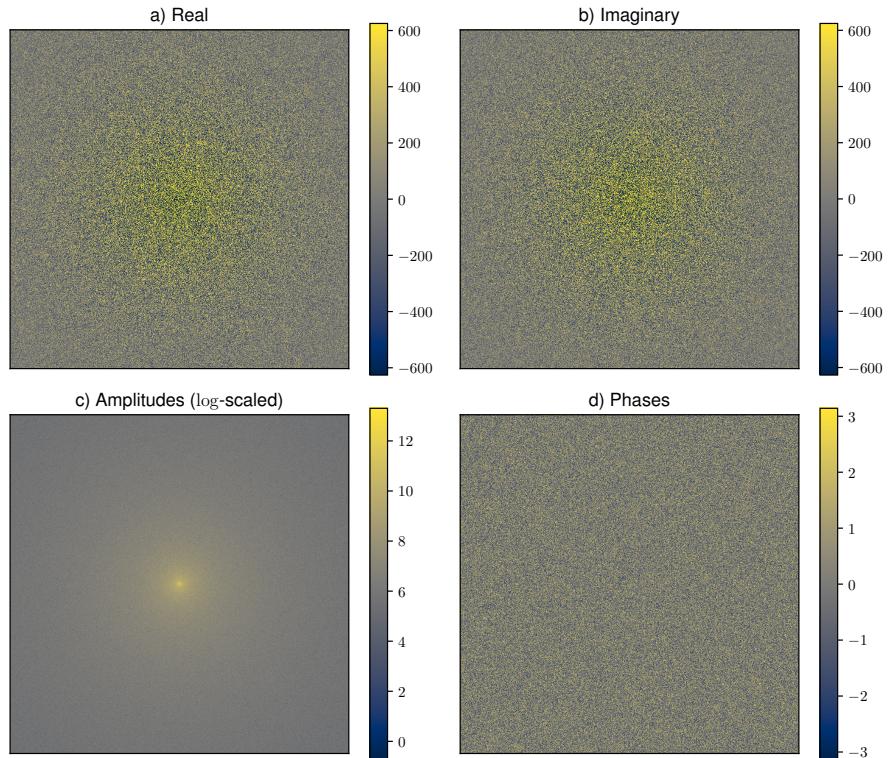


Figure 6: A figure showing the constituents of a Fourier transformed field in Fourier space. The real and imaginary parts of each mode  $\delta_{\mathbf{k}} = A_{\mathbf{k}} + iB_{\mathbf{k}}$  are plotted in panels a) and b), as well as the amplitudes and phases of the  $\delta_{\mathbf{k}} = r_{\mathbf{k}} \exp(-i\theta_{\mathbf{k}})$  of each mode in panels c) and d). Note that the ranges of the real and imaginary plots correspond to the 10% and 90% percentiles for either component respectively.

where in the first line we assume that the log-likelihood is twice differentiable. This expression shows us that the Fisher information of the data upon the model parameter  $\pi$  is exact only if the model is linear in the parameters  $\pi$ .

**Question:** *To what extent is the Fisher information quantifying the information content that the data contains on a model? Are assumptions being made about the complexity of the model in terms of the parameters?*

### 3.6 The Fisher information for the GRF likelihood

The Fisher information of our Gaussian random field data can be calculated because we know the form of the likelihood, a multivariate Gaussian, and we have the true model for the covariance of this likelihood (in Fourier space, noting that the expectation value parameterising this likelihood is zero, see [12]). Therefore we can express the information content of the field on our model parameters. The expression for the Fisher information of a Gaussian random field in Fourier space,  $\delta(\mathbf{k})$ , is given by

$$F_{\alpha\beta} = \frac{1}{2} \text{Tr}[C^{-1} C_{,\alpha} C^{-1} C_{,\beta}] \quad (20)$$

where the covariance is the covariance between independent modes in Fourier space. Note also that  $C_{,\alpha} = (\partial/\partial\pi_\alpha)C$ . Since the derivative of the expectation value in the likelihood is independent of the parameters (it is zero). Since the Gaussian field has a simple representation in Fourier space given by the power spectrum, the covariance matrix in Fourier space is given by

$$C(\mathbf{k}_i, \mathbf{k}_j) = P(k_i; \boldsymbol{\pi})\delta_{ij}, \quad (21)$$

and the derivatives of the covariance matrix with respect to the model parameters  $\boldsymbol{\pi} = (A, B)$  are therefore

$$\begin{aligned} \frac{\partial P(k; \boldsymbol{\pi})}{\partial A} &= k_i^{-B} \delta_{ij}, \\ \frac{\partial P(k; \boldsymbol{\pi})}{\partial B} &= -Ak_i^{-B} \log k_i \delta_{ij}. \end{aligned} \quad (22)$$

