The Probability of the Cosmos - Classical, Quantum, Bayesian

Oliver Friedrich, Jed Homer

Fakultät für Physik, Ludwig-Maximilians Universität München,
Geschwister-Scholl-Platz 1, 80539 München, Germany

E-mail: oliver.friedrich@physik.lmu.de

Abstract. The Universe can be seen as the outcome of a probabilistic experiment. We will trace different aspects of cosmology in which this probabilistic nature of the cosmos shows up - from the likelihood functions of cosmological observables, over the functional probability density of the entire cosmic density field to the Quantum probabilities of curvature perturbations in the early Universe - and we will study the gradual transition between them. On our journey we will learn about cutting edge concepts in the theory of cosmic structure formation. At the same time, we will face interesting philosophical questions: What is probability - classical, quantum, Bayesian?

Contents

Part I:
The Cosmic Density Field as a Random Field

1 Introduction: The inhomogeneous Universe

Let us start this course by jumping right into the matter. Figure 1 displays the angular positions of galaxies on the sky as observed with year-1 data of the Dark Energy Survey (DES, https://www.darkenergysurvey.org). DES is a wide area galaxy survey that covers about 5000 square degrees on the sky, so Figure 1 only shows a small subset of the DES data. It also only covers a small cutout of the full redshift range that was observed by DES ($0.6 < 1 < 0.65$ as opposed to the full DES range $0 < z \lesssim 1.5$). The galaxies that have been used for the figure are so called luminous red galaxies (LRGs) which are large and old galaxies that are often considered in cosmology because it is particularly easy to determine their redshift and hence their distance to us. But they are only a small subset of all the galaxies that populate the cosmos.

The subject of large-scale structure cosmology is studying this large-scale distribution of objects in the Universe and tries to explain it. **(JH: Possible rephrase: *Research into large-scale structure in cosmology focuses on the large-scale distribution of objects in the universe to explain the appearance of structure.*)** This leads us to a first exercise for you:

Exercise 1
What is there to explain about in Figure 1? Take some time to formulate 3 questions or observations about the figure that in your opinion need answering / an explanation.

Whatever questions or observations you came up with - it is probably impossible to give answers and explanations to them that are 100% certain. To understand why this is, look e.g. at Figure 2. The left hand panel of that figure shows the same galaxy positions as Figure 1 while the right hand panel shows positions that have been randomly drawn from a uniform distribution but with the same overall number density. Is it **possible**, that the positions of the galaxies on the left hand panel of Figure 2 are also the result of such a uniform random draw? The answer is clearly *yes*. But is it also **likely** that they are the result of such a random draw? The answer to that may very well be *no*. To figure that out we will study the cosmic matter density field

$$\rho(t, \boldsymbol{r}) \equiv \bar{\rho}(t)\ (1 + \delta(t, \boldsymbol{r}))\ . \tag{1.1}$$

Here $\delta(t, \boldsymbol{r})$ is the so called matter density contrast, which quantifies the relative deviation of the density field at a location $\boldsymbol{r}$ and time t from the average cosmic density $\bar{\rho}(t)$ at that time.

1.1 A homogeneous, expanding cloud of dust

Let us in a first step assume that $\delta = 0$ everywhere, i.e. that the Universe is perfectly homogeneous. Then how does $\bar{\rho}$ evolve with time? We can answer this with a pseudo-Newtonian treatment: Assume that the Universe is a homogeneous, spherical dust cloud of radius $R = R(t)$. It is then then clear that

$$\bar{\rho}(t) = \bar{\rho}(t_0)\ \frac{R(t_0)^3}{R(t)^3} \equiv \bar{\rho}_0\ \frac{R_0^3}{R(t)^3}\ , \tag{1.2}$$

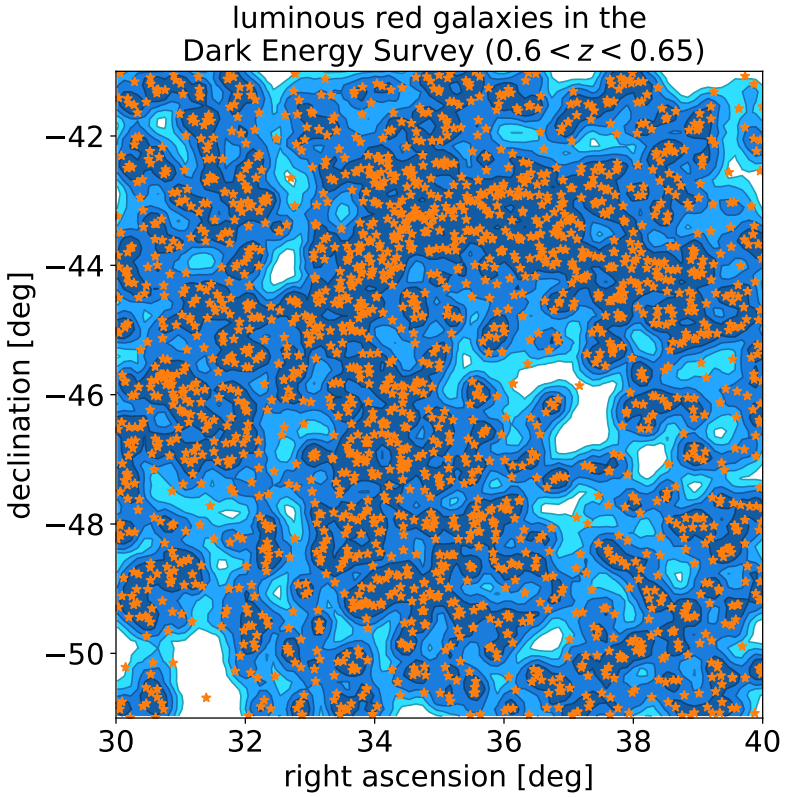

Figure 1. Angular positions of luminous red galaxies (LRGs) on the sky as observed with year-1 data of the Dark Energy Survey (DES). This is just a tiny fraction of the area covered by DES and also only a tiny fraction of the redshift range that DES has observed. Furthermore, LRG are only a small subset of the galaxies present in the cosmic web. The blue density contours represent a Gaussian smoothing of the galaxy positions in order to obtain a smooth density field.

where t_0 is some fixed moment in time, which we will take to represent today, and the last equality serves as a definition of $\bar{\rho}_0$ and R_0. So the time evolution of $\bar{\rho}$ can be reduced to the evolution of R. Now consider a small mass element dm which is located directly at the edge of the cloud. It's distance to the center is exactly R and Newtonian gravity tells us that this distance evolves according to the differential equation

$$\ddot{R} = -G\frac{M}{R^2} + \lambda M R \ , \tag{1.3}$$

where M is the total mass of the cloud and G and λ are two constants quantifying the force strength. As you have surely spotted, only the first term on the right hand side of this equation actually represents Newtonian gravity. But the second term has in fact been considered by Newton as a possible extension to his theory because it shares an interesting symmetry property with the first term [16], and we will keep (and need!) it in the following. The mass of the cloud is given by

$$M = \frac{4\pi G}{3}\bar{\rho}R^3 = \frac{4\pi G}{3}\bar{\rho}_0 R_0^3 \tag{1.4}$$

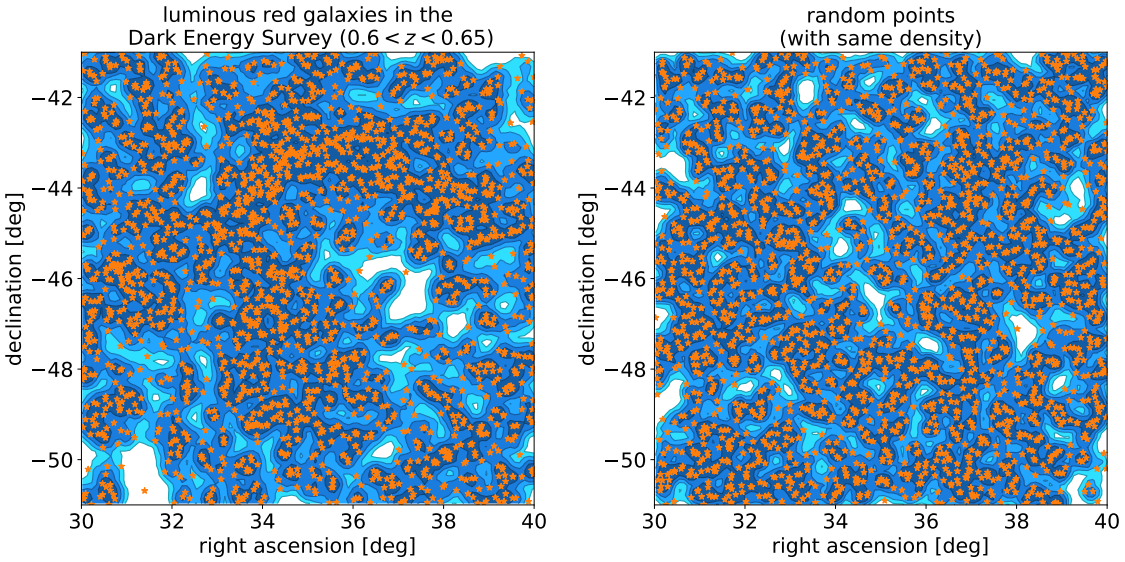

Figure 2. The left hand panel shows the same galaxy positions as Figure 1 while the right hand panel shows positions that have been randomly drawn from a uniform distribution.

and in particular it is conserved over time. Using this, we can re-express Equation 1.3 as

$$\ddot{R} = -\frac{4\pi G R_0^3 \bar{\rho}_0}{3}\frac{1}{R^2} + \frac{4\pi\lambda R_0^3\bar{\rho}_0}{3}R \,. \tag{1.5}$$

Despite the fact that we have started with a Newtonian equation, we are actually about to obtain the differential equation that governs the expansion of the Universe in the full theory of General Relativity. To arrive there, let us first change from R to the unit-less function

$$a(t) \equiv \frac{R(t)}{R_0} \,, \tag{1.6}$$

which is called the *scale factor* of the Universe. In terms of a Equation 1.5 reads

$$\begin{aligned}\ddot{a} &= -\frac{4\pi G\bar{\rho}_0}{3}\frac{1}{a^2} + \frac{4\pi\lambda R_0^3\bar{\rho}_0}{3}a \\ &\equiv -\frac{4\pi G\bar{\rho}_0}{3}\frac{1}{a^2} + \frac{\Lambda c^2}{3}a \,.\end{aligned} \tag{1.7}$$

Here c is the speed of light, and we have introduced a new constant Λ. Equation 1.7 is exactly the Friedmann equation, which describes the expansion of a homogeneous universe within General Relativity + a cosmological constant Λ.

What does this equation tell us? The scale factor a describes how the Universe expands over time, so $\ddot{a}$ quantifies the acceleration of the expansion. The first term on the right hand side of Equation 1.7 (the term appearing in the standard version of Newtonian gravity) is in fact always negative, i.e. it tends to cause a deceleration of the expansion. This makes sense, because by its own gravity our dust cloud should slow down its expansion. But for positive values of Λ the second term can counteract this. Cosmological observations have with very high certainty shown that this term indeed exists [46] and that Λ is actually large enough to lead to an overall positive acceleration of the cosmic expansion.

Exercise 2

Why were we able to start with a classical equation and still arrive at the relativistic result? To what extent does or doesn't our derivation of Equation 1.7 depend on the absolute size we assumed for the Universe?

Exercise 3

Formulate further questions you may have about the above material.

1.2 Perturbations of the density field

Let us now consider inhomogeneities $\delta \neq 0$ in the cosmic matter density field. We will treat our dust cloud as a fluid, which in classical mechanics is described by the quantities

- $\rho(t, \boldsymbol{r})$, the mass density of the fluid,
- $\mathbf{v}(t, \boldsymbol{r})$, the streaming velocity of the fluid,
- $p(t, \boldsymbol{r})$, the pressure of the fluid,

and potentially of other quantities such as the shear viscosity, which we will ignore in the following. A fluid, as described by the above functions, evolves according to the continuity equation

$$\frac{\partial \rho}{\partial t} + \boldsymbol{\nabla}_{\boldsymbol{r}}(\rho \mathbf{v}) = 0 \tag{1.8}$$

and the Euler equation

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \boldsymbol{\nabla}_{\boldsymbol{r}})\mathbf{v} + \frac{1}{\rho}\boldsymbol{\nabla}_{\boldsymbol{r}} p = \boldsymbol{g} \ , \tag{1.9}$$

where $\boldsymbol{g}$ is the acceleration that fluid elements experience due to external forces. The continuity equation says that the rate at which mass flows out of an infinitesimal volume should be equal to minus the time derivative of the mass in that volume, i.e. it states that mass is conserved. The Euler equation says that the acceleration of any particular fluid element should be given by the sum of all the forces acting on that element. This can be most straight forwardly seen by noting that

$$\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = \frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \boldsymbol{\nabla}_{\boldsymbol{r}})\mathbf{v} \ . \tag{1.10}$$

The continuity and Euler equations are a set of 4 partial differential equations, i.e. they can be used to solve for the evolution of 4 degrees of freedom. If we take those degrees of freedom to be ρ and $\mathbf{v}$ (the 3-dimensional velocity vector), then we still need further information to fix the behaviour of p and $\boldsymbol{g}$. In an ideal fluid the pressure p would typically be specified as a function of ρ via an equation of state

$$p = p(\rho) \ . \tag{1.11}$$

For most of this lecture, we will actually assume the $p = 0$, i.e. that there is no pressure at all. This is a very good approximation for the dark matter that makes up most matter in the Universe. The acceleration $\boldsymbol{g}$ will in our situation be caused purely through gravitational interaction between the different fluid elements. In Newtonian gravity this means that

$$\boldsymbol{g}(t, \boldsymbol{r}) = -\boldsymbol{\nabla}_{\boldsymbol{r}}\phi(t, \boldsymbol{r}) \tag{1.12}$$

where the gravitational potential ϕ is given in terms of the fluid density as

$$\phi(t, \boldsymbol{r}) = -\int \mathrm{d}^3 r' \; \rho(t, \boldsymbol{r}') \left(\frac{G}{|\boldsymbol{r} - \boldsymbol{r}'|} + \frac{\lambda}{2} |\boldsymbol{r} - \boldsymbol{r}'|^2 \right) \tag{1.13}$$

$$\Leftrightarrow \; \Delta_{\boldsymbol{r}} \phi(t, \boldsymbol{r}) = 4\pi G \rho(t, \boldsymbol{r}) - \Lambda \; , \tag{1.14}$$

where the second equality is the Poisson equation and we have again taken into account the modification to purely Newtonian gravity given by the second term of Equation 1.3.