These equations are substituted into Equation 20 to define the Fisher information of the GRF in Fourier space. Our equation for the Fisher information of a Gaussian likelihood with model parameter dependent covariance tells us only these first derivatives are required.

The derivation of the Fisher information of the Gaussian random field likelihood uses the fact that the Fourier transform conserves the information content of either field (and the transformation gives us a simple distribution we can compute likelihoods with). The representation in Fourier space has a diagonal covariance matrix which makes the calculation much more simple.

**Question:** Derive Equation 20. You may find [12] useful. Also note that  $C^{-1}C = \mathbb{I} \implies \mathbb{I}_{,\alpha} = 0 \implies C_{,\alpha}C^{-1} + CC_{,\alpha}^{-1} = 0$ .

## 4 Simulation-based inference

For many applications, the likelihood function is not known. One can assume a form such as a Gaussian, but to calculate the likelihood accurately the expectation value and covariance must be known accurately. Since one or both may depend on the model parameters, they must be known accurately for each parameter realisation to register the correct fit when the likelihood is applied with measured data. Without an accurate likelihood function the information we can extract upon our models from data is distorted due to inaccuracies in a statistical model (even if the statistical model uses the correct distribution for the uncertainty).

Simulation-based inference (SBI) [2] is a method for fitting a likelihood function for Bayesian inference using simulations instead of approaching the problem from first-principles or estimating the parameters of a distribution. These methods typically don't assume the form of a likelihood, but instead fit a likelihood function to a set of simulations of the data. The simulations themselves are the output of a probabilistic simulator which is usually assumed to be accurate in representing the signal in a noisy measured datavector. You can think of the simulation as a theory prediction that results from inputting the model parameters into the program that simulates the data.

**Question:** Can you think of any problems with the approach outlined above? Is it useful to have an analytic likelihood you can write down?

### 4.1 Density estimation for SBI

In this lab we will use density-estimation SBI [8, 7]. Put simply, this class of methods for SBI uses neural networks to estimate the likelihoods of simulation-parameter pairs, particularly when the form of the likelihood is not known. This allows us to do a Bayesian inference with a likelihood that is not based on assumptions about the data. Also we avoid using a likelihood made from assuming a distribution and estimating the parameters of this distribution from simulations. Due to the flexibility of the models used to fit the complex data distributions, we can also fit the posterior directly. There is as of yet no preferred way between fitting the likelihood or the posterior.

### 4.2 Normalizing flows

A normalizing flow [13] is, put simply, a change-of-variables between two probability distributions parameterised by a neural network. What does this mean? Suppose a variable  $\mathbf{x}$  with PDF  $p(\mathbf{x})$  is mapped to another variable  $\mathbf{y} = f(\mathbf{x})$