Exercise 4

Derive Equations 1.13 and 1.14. Hint: start with Equations 1.3 and 1.12, and also use the fact that $\Delta_{\boldsymbol{r}} |\boldsymbol{r} - \boldsymbol{r}'|^{-1} = -4\pi \delta_D(\boldsymbol{r} - \boldsymbol{r}')$ *where* δ_D *is the Dirac delta function.*

Cosmological observations have demonstrated that the Universe is very homogeneous on scale of $\gtrsim 100$ Mpc. Hence, when looking at the cosmic density field with a coarse resolution of about 100 Mpc its evolution will be well described by Equation 1.7. Let us try to factor out this large-scale motion of matter in the Universe (the so called *Hubble flow*) and derive equations of motions for only the perturbations on top of that motion. We had already decomposed the matter density field in terms of the density contrast and the average cosmic density,

$$\rho = (1 + \delta) \; \bar{\rho} \; . \tag{1.15}$$

We can also decompose the velocity field as

$$\mathbf{v}(t, \boldsymbol{r}) = \bar{\mathbf{v}}(t, \boldsymbol{r}) + \boldsymbol{v}(t, \boldsymbol{r}) \; , \tag{1.16}$$

where $\bar{\mathbf{v}}$ would be the velocity field of a perfectly homogeneous, expanding gas cloud and $\boldsymbol{v}$ is the perturbation to that in the real, inhomogeneous Universe. And finally, also the gravitational potential will pick up a perturbation

$$\phi(t, \boldsymbol{r}) = \bar{\phi}(t, \boldsymbol{r}) + \varphi(t, \boldsymbol{r}), \tag{1.17}$$

where $\bar{\phi}(t, \boldsymbol{r})$ would be the potential of a perfectly homogeneous gas cloud. The potential perturbation φ is connected to the density contrast δ via the modified Poisson equation

$$\Delta_{\boldsymbol{r}} \varphi(t, \boldsymbol{r}) = 4\pi G \bar{\rho}(t) \delta(t, \boldsymbol{r}) \; . \tag{1.18}$$

To obtain particularly concise equations for the perturbations δ, $\boldsymbol{v}$ and φ it is useful to a set of new time and spatial variables. First, let us introduce *co-moving coordinates* $\boldsymbol{x}$ through

$$\boldsymbol{x}(t, \boldsymbol{r}) \equiv \boldsymbol{r} / a(t) \; , \tag{1.19}$$

where a is the scale factor of the homogeneous background expansion. The new coordinates $\boldsymbol{x}$ are co-moving with the homogeneous background flow $\bar{\mathbf{v}}(t, \boldsymbol{r})$ (hence the name). Furthermore, let us introduce the so called *conformal time*, which can e.g. be defined as

$$\eta(t) = \int_{t_0}^{t} \mathrm{d}t' \; \frac{1}{a(t')} \; . \tag{1.20}$$

Derivatives in these new variables are related to derivatives in the old variables via

$$\boldsymbol{\nabla}_{\boldsymbol{x}} = a \boldsymbol{\nabla}_{\boldsymbol{r}} \; , \; \Delta_{\boldsymbol{x}} = a^2 \Delta_{\boldsymbol{r}} \tag{1.21}$$

and

$$\left.\frac{\partial}{\partial \eta}\right|_{\boldsymbol{x}=\text{const.}} = a \left\{ \left.\frac{\partial}{\partial t}\right|_{\boldsymbol{r}=\text{const.}} + \bar{\mathbf{v}} \boldsymbol{\nabla}_{\boldsymbol{r}} \right\} . \tag{1.22}$$

Exercise 5

Derive Equations 1.21 and 1.22.

With these definitions and relations, and taking into account that the scale factor a satisfies Equation 1.7, one can show that the continuity equation for the perturbations takes the form

$$\frac{\partial \delta}{\partial \eta} + \boldsymbol{\nabla}_{\boldsymbol{x}} \left[(1+\delta) \boldsymbol{v} \right] = 0 \tag{1.23}$$

while the Euler equation becomes

$$\frac{\partial \boldsymbol{v}}{\partial \eta} + \mathcal{H} \boldsymbol{v} + (\boldsymbol{v} \cdot \boldsymbol{\nabla}_{\boldsymbol{x}}) \boldsymbol{v} = -\boldsymbol{\nabla}_{\boldsymbol{x}} \varphi . \tag{1.24}$$

Here $\mathcal{H}$ is the socalled conformal expansion rate, which is defined as

$$\mathcal{H} = \frac{1}{a} \frac{\partial a}{\partial \eta} . \tag{1.25}$$

Exercise 6

Derive Equations 1.23 and 1.24. Hint: use the fact that the background quantities $\bar{\rho}$, $\mathbf{v}$ and $\bar{\phi}$ (i.e. the density, velocity and potential of a homogeneous, unperturbed gas cloud) also satisfy the continuity, Euler and Poisson equations. Especially, $\bar{\rho}(t) \sim 1/a(t)^3$ and the scale factor $a(t)$ satisfies the Friedmann equation.

Before we attempt to solve Equations 1.23 and 1.24, let us pause for a moment. What exactly have we achieved so far? We have derived concise evolution equations for δ and $\boldsymbol{v}$, but how does this connect to our original question?

Exercise 7

Formulate at least one idea of how observations like those of Figure 1 could be used to test the validity of Equations 1.23 and 1.24. In what sense would or wouldn't such a test constitute an explanation of Figure 1?

1.3 Linear evolution of density perturbations

At leading order Equations 1.23 and 1.24 will be dominated by terms that are linear in the perturbations, i.e.

$$\frac{\partial \delta}{\partial \eta} + \boldsymbol{\nabla}_{\boldsymbol{x}} \boldsymbol{v} \approx 0 \tag{1.26}$$

and

$$\frac{\partial \boldsymbol{v}}{\partial \eta} + \mathcal{H} \boldsymbol{v} \approx -\boldsymbol{\nabla}_{\boldsymbol{x}} \varphi . \tag{1.27}$$

Taking the divergence $\boldsymbol{\nabla}_{\boldsymbol{x}}$ of the linearized Euler equation and inserting both the linearized continuity equation and the perturbed Poisson equation into it leads to a closed equation for δ only:

$$\frac{\partial^2 \delta}{\partial \eta^2} \approx 4\pi G a^2 \bar{\rho} \delta - \mathcal{H} \frac{\partial \delta}{\partial \eta} . \tag{1.28}$$

Exercise 8

Derive Equation 1.28.

Equation 1.28 has a very simple interpretation. The first term on its right hand side acts as a driver for gravitational collapse: if a region in the Universe is overdense ($\delta > 0$) then this region will collapse under its own gravity which will lead to a positive acceleration of density ($\partial^2\delta/\partial\eta^2 > 0$) and the region will become even more overdense. Similarly, an underdense region ($\delta < 0$) loses its matter to surrounding overdense structures and will become even more underdense in the future ($\partial^2\delta/\partial\eta^2 < 0$). In contrast to that, the last term on the right hand side of Equation 1.28 acts as a friction that dampens gravitational collapse (recall that friction forces acting on an object are typically proportional to the velocity of that object). This term quantifies how much the expansion of the Universe (i.e. the Hubble flow) suppresses the collapse of structures.

We can solve Equation 1.28 with the ansatz

$$\delta(t, \boldsymbol{x}) = \delta(t_i, \boldsymbol{x}) \frac{D(t)}{D(t_i)} \ , \tag{1.29}$$

where t_i could e.g. be some initial point in time and the function $D(t)$ is the so called growth factor that satisfies a differential equation equivalent to Equation 1.28. So, as long as the linear approximation is justified, we can think of the cosmic density field as an initial density field that is re-scaled by an over all factor $D(t)/D(t_i)$. As we will see later in the course, this point of view is oversimplified and the non-linearity of gravitational collapse becomes relevant at times and scales that are important for modern cosmology. But the ansatz in Equation 1.29 is nevertheless highlighting an important distinction between the properties of the initial density field $\delta(\boldsymbol{x}) = \delta(t_i, \boldsymbol{x})$ at some early time during the Universe's history and how gravity is modifying these properties at later times. This conceptual split is also highlighted in Figure 3 which shows the evolution matter density contrast in a simulated, 1-dimensional Universe.

Exercise 9

Does the ansatz of Equation 1.29 map in any way to the answers you gave to Exercise 7?

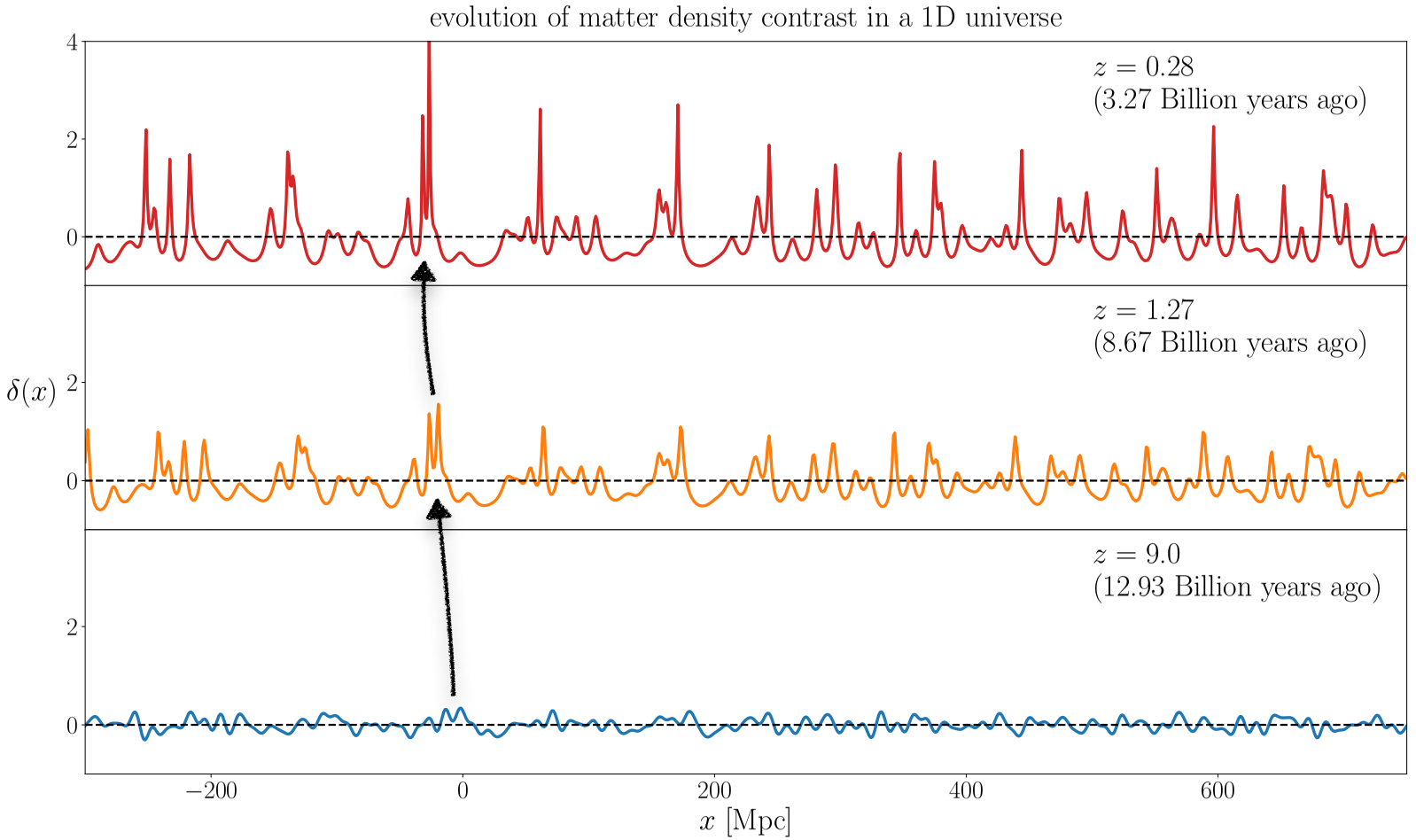

Figure 3. Evolution of the matter density contrast field $\delta(\boldsymbol{x})$ in a simulated, 1-dimensional universe. The subject of large-scale structure cosmology can largely be split into two questions: What were the statistical properties of the initial density field (blue line in the left most panel)? And how did the cosmic density field evolve after that (transition from left to right)? The two most important aspects of the latter question are gravity & the expansion of the Universe. Gravity leads to the collapse of tiny initial structures into larger and larger overdensities, while the cosmic expansion act as a friction that slows down structure growth (cf. Equation 1.28 and the text thereafter).

We are going to close this section with a symbolic outlook on upcoming parts of this course. Imagine that some cosmological observable can be considered as a functional of the density field, i.e. $\mathcal{O}(t) = \mathcal{O}[\delta(t, \cdot)]$. **(AH: Question: what is the dot in the argument of δ standing for? A single location, set of locations?) (JH: Is an example of $\mathcal{O}$ the galaxy field - i.e. some tracer of total density field δ?)**The expectation value of $\mathcal{O}$ can then be computed as

$$\langle \mathcal{O}(t) \rangle = \int \mathcal{D}\delta \; \mathcal{P}[\delta|t] \; \mathcal{O}[\delta] \; , \tag{1.30}$$

where $\mathcal{D}\delta$ represents functional integration over all possible configurations of δ and $\mathcal{P}[\delta|t]$ is the probability density functional (PDFunctional) of δ at the time t. If we do not know $\mathcal{P}$ then it may still be possible to evaluate the above integral if we know the PDFunctional of the initial density field, i.e. one could instead consider the integral **(AH: can one show the steps in between how to go from 1.30 to 1.31?)**

$$\langle \mathcal{O}(t) \rangle = \int \mathcal{D}\delta_i \; \mathcal{P}_i[\delta_i] \; \mathcal{O}\left[\delta(\cdot|\delta_i, t)\right] \; , \tag{1.31}$$

where the density contrast field at time t is now expressed as a functional of the initial density field δ_i and $\mathcal{P}_i$ is the PDFunctional of δ_i. We can even obtain the full probability density function (PDF) of the observable $\mathcal{O}$ from $\mathcal{P}_i$ as **(AH: Actually, this is great because I did not even know about this relation between the PDFunctional and normal PDF! But why is this the case?)**

$$p(\mathcal{O} = o|t) = \int \mathcal{D}\delta_i \; \mathcal{P}_i[\delta_i] \; \delta_D \left(o - \mathcal{O}\left[\delta(\cdot|\delta_i, t)\right]\right) \; . \tag{1.32}$$

(JH: Here o is a realization of the random variable $\mathcal{O}$?) Here δ_D is the Dirac delta distribution (not to be confused with the matter density contrast δ).

(JH: What exactly is the definition of a functional and a PDFunctional? I have a rough definition of the former from the least-action principle and Lagrangians but the idea is more interesting here so a proper definition would be useful.)

2 Random Variables and Probability Distributions

Recall that in the previous section we had said that it is the task of large-scale structure cosmology to "explain" observations about the cosmic matter density field and its tracers, such as the galaxy density field. We then found that this task can be split into two parts (cf. Figure 3 as well as Equation 1.29) which are

a) understanding the origin of the initial density contrast field $\delta_i(\boldsymbol{x}) = \delta(t_i, \boldsymbol{x})$ and

b) understanding how the initial density field was modified during the subsequent evolution of the cosmos.

We have somewhat addressed point b) in Section 1 by deriving the differential equations that govern the evolution of the cosmic density and velocity fields. As we will see in the second half of the course, there are reasons to believe that the exact configuration of the initial density field is the outcome of a genuinely random process. Hence, there is no way to theoretically predict exactly how the initial density contrast field δ_i looked like. All we can hope to achieve is to predict the statistical properties of the random process from which δ_i resulted. **(JH: this is very cool)**

To prepare for this task, we are going to look at random variables and their mathematical description. A particularly simple example of a random variable X is one that can take on two different values $\{H, T\}$. Such an X could e.g. model the outcome of a coin flip with H representing "heads" and T representing "tails". Or it could model the outcome of a Stern-Gerlach experiment, where one shoots an electron whose spin is aligned along the x-axis of a lab frame through a magnetic field aligned along the z-axis and then measures the deflection of the electron trajectory.

Exercise 10

Find more examples of random processes whose outcomes can be modelled by a random variable with two possible values.

To fully characterise a binary variable we need to specify probabilities for its two possible outcomes. A probability of 1 is usually taken to represent absolute certainty that a certain outcome will occur, while a probability of 0 represents absolute certainty that it does not occur. Hence, if we assume that it is guaranteed to obtain *any* outcome, the sum of the probabilities of all outcomes should be one. So we can parameterise the outcome probabilities by

$$P(X = H) = p \quad \text{and} \quad P(X = T) = 1 - p \quad \text{with } p \in [0, 1] \; . \tag{2.1}$$

Exercise 11

Assume that $P(X = H) = p$ with $p > 0$ and $p < 1$. What does that mean? Try to answer this question either in general or within the context of the above examples of random processes.