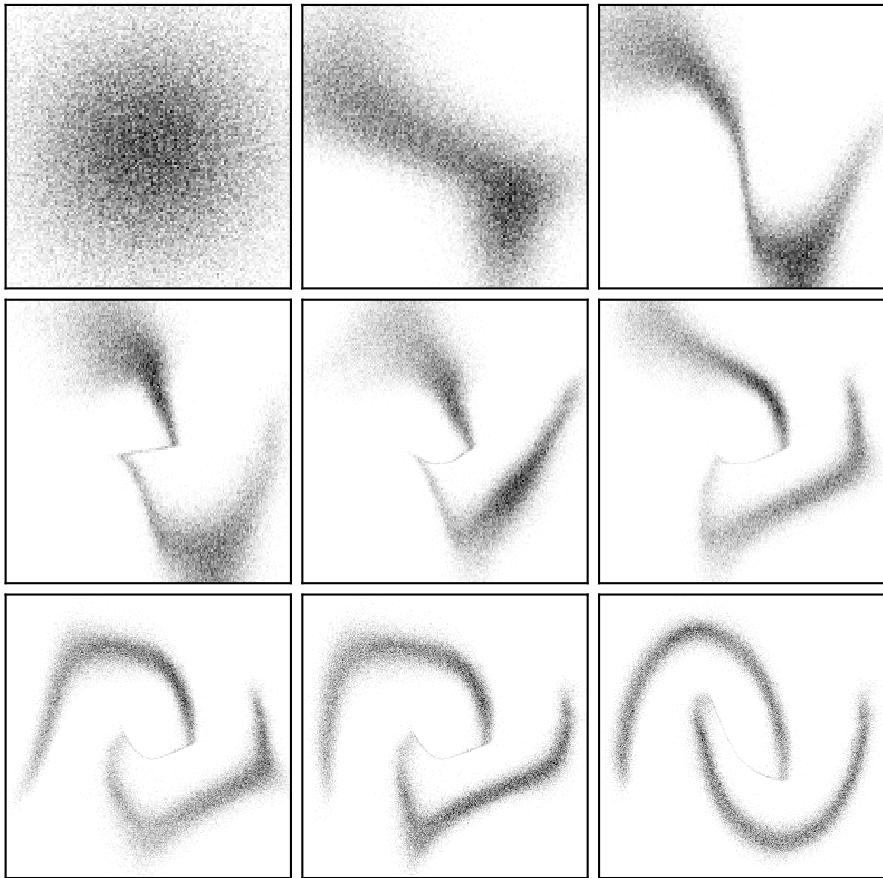


Figure 7: A figure showing a change of variables between a multivariate Gaussian (samples  $\mathbf{z}$ , top left) and a non-linear distribution (samples  $\mathbf{x}$ , bottom right) parameterised by a normalizing flows. Each plot shows one of the transforms in the flow, applied to the Gaussian samples initially, and then to the output of the step before it, for each transformation in the sequence of the normalizing flow.

by a function  $f$ . Both of these variables sit in  $N$ -dimensional space. The change of variables would be written

$$p(\mathbf{y}) = |\det(\mathbf{J}[f(\mathbf{x})])| p(\mathbf{x}) \quad (23)$$

where the Jacobian  $\mathbf{J}[f(\mathbf{x})]$  is a matrix of derivatives of the function  $f$  with respect to the inputs  $\mathbf{x}$ . The matrix would be expressed as

$$\mathbf{J}[f(\mathbf{x})] = \begin{bmatrix} \frac{\partial f_i(\mathbf{x})}{\partial \mathbf{x}_j} & \cdots & \frac{\partial f_i(\mathbf{x})}{\partial \mathbf{x}_0} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_0(\mathbf{x})}{\partial \mathbf{x}_N} & \cdots & \frac{\partial f_N(\mathbf{x})}{\partial \mathbf{x}_N} \end{bmatrix}. \quad (24)$$

In Figure 7 you can see an example of a change of variables between two distributions. The initially multivariate Gaussian variables are ‘warped’ through a function in each step of the flow displayed on the plot. This mapping between two PDFs that depends on points  $\mathbf{x}$  and  $\mathbf{y}$  is being used to map between two distributions, one known, and one unknown.

In SBI methods, typically a normalizing flow is used to fit the unknown likelihood  $p(\mathbf{x}|\boldsymbol{\pi})$  (or posterior  $p(\boldsymbol{\pi}|\mathbf{x})$ ) from a set of simulations and parameters  $\{\mathbf{x}, \boldsymbol{\pi}\}$ . The key point is that in deriving a likelihood directly from simulations of data, that may include hard-to-model effects, we do not make any assumptions on the form of the likelihood.

The next question would be, how do we calculate the likelihood of a simulation  $\mathbf{x}$  given its parameters  $\boldsymbol{\pi}$  without knowing the likelihood? This is exactly why we need a change of variables.

We can take a simple distribution such as a Gaussian, with which we can calculate the probability of an  $N$ -dimensional point  $\mathbf{z}$  as

$$\mathcal{G}[\mathbf{z}|\mathbf{0}, \mathbb{I}] = \frac{1}{\sqrt{2\pi}} \exp \left[ -\frac{1}{2} \mathbf{z}^T \mathbf{z} \right], \quad (25)$$

where  $\mathbb{I}$  is the identity matrix and the expectation value of  $\mathbf{z}$  is zero. Using this ‘base’ distribution, we can calculate the log-probability of a point  $\mathbf{x} = f(\mathbf{z})$  with an unknown PDF using a change-of-variables via

$$\log p(\mathbf{x}) = \log |\det(\mathbf{J}[f(\mathbf{z})])| + \log \mathcal{G}[\mathbf{z}|\mathbf{0}, \mathbb{I}], \quad (26)$$

and so an estimate of the probability of  $\mathbf{x}$  with the unknown likelihood depends only on the base distribution (a known, fixed quantity) and the function  $f(\mathbf{z})$ .

**Question:** Can you express the equations above with a conditional dependence on the physical model parameters  $\boldsymbol{\pi}$ , as if to model the likelihood  $p(\mathbf{x}|\boldsymbol{\pi})$ ? How are these conditional variables processed by the normalizing flow?

### 4.2.1 A simple change-of-variables

Let's take a simple example. Suppose we have a one-dimensional random variable  $x = f(z)$  where  $x = az + b$ . If  $z$  is drawn from a unit Gaussian  $\mathcal{G}[z|\mu, \sigma^2]$ , then because  $x$  is a linear function of  $z$ ,  $x$  is also Gaussian distributed with  $p(x) = \mathcal{G}[x|a\mu + b, a^2\sigma^2]$ .

The inverse mapping  $f^{-1}(x)$  maps a point  $x$  back to a point  $z$  as  $(x - b)/a$ . The Jacobian of this change-of-variables is then simply  $\partial f^{-1}(x)/\partial x = 1/a$ . The log-probability of  $x$  under the ‘unknown’ PDF  $p(x)$  is expressed as

$$\log p(x) = \log \frac{1}{a} + \log \mathcal{G}\left[\frac{x - b}{a} \middle| \mu, \sigma^2\right]. \quad (27)$$

In this simple example, the neural network we use to parameterise the Jacobian would simply have to fit the function  $z = f^{-1}(x)$ , where the model could simply estimate  $a$  and  $b$  from a set of  $x$  drawn from  $p(x)$ . The Jacobian here depends on the fact we have a linear mapping between the two distributions, and in more complicated cases where  $p(x)$  is not simply Gaussian, this will not be very useful.

### 4.2.2 Optimising a normalizing flow

In a more complex setting where we do not know the form  $p(\mathbf{x})$  for a high-dimensional variable  $\mathbf{x}$ , we fit a normalizing flow  $p_\phi(\mathbf{x})$  to a set of data<sup>6</sup>  $\{\mathbf{x}_i, \dots, \mathbf{x}_N\}$  by *maximising the log-probability of the data under our normalizing flow model distribution*. The distribution  $p_\phi(\mathbf{x})$  is parameterised by the parameters of the neural network that map from  $\mathbf{x}$  to  $\mathbf{z}$ . The Kullback-Leibler (KL) divergence quantifies how well one probability distribution represents another, and so it is a natural objective to minimise (with respect to the parameters  $\phi$ ) when we fit the normalizing flow distribution to a set of data points. We write this objective as

$$\begin{aligned} \mathcal{D}_{\text{KL}}(p(\mathbf{x}) || p_\phi(\mathbf{x})) &= \int d\mathbf{x} p(\mathbf{x}) \log \frac{p_\phi(\mathbf{x})}{p(\mathbf{x})} \\ &= \int d\mathbf{x} p(\mathbf{x}) [\log p_\phi(\mathbf{x}) - p(\mathbf{x})] \\ &= \int d\mathbf{x} p(\mathbf{x}) \log p_\phi(\mathbf{x}) + C \\ &\approx \frac{1}{N_s} \sum_{i=1}^{N_s} \log p_\phi(\mathbf{x}). \end{aligned} \quad (28)$$

In the 3rd line we drop a term that is independent of the parameters  $\phi$  and so is zero in any gradient calculation with respect to  $\phi$  and in the last line we

---

<sup>6</sup>This set of  $N$  data are drawn identically and independently from the same distribution with model parameters sampled from a prior.

approximate the integral via a Monte-carlo estimate using a batch of data of length  $N_s$ .

The batches of data are drawn from a set sampled from the unknown  $p(\mathbf{x})$  and at each iteration we maximise the log-probability via stochastic gradient descent. This is known as *maximum-likelihood* training because we maximise the likelihood of the data by varying the parameters  $\phi$  of the normalizing flow (its ‘weights and biases’).