We will take time to investigate the question of how to interpret probabilities in later lectures. But let us take a quick outlook on difficulties that can arise in such interpretations. One common way to define the meaning of p is to say: "If we were to consider many independent realisations X_i, $i \in \{1, \ldots, N\}$, of X then the fraction of times that the outcome H was obtained,

$$\hat{p}_N = \frac{\#\{X_i = H \mid i = 1, \ \ldots \ N\}}{N} \; , \tag{2.2}$$

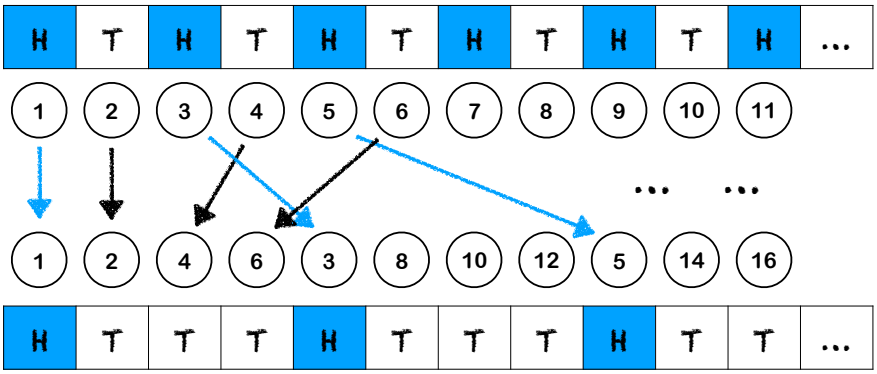

Figure 4. Sketch about difficulties when defining the meaning of probabilities: The upper part of the figure shows the outcomes of an infinite sequence of independent, identically distributed random variables. It seems like the fraction of times that the outcome H was obtained is 0.5, which is consistent with $P(X_i = H) = 0.5\ \forall i$. But a mere rearrangement of the sequence (lower part of the figure) seems to indicate a different probability of $P(X_i = H) = 0.25$. **(AH: From this figure only, this is not obvious –> the above numbers 0.5 in the first case and 0.25 in the second case are only true if we seemingly restrict our attention to only the first 8 tosses. I.e., we do not toss infinitely many times to calculate these numbers. Shouldn't we explicitly mention this?) (OF: This reshuffling is meant to go on forever (and it indeed can).)**

approaches the probability p as N approaches infinity." This statement is still somewhat vague, but we can try to put it into precise mathematical terms as follows:

Let $\{X_i = H \mid i = 1, \ldots N\}$ be N independent, identically distributed random variables with $P(X_i = H) = p\ \forall\ i$. Define the relative fraction $\hat{p}_N$ of outcomes H among the X_i as in Equation 2.2. Then for every set of small parameters ϵ and δ there exists a sufficiently high $\tilde{N}$ such that

$$P(\hat{p}_{\tilde{N}} \in [p - \epsilon, p + \epsilon]) > 1 - \delta\ . \tag{2.3}$$

In other words, even if we choose ϵ to be arbitrarily small, the probability of finding $\hat{p}_N$ to be less than ϵ away from p becomes arbitrarily close to 1 if we only choose N high enough. **(JH: Really interesting. Is it not that the probability of $\hat{p}_N$ being ϵ away from p becomes arbitrarily close to 0 (and not 1) for high N? Otherwise it seems that for high N we are certain $\hat{p}_N$ is *not* close to p? What happens with δ here also?)**

Exercise 12

Do you see any problem with the above definition of probability?

One potential problem in the above definition is the fact that Equation 2.3 is a probability statement itself **(AH: i.e. we are talking about the probability P of $\hat{p}_N$, where the latter is already our definition of probability!)**. So, in order to define what we mean by assigning probabilities to the outcomes of one random process, we were using the probabilities assigned to the outcomes of another random process. This clearly appears like a cyclic argument. A potential way around this cyclicity would be to skip the approach to larger and larger N in the above definition and consider infinite sequences of random draws $\{X_i = H \mid i = 1, \ldots \infty\}$ right away. We could then say that the only possible draws of such

a random sequence are draws in which $\hat{p}_\infty$ is *exactly* given by p.[1] But such an attempt to define p comes with problems of its own. Assume e.g. that $p = 0.5$ and consider Figure 4. The upper part of that figure displays a draw of a random sequence in which for every odd i $X_i = H$ while for any even i $X_i = T$. For $p = 0.5$ such a sequence of outcomes seems to be allowed according to the above statement. But this view is challenged by the lower part of Figure 4. It shows a mere rearrangement of the sequence, which still covers all of the drawn random variables X_i. But in that rearranged sequence the fraction of outcomes H, $\hat{p}_\infty$, seems to equal 0.25 which seems to be in conflict with $p = 0.5$. What has gone wrong?

Exercise 13

Try to fix the above attempts of defining probability with a definition of your own.

Exercise 14

What could others potentially criticise about your definition from Exercise 13?

2.1 PDFs and cumulant generating functions of random variables

Let us move away from binary variables and consider a continuous random variable X that can take any value on the real line $\mathbb{R}$ and that is distributed according to some probability density function (PDF) p on $\mathbb{R}$. This means that the probability of finding $X \in [x, x + dx]$ after a draw of X is given by

$$P(X \in [x, x + dx]) = p(x)\mathrm{d}x \ . \tag{2.4}$$

Since the probability of finding X anywhere on the real line should be 1, the PDF needs to be normalised as

$$1 = \int \mathrm{d}x \ p(x) \ . \tag{2.5}$$

Correspondingly, the probability of finding X in any finite interval $[x_1, x_2]$ must be smaller than or equal to 1, i.e.

$$1 \geq \int_{x_1}^{x_2} \mathrm{d}x \ p(x) \equiv P(X \in [x_1, x_2]) \ . \tag{2.6}$$

Important examples of continuous random variables are e.g. Gaussian random variables, whose PDF is given in terms of their expectation value μ and their standard deviation σ as

$$p_{\text{Gauss}}(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \ \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \ . \tag{2.7}$$

A related type of variable that is commonly used to model stochastic processes are the so called log-normal random variables, whose PDF is given by

$$p_{\text{LN}}(z|\mu, \sigma) = \frac{1}{z\sqrt{2\pi\sigma^2}} \ \exp\left(-\frac{(\ln z - \mu)^2}{2\sigma^2}\right) \ . \tag{2.8}$$

Log-normal variables Z are exactly those random variables whose logarithm $X = \ln Z$ is Gaussian. Figure 5 compares a Gaussian PDF with $\mu = 1$ and $\sigma = 0.4$ to a log-normal PDF with the same expectation value and standard deviation (note that this requires values for μ

[1] The point here is that we have no problem defining what a probability of $p = 1$ means.

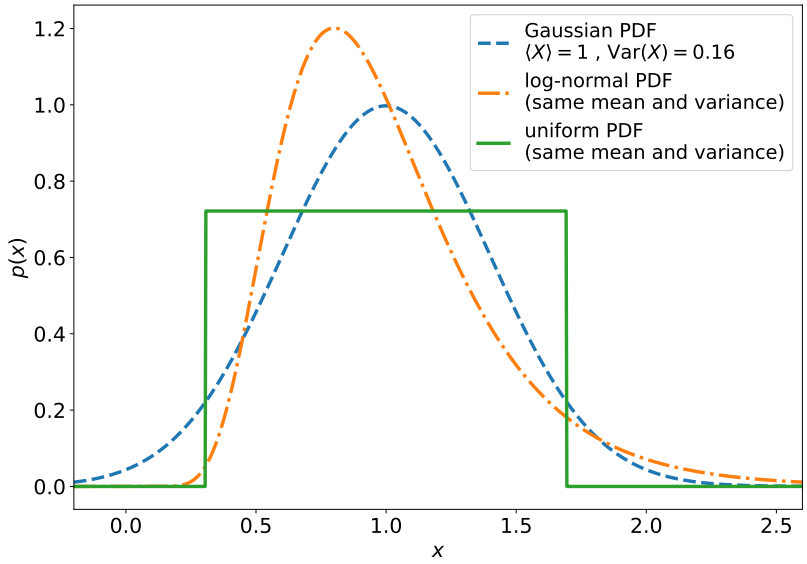

Figure 5. Comparing Gaussian, log-normal and uniform probability density functions. All of the shown PDFs have the same expectation value and the same standard deviation.

and σ in Equation 2.8 that are different from 1 and 0.4). That figure also shows the PDF of a uniform random variable,

$$p_{\text{uni}}(x|a,b) = \begin{cases} \frac{1}{b-a} & \text{for } x \in [a,b] \\ \\ 0 & \text{else} \end{cases} , \tag{2.9}$$

where the parameters a and b where again chosen to yield the same expectation value and standard deviation as the other two PDFs.

Exercise 15

Calculate the values of a and b for which the expectation value and variance of the uniform PDF $p_{\text{uni}}(x|a,b)$ take the values $\langle X \rangle = 1$ and $\text{Var}(X) \equiv \langle (X - \langle X \rangle)^2 \rangle = 0.16$.

An important characteristic of a continuous random variable X are its moments. The n-th moment of X is given by the expectation value

$$\langle X^n \rangle = \int \mathrm{d}x \; x^n \; p(x) \; . \tag{2.10}$$

If all moments of a random variable are finite (which is not necessarily the case) then one can attempt to summarise them within a single object - the so called *moment generating function* (MGF) $\psi = \psi(\lambda)$ which is defined as

$$\psi(\lambda) \equiv \langle e^{\lambda X} \rangle \tag{2.11}$$

$$\equiv \sum_{n=0}^{\infty} \frac{\langle X^n \rangle}{n!} \; \lambda^n \tag{2.12}$$

$$\equiv \int \mathrm{d}x \; e^{\lambda x} \; p(x) \; . \tag{2.13}$$

Note that the MGF is not well defined for all random variables, i.e. that the sum in Equation 2.12 does not converge for all random variables. Note also that Equation 2.13 indicates that PDF and MGF are related via a Laplace transform. In particular, you can think of the pair (λ, x) as an analog to the pair of variables (k, x) that appears when Fourier transforming a function $f(x)$.

A utility of the MGF lies in the fact that it can be used to calculate arbitrary moments of X via differentiation instead of integration. Equation 2.12 can be used to check that the n-th moment of X is connected to the n-th derivative of ψ via

$$\langle X^n \rangle = \left. \frac{\mathrm{d}^n \psi(\lambda)}{\mathrm{d}\lambda^n} \right|_{\lambda=0} \quad . \tag{2.14}$$

But the MGF has a merit beyond this relation. As we will see later in the course, for some cosmological observables (if they are considered as draws of random variables) it is easier to obtain a direct theoretical prediction for $\psi(\lambda)$ than for $p(x)$. Given a prediction for $\psi(\lambda)$ one can then invert the Laplace transform of Equation 2.13 to obtain a model for $p(x)$.

Exercise 16

Show analytically that the MGF of the Gaussian PDF p_{Gauss} is given by

$$\psi_{\mathrm{Gauss}}(\lambda) = \exp\left(\mu\lambda + \frac{\sigma^2}{2}\lambda^2\right) \quad . \tag{2.15}$$

Exercise 17

On your computer, setup a numerical experiment to confirm the result of Exercise 16. Hint: You can e.g. use the `python` *package* `numpy` *to write a* `jupyter` *notebook. The function* `numpy.random.normal` *allows you to draw sets of Gaussian random variables with identical μ and σ. Given such a set you can directly estimate the expectation value on the right hand side of Equation 2.11.*

Exercise 18

With your Gaussian random draws from Exercise 17 you can also generate log-normal random draws. Use those to estimate the MGF of a log-normal PDF. What do you observe, and why? Hint: You can use the function `numpy.random.seed` *to control the initial state of* `numpy`*'s random number generator. Try out different seed values.*

Later in the lecture, a slight variation of the MGF will be useful for us - the so called *cumulant generating function* $\varphi(\lambda)$ (not to be confused with the perturbation of the gravitational potential we had looked at in the previous section) which is defined as

$$\varphi(\lambda) = \ln\left(\psi(\lambda)\right) \quad . \tag{2.16}$$

We can express $\varphi(\lambda)$ as a power series in λ as

$$\varphi(\lambda) = \sum_{n=1}^{\infty} \frac{\langle X^n \rangle_c}{n!} \lambda^n \tag{2.17}$$

which serves as a definition of the so called *cumulants* $\langle X^n \rangle_c$ (which are also sometimes called the *connected moments*).

Exercise 19

Show that for any random variable

$$\langle X\rangle_c = \langle X\rangle \tag{2.18}$$
$$\langle X^2\rangle_c = \langle X^2\rangle - \langle X\rangle^2 \tag{2.19}$$
$$\langle X^3\rangle_c = \langle X^3\rangle - 3\langle X\rangle\langle X^2\rangle + 2\langle X\rangle^3 \ . \tag{2.20}$$

Hint: Taylor expand $\ln(\psi(\lambda))$ *in powers of* λ .

2.2 Gaussian random fields in 3D space

The initial configuration of the Universe's matter density contrast, $\delta_i(\boldsymbol{x})$, is much more complex than any individual random variable. In fact, at any point $\boldsymbol{x}$ in space $\delta_i(\boldsymbol{x})$ is a random variable, so we are dealing with an infinite (even uncountable) collection of random variables! In other words: δ_i is a so called *random field*. Random fields are ubiquitous in the world, and other examples would include: the field that assigns to each point on the Earth's surface the current temperature at that point, or the field that assigns to each location on the surface of a lake the difference between the surface height at that point and the mean height of the lake's surface.

Exercise 20

Find more examples of random fields! Can you also find examples for genuine random fields (as opposed to the above examples which are actually only draws of random fields)?

Before dealing with such large collections of random variables, let us first consider finite random vectors $\boldsymbol{F} = (F_1\ ,\ \ldots\ ,\ F_N)$, where each of the F_i is a single, real random variable. We are specifically using the letter F for these variables, because later in this section we want to transition from vectors $\boldsymbol{F}$ with elements F_i to functions f with "elements" $f(\boldsymbol{x})$, i.e. we want to think of functions as vectors with infinitely many elements.

The PDF of a finite dimensional random vector $\boldsymbol{F}$ will be a function

$$p : \mathbb{R}^N \longrightarrow \mathbb{R} \tag{2.21}$$

which satisfies

$$p \geq 0\ \ , \quad \int \mathrm{d}^N f\ p(\boldsymbol{f}) = 1\ . \tag{2.22}$$

The probability of finding $\boldsymbol{F}$ within any subvolume $V \subset \mathbb{R}^N$ is then given by

$$P(\boldsymbol{F} \in V) = \int\limits_{\boldsymbol{f}\in V} \mathrm{d}^N f\ p(\boldsymbol{f})\ . \tag{2.23}$$

An important example of a such random vectors are multivariate Gaussian random variables, whose PDF is given by

$$p_{\text{multi.Gauss.}}(\boldsymbol{f}|\boldsymbol{\mu},\boldsymbol{C}) = \frac{1}{|2\pi\boldsymbol{C}|^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{f}-\boldsymbol{\mu})^T\boldsymbol{C}^{-1}(\boldsymbol{f}-\boldsymbol{\mu})\right\}\ . \tag{2.24}$$

Here $\boldsymbol{\mu} = \langle\boldsymbol{F}\rangle$ is the (N-dimensional) expectation value of $\boldsymbol{F}$, $\boldsymbol{C}$ is the covariance matrix of $\boldsymbol{F}$ with the elements $C_{ij} = \langle(F_i-\mu_i)(F_j-\mu_j)\rangle$ and $|2\pi\boldsymbol{C}|$ is the determinant of the matrix $2\pi\boldsymbol{C}$.

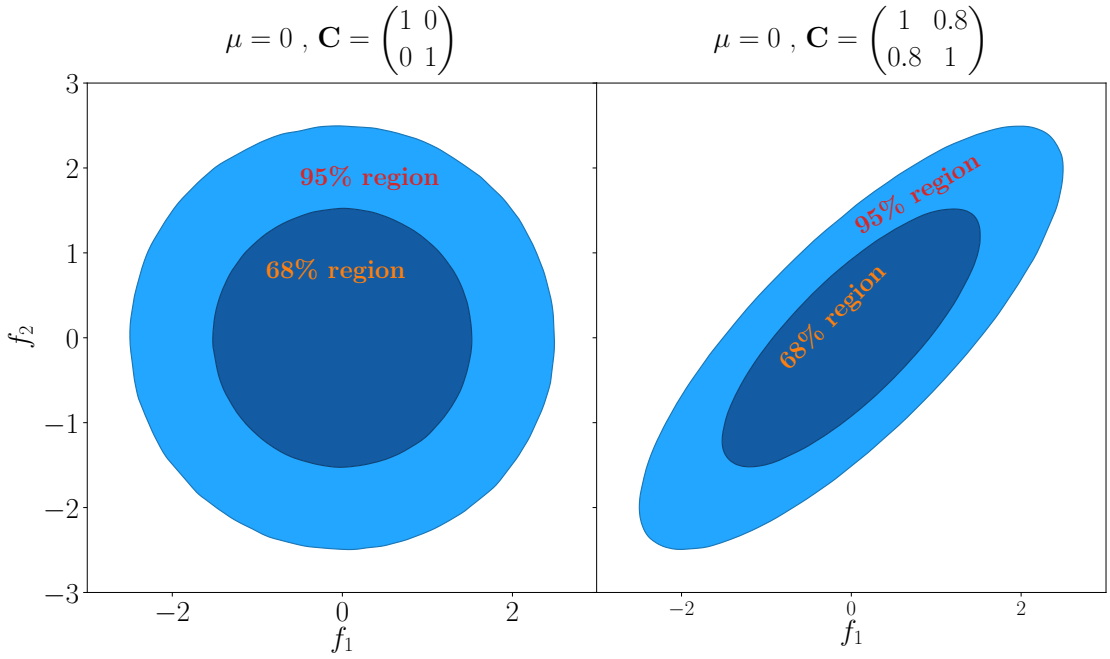

Figure 6. Visualising the PDFs of two different, 2-dimensional Gaussian random vectors (both with zero mean). In the left panel the covariance matrix $\boldsymbol{C}$ is diagonal such that the two components of the random vector $F = (F_1, F_2)$ are independent of each other. In the right panel, the covariance matrix has a significant off-diagonal component such that F_1 and F_2 become strongly correlated. For both cases we display the regions that contain 68% and 95% of the probability of the PDF. (These regions are often called the $1-\sigma$ and $2-\sigma$ confidence regions.)