**Question:** *What exactly is the KL-divergence? Why is it useful? Can you calculate it for the one-dimensional change of variables we demonstrated?*

#### 4.2.3 NVP normalizing flows

Much of the research into normalizing flows depends on exactly how to calculate the Jacobian efficiently in high-dimensional spaces. There are many ways to do this but we will use one of the simplest forms, the RealNVP [3] normalizing flow, or NVP for short.

The NVP flow describes a method of building an invertible neural network that bijectively maps between  $\mathbf{z}$  and  $\mathbf{x}$ ; meaning  $\mathbf{x} = f_\phi(\mathbf{z})$  and  $\mathbf{z} = f_\phi^{-1}(\mathbf{x})$ . The functions are made using a simple affine (linear) operation that only affects half of the input dimensions at a time. The other half of the input vector is used to compute the shift and scale parameters of the linear operation on the first half of the input. This is written as

$$\begin{aligned}\mathbf{x}_{1:h} &= \mathbf{z}_{1:h}, \\ \mathbf{x}_{h+1:D} &= \mathbf{z}_{h+1:D} \odot \exp[s_\phi(\mathbf{x}_{1:h})] + t_\phi(\mathbf{x}_{1:h}).\end{aligned}\tag{29}$$

where  $D$  is the dimensionality of  $\mathbf{x}$ ,  $h = D/2$  and  $\odot$  is the element-wise product. The reverse computation is as simple as our linear example before and is written

$$\begin{aligned}\mathbf{z}_{1:h} &= \mathbf{x}_{1:h}, \\ \mathbf{z}_{h+1:D} &= (\mathbf{x}_{h+1:D} - t_\phi(\mathbf{z}_{1:h})) \odot \exp[-s_\phi(\mathbf{z}_{1:h})].\end{aligned}\tag{30}$$

Where  $\odot$  denotes an *element-wise* product. Note a very important feature of this design. The NVP flow is invertible, but built using non-invertible neural networks, which can compute the parameters of the forward and reverse operations. This is very important because these neural networks can fit complicated mappings based on the high-dimensional input data and are not restricted in size and complexity. Again; these two sets of equations define our invertible and bijective mapping between  $\mathbf{z}$  and  $\mathbf{x}$ .

The Jacobian of the mapping  $\mathbf{x} = f_\phi(\mathbf{z})$  is given by an  $D \times D$  matrix with entries

$$J[f_\phi(\mathbf{z})] = \begin{bmatrix} \mathbb{I}_h & \mathbf{0}_{h \times (D-h)} \\ \frac{\partial \mathbf{x}_{h+1:D}}{\partial \mathbf{z}_{1:h}} & \mathbb{I}_{D-h} \exp[s_\phi(\mathbf{z}_{1:h})] \end{bmatrix}. \quad (31)$$

Where  $\mathbb{I}_d$  is the identity matrix of dimension  $d$ . This is certainly not easy to interpret, but it is easily derived from the initial expression of  $\mathbf{x}$  in terms of  $\mathbf{z}$ . The most important point to take away from the Jacobian here is that it is a *lower-diagonal* matrix, meaning that its determinant can be computed as the product of the elements on the diagonal of the matrix. This operation is linear in dimensionality of the data which is not true for a general square matrix. The determinant of the Jacobian which we use in the log-probability calculation for the flow is given by

$$\begin{aligned} \log \det(\mathbf{J}[f_\phi(\mathbf{z})]) &= \log \prod_{i=h}^{D-d} \exp[s_\phi(\mathbf{z}_{1:h})] \\ &= \sum_{i=h}^{D-d} s_\phi(\mathbf{z}_{1:h}). \end{aligned} \quad (32)$$

In practise it is much easier to split the variables using a masking strategy, where ones and zeros within the masking matrix  $M$  dictate the splitting in the linear operation for the NVP flow defined in the previous equations. This is expressed as

$$\mathbf{x} = M \odot \mathbf{z} + (1 - M) \odot [\mathbf{z} \odot \exp[s(M \odot \mathbf{z})] + t(M \odot \mathbf{z})]. \quad (33)$$

**Question:** Can you derive the Jacobian matrix above? The derivative is zero if the function being differentiated is not a function of the variable the derivative is taken with respect to e.g.  $\partial f(\mathbf{x}_{1:h})/\partial \mathbf{x}_{h:D} = 0$ .