Figure 6 visualises the PDF of 2-dimensional Gaussian random vectors with two different covariance matrices.

Both cosmological observations [21] and theoretical considerations [40, 41] indicate, that the initial density contrast field δ_i is well described as a *Gaussian random field*. We will see in the 2nd part of the course why this is the case, but for now let us only ask what it means for a random field to be *Gaussian*. A random field $f = f(\boldsymbol{x})$ is called a Gaussian random field if for any N locations in space $\boldsymbol{x}_1$, ... , $\boldsymbol{x}_N$ the vector

$$\boldsymbol{F} = (F_1 \ , \ \dots \ , \ F_N) \equiv (f(\boldsymbol{x}_1) \ , \ \dots \ , \ f(\boldsymbol{x}_N)) \tag{2.25}$$

is a multivariate Gaussian random variable. This is especially required to hold for any finite number N. Please note that for the sake of a simple notation we will give up some of the mathematical rigor we have employed before. Previously we have made sure to distinguish between random variables - as denoted by capital letters X, Z, F etc. - and concrete draws from those variables - as denoted with lower case letters x, z, f etc. Now - depending on the context - we will denote with (lower case) $f(\boldsymbol{x})$ both the random variable which is given by evaluating the random field f at $\boldsymbol{x}$ as well as concrete draws of the random field at that location. In particular, our notation will not distinguish between the Universe's initial density contrast δ_i and the random field from which this initial field has been drawn (by e.g. introducing the notation Δ_i for that random field).

In complete analogy to the multivariate Gaussian case, a Gaussian random field is fully

characterised by its expectation value

$$\mu(\boldsymbol{x}) = \langle f(\boldsymbol{x}) \rangle \tag{2.26}$$

and by the covariance between its values at every two locations $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$

$$C(\boldsymbol{x}_1, \boldsymbol{x}_2) = \langle [f(\boldsymbol{x}_1) - \mu(\boldsymbol{x}_1)] [f(\boldsymbol{x}_2) - \mu(\boldsymbol{x}_2)] \rangle \ . \tag{2.27}$$

The function $C(\boldsymbol{x}_1, \boldsymbol{x}_2)$ is also called the 2-point correlation function of the field, because it quantifies how much the values of the field f at any two locations are correlated.

The initial density contrast δ_i is actually a rather special Gaussian field, because it has the following additional properties.

- **zero-mean:**

  The expectation value of $\delta_i(\boldsymbol{x})$ is zero everywhere, i.e. $\mu(\boldsymbol{x}) \equiv 0$. This is a consequence of our very definition of δ as

  $$\delta = \frac{\rho - \langle \rho \rangle}{\langle \rho \rangle}$$

  (cf. Equation 1.1).

- **Homogeneity:**

  Cosmological observations have shown that the cosmic density field is statistically homogeneous on scales $\gtrsim 100$ Mpc, i.e. that the statistical properties of the field on these scales are the same at every location in the Universe. As a consequence, also the initial density field must have been a homogeneous random field. For a Gaussian random field, this means that the covariance function (or 2-point correlation function) for any two points $\boldsymbol{x}_1$, $\boldsymbol{x}_2$ only depends on the distance vector between the two points, i.e.

  $$C(\boldsymbol{x}_1, \boldsymbol{x}_2) = C(\boldsymbol{x}_1 - \boldsymbol{x}_2) \ . \tag{2.28}$$

  (It actually also means that $\mu(\boldsymbol{x})$ is independent of $\boldsymbol{x}$, but this is anyway satisfied by δ_i because it is a zero-mean field.)

- **Isotropy:**

  Observations have also show that the statistical properties of the cosmos are invariant under spatial rotations or equivalently - they are independent of direction. For a homogeneous Gaussian random field this means that the 2-point function between two points actually only depends on the absolute value of the distance between the two points, i.e.

  $$C(\boldsymbol{x}_1, \boldsymbol{x}_2) = C(|\boldsymbol{x}_1 - \boldsymbol{x}_2|) \ . \tag{2.29}$$

So to summarize: we have reasons to believe that δ_i is (a draw from) a zero-mean, homogeneous, isotropic Gaussian random field.

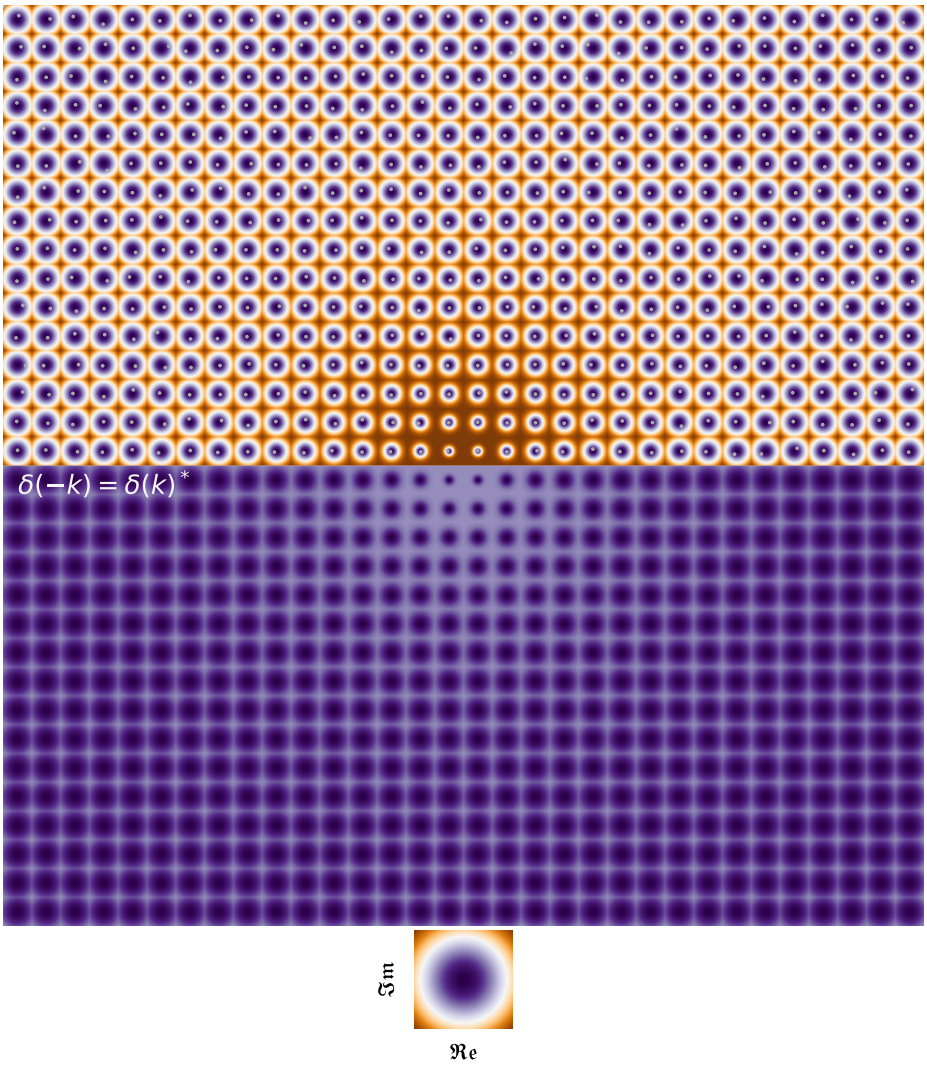

Figure 7. **FIGURE STILL WORK IN PROGRESS** A diagram for the density field in Fourier space.

Exercise 21

The initial density contrast in Figure 3 (blue line) was indeed drawn from a Gaussian random field. What about the later stages in the evolution of that simulated universe? Do you think they can also be interpreted as draws from (potentially different) Gaussian fields? What are your reasons for your answer?

Symbolically, we can think of δ_i as a huge vector with one entry for each location $\boldsymbol{x}$. The two-point function is the covariance matrix of that random vector. If we were able to find an inverse C^{-1} to that covariance, then the probability density function of δ_i would be given by

$$\text{“ } p(\delta_i) = \frac{1}{|2\pi C|^{1/2}} \exp\left\{-\frac{1}{2}\sum_{\boldsymbol{x}_1}\sum_{\boldsymbol{x}_2}\delta_i(\boldsymbol{x}_1)C^{-1}(\boldsymbol{x}_1,\boldsymbol{x}_2)\delta_i(\boldsymbol{x}_2)\right\} \text{ ”} \,. \tag{2.30}$$

Of course, there is hardly such thing as a “sum over all $\boldsymbol{x}$” and it is also hard to define a

"determinant" of a matrix that has an uncountably infinite number of elements. But let us nevertheless play around a bit with this analogy and try to make it more accurate. We could e.g. look at a regular grid of points in space as opposed to the entire continuum $\mathbb{R}^3$. For that purpose, let capital letters like I denote 3-dimensional indices $I = (i_1, i_2, i_3)$ and let Δ be some small distance. We can then consider the points

$$\{\boldsymbol{x}_I = (i_1\Delta, i_2\Delta, i_3\Delta) \; | i_1, i_2, i_3 \in \mathbb{Z}\} \; . \tag{2.31}$$

These are literally the points on a regular grid that spans the entire Universe and has a grid width of Δ . Let us denote the 2-point function restricted to these points by $\boldsymbol{C}_\Delta$ and the density contrast field restricted to these points as $\boldsymbol{\delta}_{i,\Delta}$. Both $\boldsymbol{C}_\Delta$ and $\boldsymbol{\delta}_{i,\Delta}$ now still have infinitely many elements, but at least it is only a countable infinity. Especially, we can now sum over all these elements, and attempt to write down a PDF for $\boldsymbol{\delta}_{i,\Delta}$ as

$$\text{"} \; p(\boldsymbol{\delta}_{i,\Delta}) = \frac{1}{|2\pi \boldsymbol{C}_\Delta|^{1/2}} \; \exp\left\{ -\frac{1}{2} \sum_I \sum_J \delta_i(\boldsymbol{x}_I) \left(C_\Delta^{-1}\right)_{IJ} \delta_i(\boldsymbol{x}_J) \right\} \; \text{"} \; . \tag{2.32}$$

If we define

$$C^{-1}(\boldsymbol{x}_I, \boldsymbol{x}_J) \equiv \frac{1}{\Delta^6} \; \left(C_\Delta^{-1}\right)_{IJ} \tag{2.33}$$

then this can be approximated as

$$\begin{aligned} \text{"} \; p(\boldsymbol{\delta}_{i,\Delta}) &= \frac{1}{|2\pi \boldsymbol{C}_\Delta|^{1/2}} \; \exp\left\{ -\frac{1}{2} \sum_I \Delta^3 \sum_J \Delta^3 \; \delta_i(\boldsymbol{x}_I) \frac{1}{\Delta^6} \left(C_\Delta^{-1}\right)_{IJ} \delta_i(\boldsymbol{x}_J) \right\} \\ &\approx \frac{1}{|2\pi \boldsymbol{C}_\Delta|^{1/2}} \; \exp\left\{ -\frac{1}{2} \int \mathrm{d}^3 x_1 \int \mathrm{d}^3 x_2 \; \delta_i(\boldsymbol{x}_1) C^{-1}(\boldsymbol{x}_1, \boldsymbol{x}_2) \delta_i(\boldsymbol{x}_2) \right\} \; \text{"} \; . \end{aligned} \tag{2.34}$$

This is actually still not a well defined PDF, as you will explore in the following exercise. But it provides us with a symbolic way of thinking about the entire field δ_i as one random quantity. As we will see in Section 4, this ill-defined way of thinking is quite powerful and can be used to derive concrete, non-trivial results regarding today's matter density field.

Exercise 22

Let $\boldsymbol{F}$ be an N-dimensional random vector with a multivariate Gaussian distribution with covariance matrix $\boldsymbol{C}$ and mean $\boldsymbol{\mu}$. And important 1-dimensional random variable that can be formed from $\boldsymbol{F}$ is the so called χ^2 statistic

$$\chi^2 = (\boldsymbol{F} - \boldsymbol{\mu})^T \boldsymbol{C}^{-1} (\boldsymbol{F} - \boldsymbol{\mu}) \; . \tag{2.35}$$

The PDF of χ^2 is in fact independent of $\boldsymbol{\mu}$ and $\boldsymbol{C}$ and depends only on N (which in the context of the χ^2 distribution is also called the number of degrees of freedom*). It can be shown that χ^2 has the expectation value*

$$\langle \chi^2 \rangle = N \tag{2.36}$$

and the standard deviation

$$\sqrt{\langle (\chi^2 - \langle \chi^2 \rangle)^2 \rangle} = \sqrt{2N} \; . \tag{2.37}$$

What can you conclude from this for typical value of the exponent in Equation 2.32? What does this mean for our attempt to define a PDF for the Gaussian random field?

Exercise 23
Using the fact that

$$\sum_K \left(C_\Delta^{-1}\right)_{IK} (C_\Delta)_{KJ} \;=\; \delta_{IJ} \;\equiv\; \begin{cases} 1 & \text{for } I = J \\ 0 & \text{else} \end{cases} \;, \tag{2.38}$$

show that in the limit $\Delta \to 0$ *the function* $C^{-1}(\boldsymbol{x}_1, \boldsymbol{x}_2)$ *which was defined in Equation 2.33 is related to the two-point function* $C(\boldsymbol{x}_1, \boldsymbol{x}_2)$ *via*

$$\int \mathrm{d}^3 y \; C^{-1}(\boldsymbol{x}_1, \boldsymbol{y}) C(\boldsymbol{y}, \boldsymbol{x}_2) = \delta_D^{(3)}(\boldsymbol{x}_1 - \boldsymbol{x}_2) \;, \tag{2.39}$$

where $\delta_D^{(3)}$ *is the 3-dimensional Dirac delta function.*

2.3 Fourier transform & the power spectrum

There is one more, very important question we will address about δ_i: how does it behave in Fourier space? The Fourier transform of δ_i is given by

$$\tilde{\delta}_i(\boldsymbol{k}) = \int \mathrm{d}^3 x \; \delta_i(\boldsymbol{x}) \; e^{-i\boldsymbol{k}\boldsymbol{x}} \;. \tag{2.40}$$

Since $\tilde{\delta}_i$ is a sum (or rather: an integral) over Gaussian random variables, we can expect it to be Gaussian as well. It is further more clear that $\langle \tilde{\delta}_i \rangle = 0$ because the Fourier transform is a linear operation. But how about the 2-point function of $\tilde{\delta}_i$? Note that $\tilde{\delta}_i$ can in general be complex valued, even if δ_i is real valued. In this situation the covariance is most conveniently defined as

$$\tilde{C}(\boldsymbol{k}_1, \boldsymbol{k}_2) = \langle \tilde{\delta}_i(\boldsymbol{k}_1) \tilde{\delta}_i(\boldsymbol{k}_2)^* \rangle \;, \tag{2.41}$$

where $*$ means complex conjugation. Inserting Equation 2.40 into this definition yields

$$\begin{aligned} \tilde{C}(\boldsymbol{k}_1, \boldsymbol{k}_2) &= \int \mathrm{d}^3 x_1 \mathrm{d}^3 x_2 \; \langle \delta_i(\boldsymbol{x}_1) \delta_i(\boldsymbol{x}_2) \rangle \; e^{-i\boldsymbol{k}_1\boldsymbol{x}_1 + i\boldsymbol{k}_2\boldsymbol{x}_2} \\ &= \int \mathrm{d}^3 x_1 \mathrm{d}^3 x_2 \; C(\boldsymbol{x}_2 - \boldsymbol{x}_1) \; e^{-i\boldsymbol{k}_1\boldsymbol{x}_1 + i\boldsymbol{k}_2\boldsymbol{x}_2} \;, \end{aligned} \tag{2.42}$$

where we have use the fact that δ_i is a homogeneous random field such that C only depends on the difference $\boldsymbol{x}_2 - \boldsymbol{x}_1$. This allows us to perform the variable transformation $\boldsymbol{x_2} \to \boldsymbol{y} = \boldsymbol{x}_2 - \boldsymbol{x}_1$ in the above integral, which leads to

$$\begin{aligned} \tilde{C}(\boldsymbol{k}_1, \boldsymbol{k}_2) &= \int \mathrm{d}^3 x_1 \mathrm{d}^3 x_2 \; C(\boldsymbol{y}) \; e^{-i\boldsymbol{k}_1\boldsymbol{x}_1 + i\boldsymbol{k}_2(\boldsymbol{y} + \boldsymbol{x}_1)} \\ &= \left(\int \mathrm{d}^3 x_1 \; e^{i(\boldsymbol{k}_2 - \boldsymbol{k}_1)\boldsymbol{x}_1} \right) \left(\int \mathrm{d}^3 y \; C(\boldsymbol{y}) \; e^{i\boldsymbol{k}_2\boldsymbol{y}} \right) \\ &\equiv (2\pi)^3 \; \delta_D^{(3)}(\boldsymbol{k}_2 - \boldsymbol{k}_1) \; P(\boldsymbol{k}_2) \;. \end{aligned} \tag{2.43}$$

Here we have use the fact that the Diract delta function can be expressed as

$$\delta_D^{(3)}(\boldsymbol{k}) = \frac{1}{(2\pi)^3} \int \mathrm{d}^3 x \; e^{i\boldsymbol{k}\boldsymbol{x}} \tag{2.44}$$

and the have denoted the Fourier transform of the 2-point function $C(\boldsymbol{x})$ (when viewed as a function of the difference between two points) by $P(\boldsymbol{k})$. The latter is called the *power spectrum* of the density contrast field. It actually only depends on the absolute magnitude k of the wave vector $\boldsymbol{k}$, and we can hence write it as $P(k)$. As you will show in the following exercise, this is a consequence of the fact that δ is not only a homogeneous but also an isotropic random field.

Exercise 24

NOT graded: *Show that $P(\boldsymbol{k})$ only depends on $k = |\boldsymbol{k}|$ if $C(\boldsymbol{x})$ only depends on $|\boldsymbol{x}|$*

The power spectrum is a **central quantity in cosmology**, and there are two points I would like you to take away from its above introduction:

A) The Fourier transform is an invertible operation. Hence, knowledge of P is equivalent to knowledge of C. Especially, the statistical properties of a homogeneous Gaussian random field are fully characterised by either the power spectrum P or the 2-point function C. This is crucial, because cosmological theory can predict the shape of the power spectrum of δ_i (cf. next lecture).

B) The Fourier transform diagonalises the covariance "matrix" if δ_i, as can be seen from the delta function appearing in Equation 2.43. This has the convenient consequence that we can actually invert $\tilde{C}$ (when seen as a matrix):

$$\tilde{C}^{-1}(\boldsymbol{k}_1, \boldsymbol{k}_2) = \frac{1}{(2\pi)^3} \frac{1}{P(k_2)} \, \delta_D^{(3)}(\boldsymbol{k}_2 - \boldsymbol{k}_1) \; . \tag{2.45}$$

Let us close this section by spelling out our symbolic view of the probability distribution of δ_i in a slightly more accurate way. Since δ_i is a function (and not just some vector in $\mathbb{R}^N$) this distribution will not be characterised by a probability density function but by a probability density functional[2]. In real space this functional is given by

$$\mathcal{P}[\delta_i] = \frac{1}{|2\pi C|^{1/2}} \, \exp\left\{ -\frac{1}{2} \int \mathrm{d}^3 x_1 \int \mathrm{d}^3 x_2 \; \delta_i(\boldsymbol{x}_1) C^{-1}(\boldsymbol{x}_1, \boldsymbol{x}_2) \delta_i(\boldsymbol{x}_2) \right\} \tag{2.46}$$

whereas in Fourier space it takes the form

$$\mathcal{P}[\tilde{\delta}_i] = \frac{1}{|2\pi \tilde{C}|^{1/2}} \, \exp\left\{ -\frac{1}{2} \int \frac{\mathrm{d}^3 k}{(2\pi)^3} \, \frac{|\tilde{\delta}_i(\boldsymbol{k})|^2}{P(k)} \right\} \; . \tag{2.47}$$

3 How likely is my power spectrum measurement?

3.1 The power spectrum of cosmic density fluctuations and its evolution

In the last section we have considered the initial density contrast δ_i of the Universe as a draw from a zero-mean, homogeneous, isotropic Gaussian random field. We have seen that the probability distribution function of such a field is completely characterised by its 2-point correlation function or equivalently, by its power spectrum which measures the variance of the Fourier modes of δ_i (cf. Equation 2.43).

[2] You can think of functionals as functions whose arguments are themselves functions. We will typically denote them with capital, calligraphic font and their arguments will be enclosed by square brackets. E.g. $\mathcal{F}[f]$ would be the functional $\mathcal{F}$ evaluated at the function f.

As we have explored in Exercise 21, gravitational collapse changes these initial density fluctuations in such a way that the density contrast $\delta(\boldsymbol{x}, t)$ at later times in cosmic history is not consistent with being drawn from a Gaussian random field anymore (cf. Section 4 for further insights into late-time non-Gaussianity). But we can nevertheless define the power spectrum $P(\boldsymbol{k}, t)$ of δ at any time t via the equation

$$\langle \tilde{\delta}(\boldsymbol{k}_1, t) \tilde{\delta}(\boldsymbol{k}_2, t)^* \rangle = (2\pi)^3 \delta_D(\boldsymbol{k}_2 - \boldsymbol{k}_1) P(k_1, t) \ , \tag{3.1}$$

where $*$ again denotes complex conjugation. This P is still the Fourier transform of the 2-point correlation function

$$C(|\boldsymbol{y}|) = \langle \delta(\boldsymbol{x}, t) \delta(\boldsymbol{x} + \boldsymbol{y}, t) \rangle \tag{3.2}$$

since the derivations of Equation 2.43 actually hold for any homogeneous random field. Note that the choice of the point $\boldsymbol{x}$ in Equation 3.2 is irrelevant, becaus δ is a homogeneous random field at all times t. Similarly, the orientation of the vector $\boldsymbol{y}$ Equation 3.2 is arbitrary, because δ is an isotropic random field. However, when we say that P is the Fourier transform of C then this is meant wrt. the 3-dimensional arguments $\boldsymbol{k}$ and $\boldsymbol{x}$ and not wrt. the 1-dimensional ones $|\boldsymbol{k}|$ and $|\boldsymbol{x}|$.

From Equation 3.1 we can see that $P(k, t)$ is proportional to the variance of the Fourier modes of δ at time t,

$$P(k, t) \propto \langle |\tilde{\delta}(\boldsymbol{k}, t)|^2 \rangle \ . \tag{3.3}$$

So, to understand how the shape of the power spectrum as a function of k influences the appearance of the cosmic density field it is useful to understand how different Fourier modes $\tilde{\delta}(\boldsymbol{k}, t)$ contribute to the real space density contrast $\delta(\boldsymbol{x}, t)$. Since delta is a real number at any $\boldsymbol{x}$, its Fourier transform has to satisfy

$$\tilde{\delta}(-\boldsymbol{k}, t) = \tilde{\delta}(\boldsymbol{k}, t)^* \ . \tag{3.4}$$

This allows us to simplify the relation between δ and its Fourier transform $\tilde{\delta}$ as

$$\delta(\boldsymbol{x}, t) = \int \frac{\mathrm{d}^3 k}{(2\pi)^3} \, \tilde{\delta}(\boldsymbol{k}, t) \, e^{i\boldsymbol{k}\boldsymbol{x}} \tag{3.5}$$

$$= \int_{\text{half of } \mathbb{R}^3} \frac{\mathrm{d}^3 k}{(2\pi)^3} \, 2\mathrm{Re}(\tilde{\delta}(\boldsymbol{k}, t) \, e^{i\boldsymbol{k}\boldsymbol{x}}) \ , \tag{3.6}$$

where the integration in the second line is over the one half of the space $\mathbb{R}^3$ that lies to one side of an arbitrary plane passing through $\boldsymbol{k} = 0$. If we furthermore express the complex valued Fourier modes $\tilde{\delta}(\boldsymbol{k}, t)$ as

$$\tilde{\delta}(\boldsymbol{k}, t) \equiv r(\boldsymbol{k}, t) \, e^{i\alpha(\boldsymbol{k}, t)} \tag{3.7}$$

then Equation 3.6 becomes

$$\begin{aligned} \delta(\boldsymbol{x}, t) &= \int_{\text{half of } \mathbb{R}^3} \frac{\mathrm{d}^3 k}{(2\pi)^3} \, 2r(\boldsymbol{k}, t) \mathrm{Re}(e^{i\{\boldsymbol{k}\boldsymbol{x} + \alpha(\boldsymbol{k}, t)\}}) \\ &= \int_{\text{half of } \mathbb{R}^3} \frac{\mathrm{d}^3 k}{(2\pi)^3} \, 2r(\boldsymbol{k}, t) \cos\left(\boldsymbol{k}\boldsymbol{x} + \alpha(\boldsymbol{k}, t)\right) \ . \end{aligned} \tag{3.8}$$

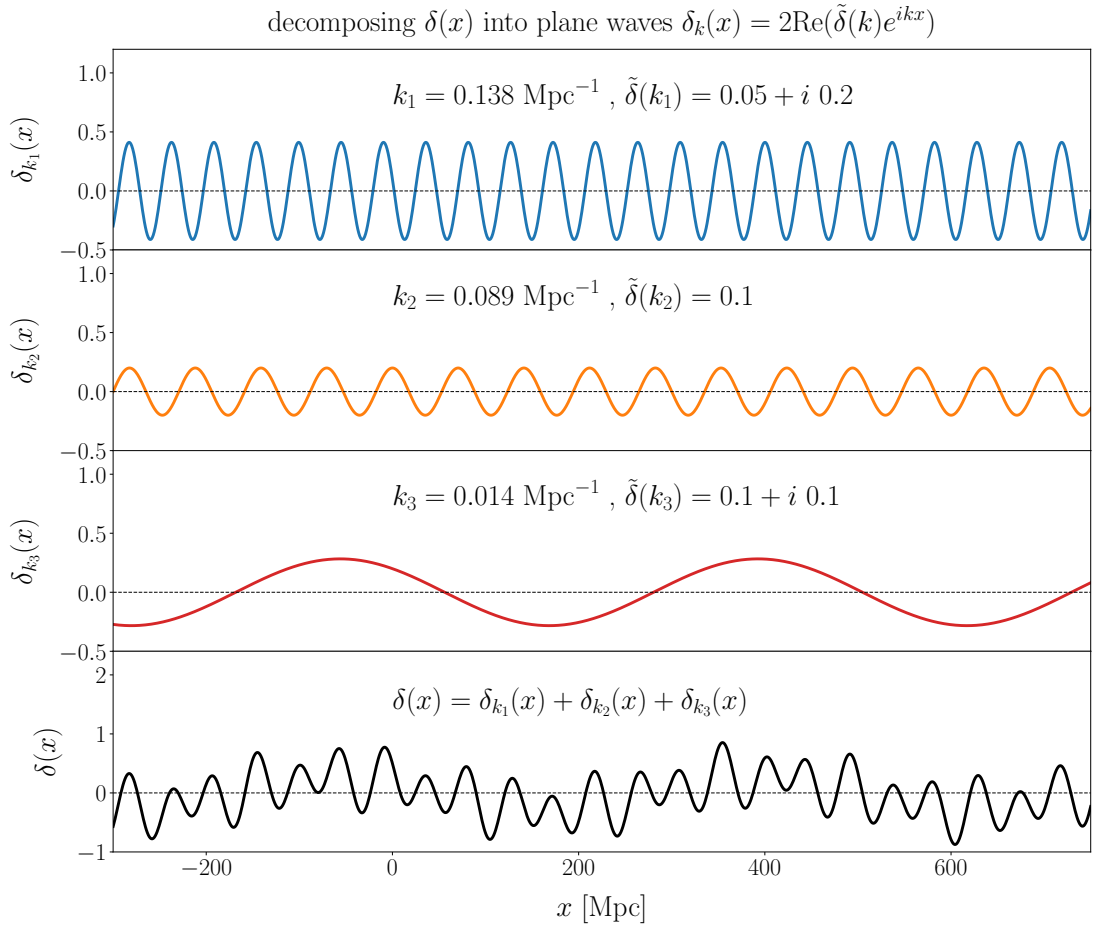

Figure 8. Combining three different sinusoidal waves (upper three panels) to form a somewhat realistic looking, 1-dimensional density contrast field (lowest panel). The real density contrast field is given by a continuous superposition of such waves, as indicated by the integral in Equation 3.8.

So the Fourier transform expresses $\delta(\boldsymbol{x}, t)$ as a superposition of cosinus-shaped waves with different wave vectors $\boldsymbol{k}$, phases $\alpha(\boldsymbol{k}, t)$ and amplitudes $2r(\boldsymbol{k},\ \mathrm{t})$, and the power spectrum $P(|\boldsymbol{k}|, t)$ determines the average amplitude of those waves. Figure 8 visualises this decomposition in the case of a 1-dimensional universe. In that figure three different sinusoidal waves are combined to for one, somewhat realistic looking density contrast field. Of course, the real density contrast field is not given by a discrete sum of waves but by a continuous superposition of waves, as indicated by the integral in Equation 3.8.

Exercise 25

Derive Equation 3.6 and Equation 3.8.

Exercise 26

Calculate the expectation value $\langle\tilde{\delta}(\boldsymbol{k}_1, t)\tilde{\delta}(\boldsymbol{k}_2, t)\rangle$, i.e. without the complex conjugation that is present in Equation 3.1. Hint: you only have to change a sign in Equation 2.43 and you have to show that $C(\boldsymbol{y}) = C(-\boldsymbol{y})$ (which is trivially true for homogeneous and isotropic random fields, but also holds for fields which are only homogeneous).