#### 4.2.4 More layers

The above derivation depends on a single layer in our normalizing flow, whilst this shallow model may work for some distributions, it is useful to stack a sequence of non-linear transforms to model more complex distributions. This is equivalent to simply nesting a sequence of functions  $\mathbf{x} = f_\phi(\mathbf{z})$ , such as

$$\mathbf{x} = f_K \circ f_{K-1} \circ \dots \circ f_1(\mathbf{z}) \quad (34)$$

**Question:** Can you write down the Jacobian for the normalizing flow with  $K$  sequential transforms (the output of one transform is the input of another in the sequence)?

## 5 Summary

Let us summarise what we know so far.

- We have a simple model for matter density fluctuations in the Universe given by a statistically homogeneous and isotropic Gaussian random field.
- We generate a field realisation by sampling a white noise field in Fourier space and scaling the variance of its Fourier modes by the a power spectrum, and then inverse Fourier transforming the field to real space.
- The power spectrum depends on parameters  $\pi = (A, B)$  that set the amplitude of the spectrum and the index of the power-law dependence of the variance of each mode with scale  $k$ .
- We wrote down an analytic likelihood for the probability of a Gaussian random field given a power spectrum (or equivalently, its parameters). We know this likelihood in Fourier space conserves the information of the likelihood in real space, but has a much simpler form.
- We discussed one method used for SBI, a process of fitting a likelihood function to a set of simulations, called ‘density estimation’. This requires a model with parameters that can be tuned - we chose a normalizing flow.
- The normalizing flow is a stacked sequence of coordinate transforms between two probability densities parameterised by neural networks, one of the densities we know (and can sample from), the other we don’t.

In short, we have a method to test the fitting of a likelihood function, made without assumptions on the data generating process, through comparing a posterior derived with this likelihood to the analytic posterior with the Fourier space GRF likelihood.

## 6 Testing SBI

### 6.1 Working with a normalizing flow

In the jupyter notebook `nvp_flow.example.ipynb` you will find a minimal implementation of an NVP normalizing flow [3]. This flow is being fit to the same distribution shown in Figure 7.

- Run the script and compare the equations in the NVP flow description in these notes to the code. The  $s_\phi(\cdot)$  and  $t_\phi(\cdot)$  networks are labelled in the code.
- Can you incorporate conditional information into the normalizing flow in this example to model  $p(\mathbf{x}|\mathbf{c})$  ( $\mathbf{c}$  is the category of the point in the spiral distribution denoting whether a point belongs to the upper spiral or the lower spiral)?
  - You might want to express the equations in the normalizing flow section of these notes in terms of a conditioning variable  $\pi$  or  $\mathbf{c}$ .