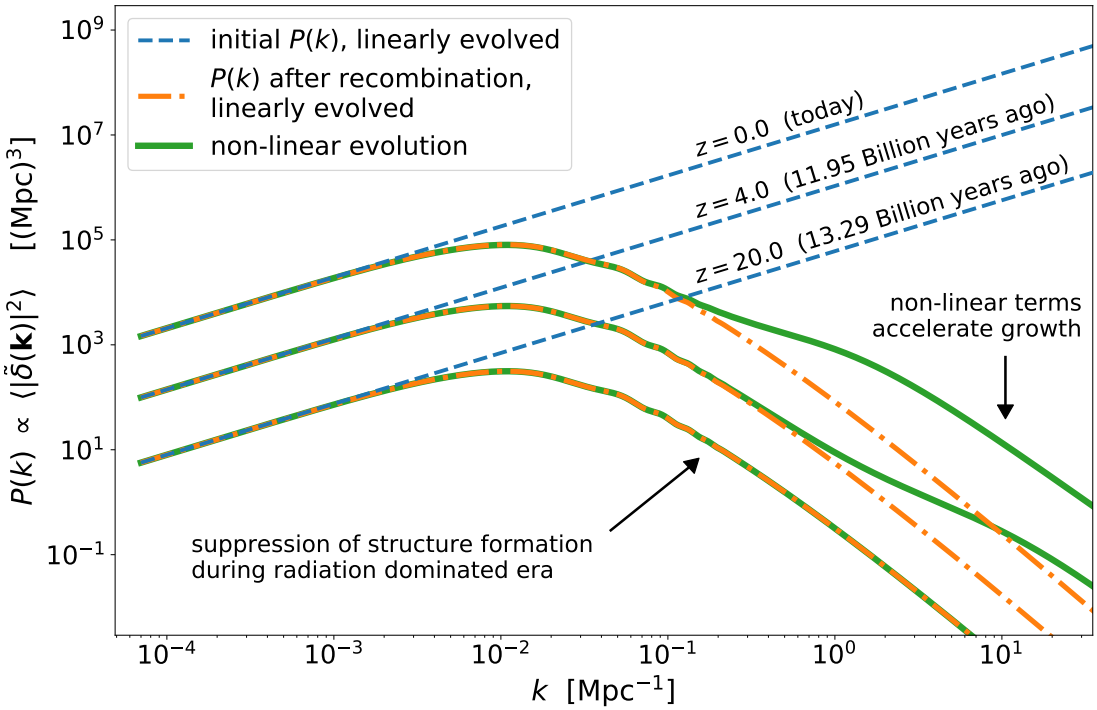

Figure 9. Evolution of the power spectrum of cosmic matter density contrast according to the cosmological standard model (see main text for details).

Exercise 27

Use your result from Exercise 26 together with Equation 3.1 to answer the following questions: Can the power spectrum $P(k)$ have complex values? When decomposing $\tilde{\delta}$ into its real and imaginary part as $\tilde{\delta} = \tilde{\delta}_{\mathrm{re}} + i\tilde{\delta}_{\mathrm{im}}$, what is the covariance of $\tilde{\delta}_{\mathrm{re}}$ and $\tilde{\delta}_{\mathrm{im}}$ (i.e. what is the expectation value $\langle\tilde{\delta}_{\mathrm{re}}\tilde{\delta}_{\mathrm{im}}\rangle$)?

Now why is it useful to decompose the real space density contrast $\delta(\boldsymbol{x}, t)$ into plane waves via Equation 3.8? We already saw one reason for this in the last section: the 2-point correlation function on Fourier space is diagonal, which significantly simplifies expressions for central quantities such as the probability density functional of the density contrast field (cf. Equations 2.46 and 2.47). Another reason is that the spatial derivatives in the continuity equation and the Euler equation (Equations 1.23 and 1.24) turn into multiplications after a Fourier transform. Hence, these equations become a system of ordinary differential equations (as opposed to a system of partial differential equations, where both time derivatives and spatial derivatives appear). This allows one to solve these equations in a systematic perturbation theory approach [see e.g. 9, for a review].

As we have seen above, the power spectrum $P(|\boldsymbol{k}|, t)$ has a very simple interpretation: it characterises the mean squared amplitude of the waves whose superposition constitutes the density contrast field. In Figure 9 you can see how the power spectrum evolves in time according to our current standard model in cosmology. It is not the goal of this course to derive in full detail how the understanding of the power spectrum that is shown in that figure comes about - other LMU courses are much better suited for that (e.g. the regularly held courses of Professor Mukhanov, Professor Weller or Dr. Sanchez). Instead, the following brief

summary will suffice for us:

- Shortly after the "big bang" (or rather: right after the era of cosmic inflation) the power spectrum is well approximated as a power law in k, i.e.

$$P_i(|\boldsymbol{k}|) = \mathcal{A}_\delta \left(\frac{k}{k_0}\right)^{n_s} \quad . \tag{3.9}$$

  Here the subscript i stands for "initial", k_0 is an arbitrary normalisation scale which for the purpose of this course we will choose to be 1 Mpc^{-1}, and $\mathcal{A}_\delta$ is a parameter characterising the overall amplitude of P_i. It can be shown theoretically that the power law index n_s should be slightly smaller than 1, and it has indeed been measured to be ≈ 0.96 [22]. This initial power spectrum is represented by the blue dashed lines in Figure 9. We will derive in detail where this shape of the initial perturbation comes from in the 2nd part of this course.

- From that initial moment in time until about 1 Billion years after the big bang the density fluctuations stay small enough for the linear approximation of Section 1.3 to be accurate. Hence, in this time the density contrast field evolved as

$$\delta(\boldsymbol{x}, t) = \delta(\boldsymbol{x}, t_i) \frac{D(t)}{D(t_i)} \ , \tag{3.10}$$

  where D is the linear growth factor we had encountered before and t_i could e.g. represent a time right after cosmic inflation. According to Section 1.3 D does not depend on $\boldsymbol{x}$, so it describes the evolution of density contrast both in real space and in Fourier space. Hence, in this time of linear growth the power spectrum evolves as

$$P_L(\boldsymbol{k}, t) = P(\boldsymbol{k}, t_i) \left(\frac{D(t)}{D(t_i)}\right)^2 \quad . \tag{3.11}$$

  There is however an important caveat to this statement, which is explained in the following point.

- Until 380,000 years after the big bang the energy (and hence matter) density of the Universe is dominated by a hot, dense plasma of radiation and other relativistic particles. Hence, in this epoch the pressure of the Universe's density field cannot - as we had done in Section 1 - be ignored. As a consequence, during that epoch the linear growth factor D *does* in fact depend on scale. In particular, when solving the evolution equations in Fourier space it becomes a function of k.

  The pressure of this primordial plasma prevents the formation of structures at small scales, i.e. for high wave numbers k. The effect of this on the power spectrum is usually encoded by the so called *transfer function* $T(k)$ as

$$P_L(\boldsymbol{k}, t) = \mathcal{A}_\delta \left(\frac{k}{k_0}\right)^{n_s} \left(\frac{D(t)}{D(t_i)}\right)^2 T(k)^2 \ . \tag{3.12}$$

  If you compare the blue dashed line to the orange dash-dotted lines in Figure 9 you can see the impact of the transfer function: $T(k)$ approaches 1 as $k \to 0$, i.e. on very large scales the growth of structures follows is unaffected by it. But at small scales the pressure of the relativistic nature of the primordial plasma significantly suppresses the formation of structures. The exact shape of $T(k)$ can be calculated by solving the so called Boltzmann equation [41].

- About 1 Billion years after the big bang (though the exact time depends on the range of k one is interested in) the linear approximation starts to fail and all terms in the continuity and Euler equation need to be considered. The non-linear terms on those equations do in fact accelerate structure formation compared to purely linear growth, as can be seen by the solid green lines in Figure 9.

The theoretical predictions for the power spectra shown in Figure 9 depend on a number of a priori unknown parameters of the cosmological standard model. Roughly speaking, those parameter characterise:

A) the initial speed of cosmic expansion right after the big bang,

B) the total matter density after the big bang,

C) the fraction of that density that is made up of dark matter,

D) the spectral index n_s of the initial power law index,

E) the amplitude $\mathcal{A}_\delta$ of the initial power spectrum.

None of these pieces of information can be directly predicted by cosmological theory (though, in the case of n_s a good theoretical guess can be derived). So to fix the cosmological standard model, these parameters need to be measured from observational data. Let us in the following investigate, in an abstract way, how such a measurement can be performed.

3.2 Fitting a model to data: How probable are different model parameters?

A major part of scientific research consists of the development of theoretical models for different aspects of nature. These models do often depend on a set of model parameters (e.g. the mass of the Higgs boson or the fine structure constant in the standard model of particle physics) and a common task of scientific data analysis is to determine the values of those model parameters for which the model fits best to observations. This typically happens along a very simple pattern: measure some data, workout theoretical predictions for that data from the model at hand, and then determine which model parameters make those predictions resemble the observed data most closely.

To obtain a better intuition for this process of parameter inference, let us consider the following, simplistic example.

- **Example 1:**

  Imagine you have a model of the Universe which contains an unknown, fundamental constant α. From cosmological observations you have obtained a measurement $\hat{\alpha}$ of this constant. You know that $\hat{\alpha}$ suffers from measurement uncertainties. But you have good reasons to believe that these uncertainties are well described by Gaussian noise, i.e. that $\hat{\alpha}$ was drawn from a Gaussian distribution with expectation value α standard deviation σ. Furthermore, σ is known to you exactly.

Exercise 28

Based on the measurement $\hat{\alpha}$ from ***Example 1****, give a range $[\alpha_{\min}, \alpha_{\max}]$ of possible values of the fundamental constant α such that you are 68.3% sure that the true value of α lies within that range.*

Exercise 29

What do you mean when you say that you are 68.3% sure that the true value of α lies within the range $[\alpha_{\min}, \alpha_{\max}]$ you determined in Exercise 28?

An answer to Exercise 28 that many people may give is

$$[\alpha_{\min}(\hat{\alpha}, \sigma), \alpha_{\max}(\hat{\alpha}, \sigma)] = [\hat{\alpha} - \sigma, \hat{\alpha} + \sigma] \ . \tag{3.13}$$

They may then answer Exercise 29 by saying: if we had N independent realisations $\hat{\alpha}_i$, $i = 1, \ldots, N$, and N corresponding intervals $[\alpha_{\min}(\hat{\alpha}_i, \sigma), \alpha_{\max}(\hat{\alpha}_i, \sigma)]$ then the true value α should lie in approximately 68.3% of those intervals if N is large. This is the interpretation of probability based on frequencies in recurrent experiments that we have already encountered in Section 2 - the so called *frequentist interpretation of probability*. Within the frequentist interpretation the right hand side of Equation 3.13 is actually not the only interval you could give to answer Exercise 28 (you could challenge yourselves to find a different interval that will contain α in the same fraction of repeated experiments). But it is indeed the smallest one.

We had seen in Section 2 that the frequentist interpretation of probability faces difficulties. If we only consider finite numbers of repeated experiments then it suffers from circularity: we would be defining probabilities in the context of one random process with probabilities in the context of another random process. If we instead try to define the probability of an outcome of a random process via the frequency of occurrence of that outcome in an infinite number of independent realisations of that process, that we face the problem that frequencies are not well defined in infinite sequences (cf. Figure 4).

Circular answers are not necessarily bad answers. It was e.g. argued by [60, 61] that science as a whole consists of cycles of incomplete but ever refined answers to scientific questions, and that a circular definition of probability would be appropriate to this process. But let us nevertheless look for an alternative definition. Such a definition can e.g. found in the approach of betting preferences [35]. Assume somebody were to propose a bet on the outcome of the experiment in **Example 1**: The bet is on whether or not α is indeed contained in the interval $[\alpha_{\min}, \alpha_{\max}]$ you gave in Exercise 28. If you spend an amount of money M on it then the person will pay you back $X \cdot M$ if $\alpha \in [\alpha_{\min}, \alpha_{\max}]$, while you loose all your money otherwise. This of course assumes that we have some way of knowing the true value of α, but let's just go with that. Now there will be some minimal value of X that is required for you to accept this bet. And a way to define the probability of $\alpha \in [\alpha_{\min}, \alpha_{\max}]$ is via this minimum payback rate would be to set

$$P(\alpha \in [\alpha_{\min}, \alpha_{\max}]) = \frac{1}{X_{\text{accept}}} \ . \tag{3.14}$$

So when you say you are 68.3% sure that $\alpha \in [\alpha_{\min}, \alpha_{\max}]$ then it would mean that you accept the above bet as long as $X > 1.0/0.683$. This is of course an extremely subjective statement. Different people may be accepting the above bet for very different rates. One might try to reduce the choice of X_{accept} to the frequentist definition of probability by saying that everyone chooses their rate of acceptance such that they would on average earn (or at

least: not loose) money when taking part in repeated bets. But: nobody is forced to decide whether or not to accept a bet based on what fraction of bets they would win in the long run. In fact, if you are aware of the circular nature of frequentist probability definitions, then you may not even ascribe meaning to the idea of making long term gains on average. So the above definition of probability is a subjective one and it is hence also part of the so called *subjective interpretations of probability*. There is only one restriction that is usually made within this approach. A person's betting preferences should be *rational* in the sense that it follows some minimal requirements of logical consistency. For example, imagine that a random process has two outcomes A and B (among other possible outcomes C, D etc.) and assume that there is a logical implication

$$A \Rightarrow B \ ,$$

i.e. whenever A occurs we can be certain that B has also occurred. Then we demand that a rational person ascribes probabilities to A and B such that

$$P(A) \leq P(B) \ . \tag{3.15}$$

More formally, we would require a rational person to ascribe probabilities such that

- $P(A) \geq 0$ for all outcomes A,
- $P(\text{“anything happens”}) = 1$,
- $P(A \text{ or } B) = P(A) + P(B)$ if A and B are mutually exclusive outcomes.

It can be shown that these three axiom - which are the standard, Kolmogorov axioms of probability - are exactly the rules you have to adhere to if you want to prevent that somebody convinces you of a bet in which you always loose money (i.e. regardless of the outcome of the random process on which the bet is based, cf. the *Dutch book argument* [35, 57]).

Within the subjective interpretation of probability any answer to Exercise 29 that adheres to the above rationality requirements is a valid one. This would also mean that any answer you gave to Exercise 28 is correct as long as it conforms with the answer you have given in Exercise 29. It would hence seem futile for me to try to derive for you what the correct answer to both exercises is. Instead, I will simply tell you which answer is most commonly given in cosmological research, and I will leave it for you to decide whether you find that answer convincing. To proceed with that, let us first spell out a more general situation for parameter inference:

i) We measure a number of data points and arrange them into a vector $\hat{\boldsymbol{d}}$, the so called *data vector*. Such a measurement will be accompanied by measurement uncertainties (“error bars”), and we usually think of $\hat{\boldsymbol{d}}$ as a realisation of a random vector with PDF $p(\hat{\boldsymbol{d}})$.

ii) We then consider a theoretical prediction $p_{\text{theo}}(\hat{\boldsymbol{d}})$ for that PDF. This prediction may depend on a number of model parameters, which we can arrange into a *parameter vector* $\boldsymbol{\pi}$.

iii) We might be confident enough in our prediction to say that (or simply make the approximation that) $p_{\text{theo}}(\hat{\boldsymbol{d}})$ would equal $p(\hat{\boldsymbol{d}})$ if we only knew the correct model parameters. We can then denote our PDF prediction as $p_{\text{theo}}(\hat{\boldsymbol{d}}) = p(\hat{\boldsymbol{d}}|\boldsymbol{\pi})$, i.e. we interpret it as the true PDF of $\hat{\boldsymbol{d}}$ if $\boldsymbol{\pi}$ were the true parameters.