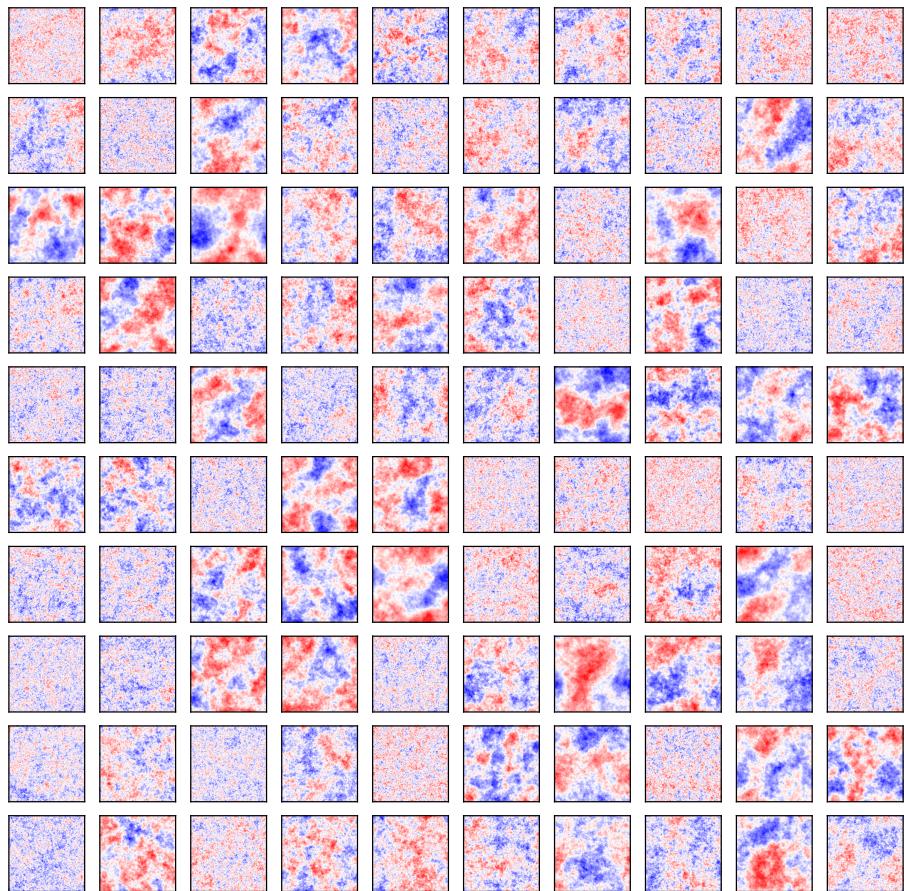


Figure 8: A set of GRFs generated with random power spectrum amplitudes and indexes generated from a uniform prior.

- Can you write a validation step in the training? This involves splitting the dataset into two subsets of training and validation fields and parameters.

## 6.2 Generate a GRF dataset

To fit our likelihood model we require a set of simulations and parameters  $\{\xi, \pi\}$ . We will create a uniform prior in both our parameters  $\pi = (A, B)$  and sample  $N$  parameter samples to generate our fields at.

Open the jupyter notebook `generate_grfs.ipynb`. Here you will find a set of functions that can generate GRFs given a sample of model parameters sampled from a prior. You can also calculate the power spectrum of the GRFs.

- Generate a training set by sampling from a prior. The prior should be uniform in  $A$  and  $B$ .
- Also generate a separate field that is not part of your training dataset. The parameters of this field should still be within the prior! This will be your observed data.
- To compare the SBI approach to the analytic approach we need to calculate likelihood of a GRF given a sample of parameters  $\pi$ .
  - Fourier transform a GRF realisation to  $k$ -space to obtain a grid like panel c) of Figure 4.
- Create a grid of evenly spaced samples (e.g. using `np.meshgrid[lowA : highA : 100j, lowB : highB : 100j]`) and see if you can evaluate the analytic likelihood and the normalizing flow likelihood over each point in the grid *given* a ‘measured’ GRF.

## 6.3 Fit a normalizing flow to a GRF dataset

In the jupyter notebook `fit_flow.ipynb` you will find an implementation of an NVP normalizing flow that is different to the first implementation. This implementation requires additional components in its architecture to deal with higher dimensional data. We will use this model to fit the posterior of  $\pi$  given a noisy measurement of a matter density field  $\delta(\mathbf{x})$ .

To make matters only a *bit* more complicated, due to the high dimensionality of the GRFs, the flow is fit to a set of compressed GRFs. These GRFs are compressed via an autoencoder [1] model (this is inspired by StableDiffusion [11]). As you will see, the autoencoder has to be pretrained.

- Fit the autoencoder and write a validation loop for its training.
- How is ‘new’ data generated with the flow-autoencoder setup?

## 6.4 Two ways of inference

To compare the result of calculating a posterior fit with simulations using a normalizing flow, we must calculate the analytic likelihood of the GRF. As we know, the GRF likelihood in real space is equivalent to the likelihood of the Fourier transformed field in  $k$ -space. The field in  $k$ -space is a Gaussian with an expectation value of zero and a diagonal covariance whose entries are the power spectrum  $P(k)$ .

- A posterior is a function of the parameters given a measurement. Create a grid of points in parameter space to evaluate the posterior on. The posterior is the product of the likelihood (from the flow or the analytic solution) and the prior.
- Plot the analytic posterior and the flow posterior over the grid. What do you see?

## 7 Looking ahead

This lab was a fairly speculative approach to simulation based inference using generative models. You could improve this method in a number of ways, by fitting a compression function to the data, using Principle Components Analysis (PCA) or by a linear compression (given that the likelihood is Gaussian). Indeed, there is a ‘hidden’ statistic of the data (and GRFs in general) that summarises the information content of the data. Can you find it and fit the flow to it? You could employ a state-of-the-art diffusion model for the likelihood fitting. You could look at adding an additional source of noise on top of the measurement to create a hierarchical model setup for the problem, looking to test how the flow responds to noise not present in the training data. If you are interested in a deeper explanation of some of the physics involved in modern cosmology then look into [5, 4] as well as the lecture courses at LMU Physik (Prof. Mukhanov, Dr. Sanchez, Prof. Weller and Dr. Friedrich).

## Acknowledgements

Thank you to Sven Krippendorf for insightful discussions whilst writing this lab. Thank you to Michael Walther for helpful comments and discussions.

## References

- [1] Francois Chollet. *Autoencoders in Keras*. 2016. URL: <https://blog.keras.io/building-autoencoders-in-keras.html>. (accessed: 01.03.2024).

- [2] Kyle Cranmer, Johann Brehmer, and Gilles Louppe. “The frontier of simulation-based inference”. In: *Proceedings of the National Academy of Sciences* 117.48 (May 2020), pp. 30055–30062. ISSN: 1091-6490. DOI: 10.1073/pnas.1912789117. URL: <http://dx.doi.org/10.1073/pnas.1912789117>.
- [3] Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. *Density estimation using Real NVP*. 2017. arXiv: 1605.08803 [cs.LG].
- [4] S. Dodelson and F. Schmidt. *Modern Cosmology*. Elsevier Science, 2020. ISBN: 9780128159484. URL: <https://books.google.de/books?id=GGjfywEACAAJ>.
- [5] Oliver Friedrich and Jed Homer. *The Probability of the Cosmos - Classical, Quantum, Bayesian*. 2022. URL: [https://www.usm.uni-muenchen.de/people/oliverf/files/probability\\_of\\_the\\_cosmos.pdf](https://www.usm.uni-muenchen.de/people/oliverf/files/probability_of_the_cosmos.pdf). (accessed: 01.03.2024).
- [6] Garrett Goon. *Gaussian Random Fields*. 2021. URL: <https://garrettgoon.com/gaussian-fields/>. (accessed: 01.03.2024).
- [7] David Greenberg, Marcel Nonnenmacher, and Jakob Macke. “Automatic Posterior Transformation for Likelihood-Free Inference”. In: *Proceedings of the 36th International Conference on Machine Learning*. Ed. by Kamalika Chaudhuri and Ruslan Salakhutdinov. Vol. 97. Proceedings of Machine Learning Research. PMLR, Sept. 2019, pp. 2404–2414. URL: <https://proceedings.mlr.press/v97/greenberg19a.html>.
- [8] George Papamakarios. *Neural Density Estimation and Likelihood-free Inference*. 2019. arXiv: 1910.13233 [stat.ML].
- [9] Will J Percival et al. “Matching Bayesian and frequentist coverage probabilities when using an approximate data covariance matrix”. In: *Monthly Notices of the Royal Astronomical Society* 510.3 (Dec. 2021), pp. 3207–3221. ISSN: 0035-8711. DOI: 10.1093/mnras/stab3540. eprint: <https://academic.oup.com/mnras/article-pdf/510/3/3207/42110821/stab3540.pdf>. URL: <https://doi.org/10.1093/mnras/stab3540>.
- [10] A. Walker R. Fisher S. Perkins and E. Wolfart. *The Probability of the Cosmos - Classical, Quantum, Bayesian*. 2003. URL: <https://homepages.inf.ed.ac.uk/rbf/HIPR2/fourier.htm>. (accessed: 01.03.2024).
- [11] Robin Rombach et al. *High-Resolution Image Synthesis with Latent Diffusion Models*. 2021. arXiv: 2112.10752 [cs.CV].
- [12] Max Tegmark, Andy N. Taylor, and Alan F. Heavens. “Karhunen-Loeve Eigenvalue Problems in Cosmology: How Should We Tackle Large Data Sets?” In: *The Astrophysical Journal* 480.1 (May 1997), pp. 22–35. ISSN: 1538-4357. DOI: 10.1086/303939. URL: <http://dx.doi.org/10.1086/303939>.
- [13] Lilian Weng. *Normalizing flow models*. 2018. URL: <https://lilianweng.github.io/posts/2018-10-13-flow-models/>. (accessed: 01.03.2024).