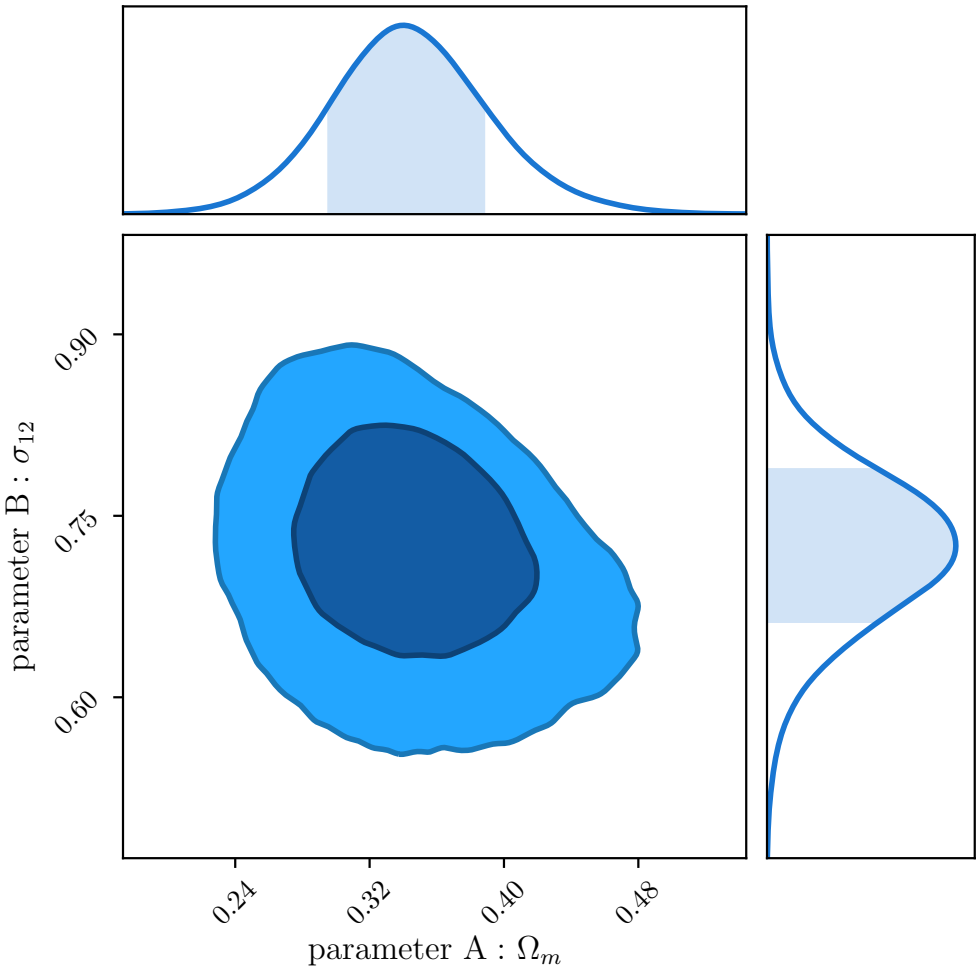

Figure 10. The posterior density derived by the Dark Energy Survey collaboration for two of their model parameters [20]: Ω_m, which in a flat Universe characterises the fraction of the Universes energy density that is made up of matter, and σ_{12} which measures how large the standard deviation of cosmic density perturbations would be on a scale of 12 Mpc, if structure growth was purely linear. The dark blue region indicates the iso-density contour that contains $\approx 68.3\%$ of posterior probability - the so called 1-σ contour. The light blue region indicates the iso-density contour that contains $\approx 95.5\%$ of posterior probability - the so called 2-σ contour.

iv) Given our measurement $\hat{\boldsymbol{d}}$ and our PDF model $p(\hat{\boldsymbol{d}}|\boldsymbol{\pi})$ we would now like to infer information about which parameters $\boldsymbol{\pi}$ are likely to be the true parameters.

These four steps may look somewhat different from the situation we have considered in **Example 1**. But that example does indeed fit into the above scheme.

Exercise 30

How does ***Example 1*** *map onto steps* i) *to* iv) *of the above inference scheme?*

Within the general situation described above, cosmological scientists would typically answer questions like those in Exercise 28 by deriving a PDF for the parameters $\boldsymbol{\pi}$ from the measured data vector $\hat{\boldsymbol{d}}$ - the so called *posterior density function* $p(\boldsymbol{\pi}|\hat{\boldsymbol{d}})$ (or for short: *the posterior*). Statements like "we are 68.3% certain that $\boldsymbol{\pi}$ lies within the sub-volume V of parameter space"

would then mean that

$$\int_{\boldsymbol{\pi}\in V} \mathrm{d}\boldsymbol{\pi}\ p(\boldsymbol{\pi}|\hat{\boldsymbol{d}}) = 0.683\ . \tag{3.16}$$

In Figure 10 you see an example of such a parameter posterior, as derived from the team of the Dark Energy Survey in their year-3 analysis [20]. To derive such posteriors, most of cosmological research employs Bayes' theorem. This theorem is concerned with joint probability distributions of pairs (X, Y) of random variables and states that

$$p(x|y)\ p(y)\ =\ p(x,y)\ =\ p(y|x)\ p(x)\ . \tag{3.17}$$

Here $p(x, y)$ is the joint PDF of X and Y while $p(x)$ and $p(y)$ are the individual PDFs of X and Y. Furthermore, $p(x|y)$ and $p(y|x)$ are the PDFs of X and Y *conditional on each other*. E.g. $p(x|y)$ is the PDF of X given the additional information that Y has taken the value y. The statement of Equation 3.17 is then almost trivial: the probability that both x and y happen is equal to the probability that y happens *times* the probability that x happens if y also happens (and vice verse for the right hand side of the equation).

Following Bayes theorem, we can derive the posterior of $\boldsymbol{\pi}$ as

$$p(\boldsymbol{\pi}|\hat{\boldsymbol{d}}) = \frac{p(\hat{\boldsymbol{d}}|\boldsymbol{\pi})p(\boldsymbol{\pi})}{p(\hat{\boldsymbol{d}})}\ . \tag{3.18}$$

Here $p(\hat{\boldsymbol{d}}|\boldsymbol{\pi})$ is our theoretical model for the PDF of the measurement noise that is present in $\hat{\boldsymbol{d}}$ (which in the context of parameter inference is also called the *likelihood function* of the parameters $\boldsymbol{\pi}$). The PDF $p(\hat{\boldsymbol{d}})$ appearing in the denominator above is called the *Bayesian evidence*. From the normalisation condition

$$1 = \int \mathrm{d}\boldsymbol{\pi}\ p(\boldsymbol{\pi}|\hat{\boldsymbol{d}}) \tag{3.19}$$

one can see that the Bayesian evidence is given by

$$p(\hat{\boldsymbol{d}}) = \int \mathrm{d}\boldsymbol{\pi}\ p(\hat{\boldsymbol{d}}|\boldsymbol{\pi})p(\boldsymbol{\pi})\ . \tag{3.20}$$

So $p(\hat{\boldsymbol{d}})$ is a $\boldsymbol{\pi}$-independent normalisation factor that does not impact the shape of the posterior. In contrast to that, the PDF $p(\boldsymbol{\pi})$ does impact the shape of the posterior. It is called the *prior density function* (or prior) and is supposed to encode a priori knowledge we had about the parameters before measuring $\hat{\boldsymbol{d}}$. So Equation 3.18 can be read as updating our knowledge: before measuring $\hat{\boldsymbol{d}}$ it was characterised by the prior density, and after measuring $\hat{\boldsymbol{d}}$ it is characterised by the posterior density.

To see how this procedure performs, let us consider a second example.

- **Example 2:**

  Imagine that you have measured two quantities $\hat{d}_1$ and $\hat{d}_2$ which you arrange into a 2-dimensional data vector $\hat{\boldsymbol{d}} = (\hat{d}_1, \hat{d}_2)$. You know that this measurement suffers from measurement uncertainties. But you have good reasons to believe that these uncertainties are well described by multivariate Gaussian noise, i.e. that $\hat{\boldsymbol{d}}$ was drawn from a

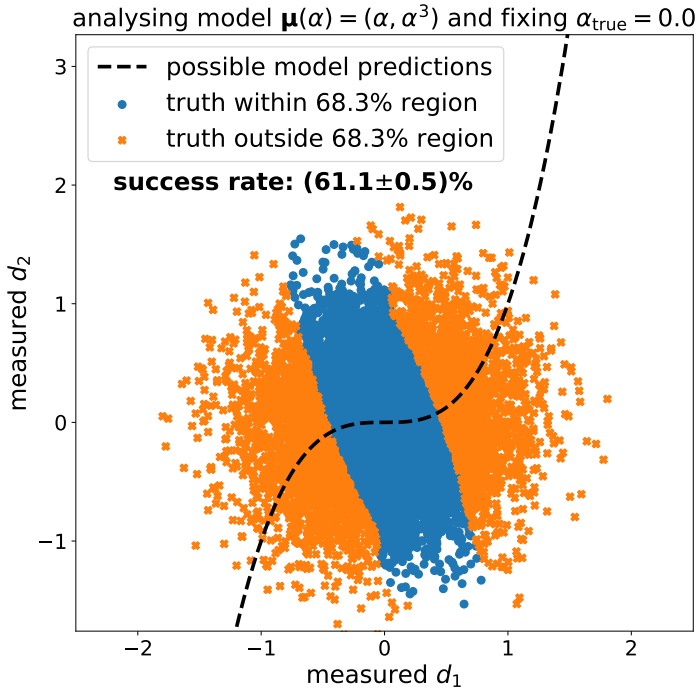

Figure 11. Measurements of the data vector $\hat{\boldsymbol{d}} = (\hat{d}_1, \hat{d}_2)$ in 10,000 parallel universes, assuming that the true parameter α in all of these Universes equals 0. The blue points correspond to universes in which this true value is included in the 68.3% confidence regions that scientists derived from their posterior density of α, while the orange points represent Universes, in which this is not the case. It turns out, that the blue points only constitute about 61% of all the considered parallel universes.

2-dimensional Gaussian distribution with some mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{C}$. Furthermore, you know that the covariance matrix is given by

$$\boldsymbol{C} = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix} , \tag{3.21}$$

i.e. $\hat{d}_1$ and $\hat{d}_2$ are uncorrelated and both have the same, known standard deviation σ. The exact mean vector $\boldsymbol{\mu}$ on the other hand is unknown to you. But you do have a theoretical model for it. This model is given by

$$\boldsymbol{\mu}(\alpha) = \begin{pmatrix} \alpha \\ \alpha^3 \end{pmatrix} , \tag{3.22}$$

where α is a free parameter of the model, which represents an unknown fundamental constant of physics.

In this example, our parameter vector $\boldsymbol{\pi} = (\alpha)$ is again one-dimensional and our PDF model $p(\hat{\boldsymbol{d}}|\alpha)$ is given by the Gaussian PDF

$$p(\hat{\boldsymbol{d}}|\alpha) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{1}{2} (\hat{d}_1 - \alpha \ , \ \hat{d}_2 - \alpha^3) \ \boldsymbol{C}^{-1} \ \begin{pmatrix} \hat{d}_1 - \alpha \\ \hat{d}_2 - \alpha^3 \end{pmatrix} \right) . \tag{3.23}$$

Within the context of that example let us carry out the following numerical experiment. We assume that the true value of the model parameter is $\alpha = 0$. Hence, the measured data vector

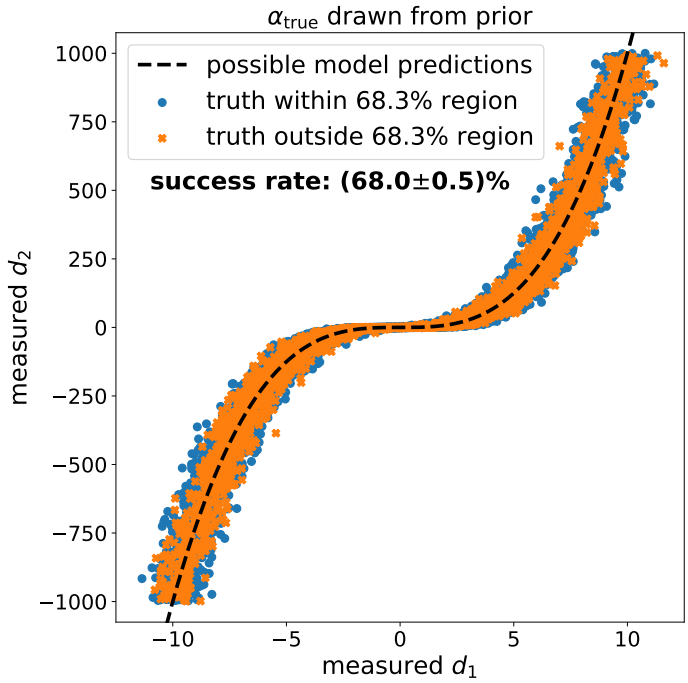

Figure 12. The same numerical experiment as in Figure 11, but this time the true value of α in each of the parallel universes is randomly drawn from the uniform prior $\alpha \in [-10, 10]$.

$\hat{\boldsymbol{d}}$ is drawn from the PDF

$$p(\hat{\boldsymbol{d}}|\alpha = 0) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{1}{2}(\hat{d}_1\ ,\ \hat{d}_2)\ \boldsymbol{C}^{-1}\ \begin{pmatrix}\hat{d}_1\\ \hat{d}_2\end{pmatrix}\right)\ . \tag{3.24}$$

We now draw 10,000 independent realisations of this measurement, assuming that $\sigma = 0.5$. These measurements are displayed in Figure 11. You could think of them as measurements of the same quantity, but carried out in 10,000 different parallel universes with different initial conditions. In each of these universes we let scientists infer posterior densities for the parameter α according to Equation 3.18. In particular, let's assume that all these scientists determine a region of 68.3% confidence for α. To apply Equation 3.18 they need a prior density $p(\alpha)$ which we assume to be a very wide uniform distribution, which would represent a lack of any prior knowledge about the value of α.

Exercise 31

On average, how many of the scientists in these parallel universes should derive 68.3% *confidence regions that include the true value of* $\alpha = 0$*?*

It turns out, that only in about 61% of our parallel Universes scientists have derived 68.3% confidence regions that include the true parameter $\alpha = 0$ (the blue points in Figure 11).

Exercise 32

What has gone wrong in our numerical experiment?

A perfectly valid answer to Exercise 32 would be to say: Why the hell did we even expect it to work?? Bayes theorem, Equation 3.18, applies to pairs of random variables. But in the

above numerical experiment the true parameter α was not a random variable, drawn from some distribution, but it was kept to $\alpha = 0$ for all of our 10,000 realisations of the data vector. There is no reason to believe that the regions of 68% credibility of our 10,000 scientists should contain this value of α in 68% of the times. In fact, any success rate (even above 68%) can be observed in appropriately altered versions of **Example 2**.

To demonstrate that Bayes' theorem indeed does work in the situations where it is supposed to, let us consider a version of the above numerical experiment where also the true value of α is drawn from a random distribution in each of our parallel universes. In particular, let us draw α from a uniform distribution on the interval $[-10, 10]$ and let us then draw the measurement $\hat{\boldsymbol{d}}$ from a Gaussian distribution with the mean vector $\boldsymbol{\mu} = (\alpha, \alpha^3)$ and with the same covariance as before. We repeat this 10,000 times and the resulting data vectors are shown in Figure 12. We again apply Bayesian parameter inference on each of these measurements. To do so, we choose the prior distribution in Equation 3.18 to be exactly the distribution from which the true α's have been drawn. This time, the fraction of our imagined scientists that find the true α of their respective universe inside of their region of 68.3% credibility is indeed about 68%.

Now what do the above findings mean? Is the method of parameter inference applied by cosmologists (including the authors of this script) just a scam? Let us postpone a discussion of the significance of the above results for cosmological parameter inference to a later time in this course. For now, we will turn to more practical matters: how would we even measure the power spectrum of cosmic density fluctuations, and what would the PDF of the measurement uncertainties of such a measurement look like?

Exercise 33

Using e.g. `python` *code, set up a version of the above experiment, where the data vector* $\hat{\boldsymbol{d}} = (\hat{d})$ *is only one-dimensional, and where scientists use a model* $\mu(\alpha) = \alpha^3$*. Assume that the true value of* α *is* 0*. How many scientists include this value within their* 68.3% *confidence regions? What happens if the true value of* α *is 1 instead? Hint: given a measurement* $\hat{d}$ *the best-fitting parameter is* $\alpha_{\rm BF} = \hat{d}^{1/3}$*. Create an array of possible* α *values around this location (e.g. using* `numpy.linspace`) and calculate the posterior density of α on this grid of points. To know how wide your array should be and how small your steps in α should be, you can use standard error propagation. (recall that the standard deviation of $\hat{d}$ is σ, which you can take to be 0.5 as before.)

3.3 The likelihood function of power spectrum measurements

Now how would we measure the power spectrum from the cosmic density contrast field $\delta(\boldsymbol{x})$ at some fixed time t? Given any particular realisation of the Fourier modes $\tilde{\delta}(\boldsymbol{k})$ we could e.g. consider the following approximate version of Equation 3.1:

$$\tilde{\delta}(\boldsymbol{k}_1)\tilde{\delta}(\boldsymbol{k}_2)^* \approx (2\pi)^3 \delta_D(\boldsymbol{k}_1 - \boldsymbol{k}_2)\ P(|\boldsymbol{k}_1|)\ . \tag{3.25}$$

If we could invert this equation, then it would provide us with an estimate of $P(|\boldsymbol{k}_1|)$. But this is of course prevented by the presence of the factor $\delta_D(\boldsymbol{k}_1 - \boldsymbol{k}_2)$. So let us try to "integrate out" the delta distribution by considering a small cube

$$V_l(\boldsymbol{k}) \equiv \left[k_x - \frac{l}{2}\ ,\ k_x + \frac{l}{2}\right] \times \left[k_y - \frac{l}{2}\ ,\ k_y + \frac{l}{2}\right] \times \left[k_z - \frac{l}{2}\ ,\ k_z + \frac{l}{2}\right] \tag{3.26}$$

of side length l around each wave vector $\boldsymbol{k} = (k_x, k_y, k_z)$, and by averaging the Fourier modes over this cube, i.e.

$$\tilde{\delta}_l(\boldsymbol{k}) \equiv \frac{1}{l^3} \int\limits_{V_l(\boldsymbol{k})} d^3k' \; \tilde{\delta}(\boldsymbol{k}') \; . \tag{3.27}$$

The variance of $\tilde{\delta}_l(\boldsymbol{k})$ is then given by

$$\begin{aligned} \langle |\tilde{\delta}_l(\boldsymbol{k})|^2 \rangle &= \frac{1}{l^6} \int\limits_{V_l(\boldsymbol{k})} d^3k'_1 \int\limits_{V_l(\boldsymbol{k})} d^3k'_2 \langle \tilde{\delta}(\boldsymbol{k}_1) \tilde{\delta}(\boldsymbol{k}_2)^* \rangle \\ &= \frac{(2\pi)^3}{l^6} \int\limits_{V_l(\boldsymbol{k})} d^3k'_1 \int\limits_{V_l(\boldsymbol{k})} d^3k'_2 \; \delta_D(\boldsymbol{k}_1 - \boldsymbol{k}_2) \; P(|\boldsymbol{k}_1|) \\ &= \frac{(2\pi)^3}{l^6} \int\limits_{V_l(\boldsymbol{k})} d^3k' \; P(|\boldsymbol{k}'|) \; . \end{aligned} \tag{3.28}$$

If we choose the cube side length l so small that $P(|\boldsymbol{k}'|)$ is approximately the same for all $\boldsymbol{k}' \in V_l(\boldsymbol{k})$ then this can be further simplified as

$$\begin{aligned} \langle |\tilde{\delta}_l(\boldsymbol{k})|^2 \rangle &\approx \frac{(2\pi)^3 P(|\boldsymbol{k}|)}{l^6} \int\limits_{V_l(\boldsymbol{k})} d^3k' \; 1 \\ &= \frac{(2\pi)^3 P(|\boldsymbol{k}|)}{l^3} \; . \end{aligned} \tag{3.29}$$

This equation can indeed be inverted and we could attempt to estimate $P(|\boldsymbol{k}|)$ as

$$\hat{P}(\boldsymbol{k}) = \frac{l^3}{(2\pi)^3} \; |\tilde{\delta}_l(\boldsymbol{k})|^2 = \frac{l^3}{(2\pi)^3} \; \left(\tilde{\delta}_{l,\mathrm{re}}(\boldsymbol{k})^2 + \tilde{\delta}_{l,\mathrm{im}}(\boldsymbol{k})^2 \right) \; , \tag{3.30}$$

where in the last equation we have expressed $\tilde{\delta}_l(\boldsymbol{k})$ in terms of its real and imaginary part. In Exercise 27 we had shown that the covariance of the real and imaginary part of the unsmoothed density contrast modes $\tilde{\delta}(\boldsymbol{k})$ is 0. This carries over to the smoothed field $\tilde{\delta}_l(\boldsymbol{k})$. One can even show that $\tilde{\delta}_{l,\mathrm{re}}$ and $\tilde{\delta}_{l,\mathrm{im}}$ are completely independent random variables (which is a stronger statement then saying that their covariance is 0). To avoid unnecessary technical details, let me only state that this is a consequence of the statistical homogeneity of the density contrast field[3]. Another consequence of homogeneity is the fact that both real and imaginary part of $\tilde{\delta}_l$ have the same variance

$$\langle \tilde{\delta}_{l,\mathrm{re}}(\boldsymbol{k})^2 \rangle = \langle \tilde{\delta}_{l,\mathrm{im}}(\boldsymbol{k})^2 \rangle = \frac{1}{2} \frac{(2\pi)^3 P(|\boldsymbol{k}|)}{l^3} \tag{3.31}$$

and indeed the same probability distribution. We will make use of this in the following derivations.

On average, Equation 3.30 does indeed give us the correct value of the power spectrum (modulo the assumption we have made about small boxes), but for each individual realisation of the cosmic density field it will only provide a very noisy estimate. To see this, let us calculate

[3] Translations in real space introduce phase factors in Fourier space, i.e. they rotate real and imaginary part into each other.

the variance of $\hat{P}$. Since $\hat{P}$ is proportional to the sum $\tilde{\delta}_{l,\mathrm{re}}(\boldsymbol{k})^2+\tilde{\delta}_{l,\mathrm{im}}(\boldsymbol{k})^2$ and since $\tilde{\delta}_{l,\mathrm{re}}(\boldsymbol{k})$ and $\tilde{\delta}_{l,\mathrm{im}}(\boldsymbol{k})$ are identically and independently distributed, it suffices to calculate that variance of say $\tilde{\delta}_{l,\mathrm{re}}(\boldsymbol{k})^2$. The latter would be given by

$$\mathrm{Var}(\tilde{\delta}_{l,\mathrm{re}}(\boldsymbol{k})^2) = \langle\tilde{\delta}_{l,\mathrm{re}}(\boldsymbol{k})^4\rangle - \langle\tilde{\delta}_{l,\mathrm{re}}(\boldsymbol{k})^2\rangle^2 \ . \tag{3.32}$$

In order to calculate this, we would need to know the 4th order moment of $\tilde{\delta}_{l,\mathrm{re}}$. Depending on the magnitude of the wave vector $\boldsymbol{k}$ this could require very technical calculation which would go beyond the scope of this course (but see e.g. [9] for a standard review). But we can nevertheless obtain an order-of-magnitude estimate by assuming that $\tilde{\delta}_{l,\mathrm{re}}(\boldsymbol{k})$ is a Gaussian random variable. A least at large physical scales (small $|\boldsymbol{k}|$) or at early cosmic times this is a good approximation. Now Wick's theorem tells us that the joint moment of four Gaussian random variables A, B, C and D with vanishing means is given by

$$\langle ABCD\rangle = \langle AB\rangle\langle CD\rangle + \langle AC\rangle\langle BD\rangle + \langle AD\rangle\langle BC\rangle \tag{3.33}$$

Setting $A = B = C = D = \tilde{\delta}_{l,\mathrm{re}}(\boldsymbol{k})$ and inserting this equation into Equation 3.32 we get

$$\mathrm{Var}(\tilde{\delta}_{l,\mathrm{re}}(\boldsymbol{k})^2) = 2\langle\tilde{\delta}_{l,\mathrm{re}}(\boldsymbol{k})^2\rangle^2 = \frac{1}{2}\frac{(2\pi)^6 P(|\boldsymbol{k}|)^2}{l^6} \ . \tag{3.34}$$

And the variance of $\hat{P}(\boldsymbol{k})$ is then given by

$$\begin{aligned}\mathrm{Var}(\hat{P}(\boldsymbol{k})) &= \frac{l^6}{(2\pi)^6}\left(\mathrm{Var}(\tilde{\delta}_{l,\mathrm{re}}(\boldsymbol{k})^2) + \mathrm{Var}(\tilde{\delta}_{l,\mathrm{im}}(\boldsymbol{k})^2)\right) \\ &= P(|\boldsymbol{k}|)^2 \ .\end{aligned} \tag{3.35}$$

Exercise 34

Proof Equations 3.34 and 3.35. (You can assume Wick's theorem as well as the fact that $\delta_{l,\mathrm{re}}$ and $\delta_{l,\mathrm{im}}$ have identical distributions.)

So our estimator $\hat{P}(\boldsymbol{k})$ has a standard deviation of $P(|\boldsymbol{k}|)$, which is the same as its expectation value! This is obviously a very bad estimator, so how can we do better than this? We can e.g. use that fact that the power spectrum only depends on the absolute magnitude of $\boldsymbol{k}$ and average over many different of the above estimators, for different wave vectors $\boldsymbol{k}$ that have similar values of $|\boldsymbol{k}|$. Figure 13 sketches an efficient procedure to do so. Let us choose some range $[k_{\mathrm{min}}, k_{\mathrm{max}}]$ and fill the shell of wave vectors $\boldsymbol{k}$ for which $|\boldsymbol{k}|$ is within this range by as many non-overlapping cubes $V_l(\boldsymbol{k})$ as possible. For each of those cubes we can calculate our previous estimator $\hat{P}(\boldsymbol{k})$ and average over all of them to get the new estimator

$$\hat{P}[k_{\mathrm{min}}, k_{\mathrm{max}}] = \frac{1}{N_{\mathrm{cubes}}}\sum_{|\boldsymbol{k}|\in[k_{\mathrm{min}},k_{\mathrm{max}}]} \hat{P}(\boldsymbol{k}) \ , \tag{3.36}$$

where N_{cubes} is the number of cube we fit into the shell. If we choose the range $[k_{\mathrm{min}}, k_{\mathrm{max}}]$ small enough such that $P(|\boldsymbol{k}|)$ is approximately the same for all of the cubes in the shell then our new estimator will have the expectation value,

$$\langle\hat{P}[k_{\mathrm{min}}, k_{\mathrm{max}}]\rangle \approx P\left(\frac{k_{\mathrm{min}} + k_{\mathrm{max}}}{2}\right) \ . \tag{3.37}$$

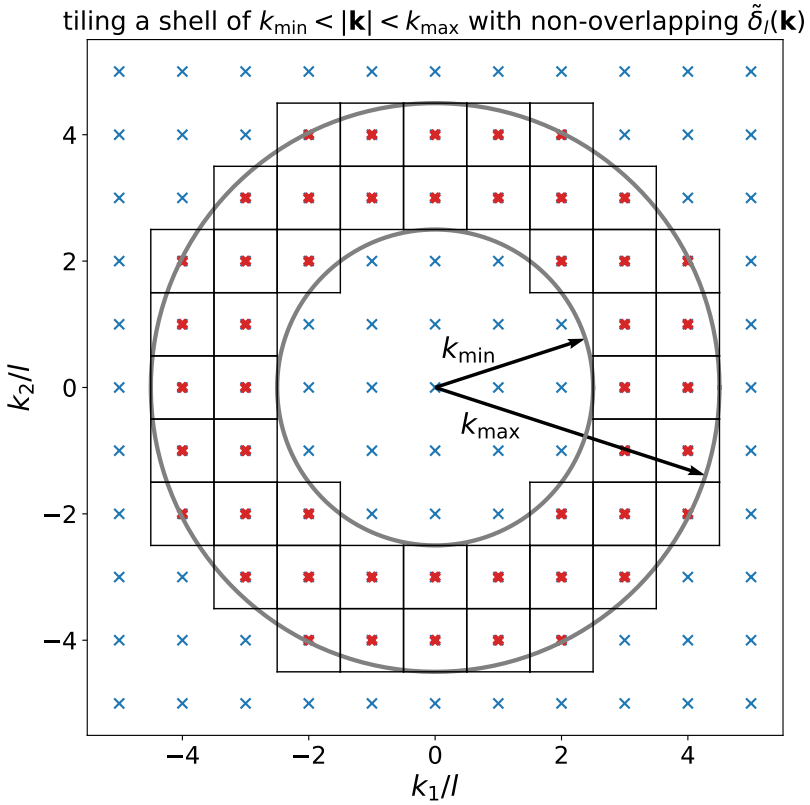

Figure 13. A small hint at the next lecture...

Let us again assume that the density contrast field is Gaussian. Then since all of the cubes in the shell are non-overlapping, the variance of $\hat{P}[k_{\mathrm{min}}, k_{\mathrm{max}}]$ will be $2/N_{\mathrm{cubes}}$ times the variance of the individual $\hat{P}(\boldsymbol{k})$, because about half of the individual $\hat{P}(\boldsymbol{k})$ are independent of each other (it is only half of them, because $\tilde{\delta}_l(-\boldsymbol{k}) = \tilde{\delta}_l(\boldsymbol{k})^*$). This results in

$$\mathrm{Var}(\hat{P}[k_{\mathrm{min}}, k_{\mathrm{max}}]) \approx \frac{2P\left(\frac{k_{\mathrm{min}}+k_{\mathrm{max}}}{2}\right)^2}{N_{\mathrm{cubes}}} \,. \tag{3.38}$$

In particular, the standard deviation of $\hat{P}[k_{\mathrm{min}}, k_{\mathrm{max}}]$ is now smaller than its expectation value by a factor of $\sqrt{2/N_{\mathrm{cubes}}}$ (the above argument based on the fact that $\tilde{\delta}_l(-\boldsymbol{k}) = \tilde{\delta}_l(\boldsymbol{k})^*$ only works if there is at least two cubes). For $N_{\mathrm{cubes}} \gg 1$ this can lead do a high signal-to-noise ratio and hence allow us to obtain a precise measurement of the power spectrum. If we choose the width of our cubes to be much smaller than the range $[k_{\mathrm{min}}, k_{\mathrm{max}}]$ then we can estimate N_{cubes} to be

$$N_{\mathrm{cubes}} \approx \frac{4\pi}{3} \frac{k_{\mathrm{max}}^3 - k_{\mathrm{min}}^3}{l^3} \,. \tag{3.39}$$

Exercise 35

It seems that we can make N_{cubes} arbitrarily large, if we just consider cubes with very small sizes l. So we can measure the power spectrum with arbitrarily high precision. Does this surprise you? Why could this statement be wrong in real observations?

As usual, we will right away discuss some thoughts about Exercise 35. So if you would like to take more time to think about it, you should stop reading here. All of the above derivations assume that we have access to our realisation of the matter density contrast field in the entire, infinitely large space $\mathbb{R}^3$. This would of course constitute an infinite amount of data, thus allowing as infinite precision. Real observations will always only inform us about a finite sub-volume of space. Say that we have e.g. observed the cosmic density field in a cube of finite size L^3. In that case we cannot apply a continuous Fourier transformation to our observed density contrast field $\delta(\boldsymbol{x})$ to obtain the Fourier modes $\tilde{\delta}(\boldsymbol{k})$. Instead, we will have to apply a discrete Fourier decomposition which will result in a discrete grid of wave vectors $\boldsymbol{k}$. This grid can be shown to have the width

$$l_{\min} = \frac{2\pi}{L} \ . \tag{3.40}$$

This constitutes a lower limit for the size l of the cubes we used to define our power spectrum estimator. If we choose l to coincide with $l_{\min}$ then the variance of $\hat{P}[k_{\min}, k_{\max}]$ becomes

$$\mathrm{Var}(\hat{P}[k_{\min}, k_{\max}]) \approx P\left(\frac{k_{\min}+k_{\max}}{2}\right)^2 \frac{3}{L^3(k_{\max}^3 - k_{\min}^3)} \ . \tag{3.41}$$

Within the Gaussian approximation we made, this is the highest precision we can achieve in our finite observed volume. A full calculation of the four-point function of $\tilde{\delta}(\boldsymbol{k})$ would actually give a larger variance, i.e. a smaller overall signal-to-noise ratio. This is because the non-linear terms in the gravitational collapse equations move cosmological information out of the power spectrum and into higher-order moments of the density field. We will investigate one way of recovering this information in Section 4.

In an analysis of observational data one would apply our estimator $\hat{P}[k_{\min}, k_{\max}]$ to a number of non-overlapping ranges $[k_{\min}, k_{\max}]$ which would result in a measurement of the data vector

$$\hat{\boldsymbol{d}} = \begin{pmatrix} \hat{P}[k_{\min,1}, k_{\max,1}] \\ \hat{P}[k_{\min,2}, k_{\max,2}] \\ \cdots \\ \hat{P}[k_{\min,N}, k_{\max,N}] \end{pmatrix} \ . \tag{3.42}$$

Within our Gaussian approximation, the different elements of $\hat{\boldsymbol{d}}$ will be independent of each other. Equation 3.41 gives us the variance of these elements, but what is their overall PDF? To see that, let us re-write Equation 3.36 as

$$\begin{aligned} \hat{P}[k_{\min}, k_{\max}] &= \frac{1}{N_{\mathrm{cubes}}} \frac{l^3}{(2\pi)^3} \sum_{|\boldsymbol{k}| \in [k_{\min}, k_{\max}]} \left(\tilde{\delta}_{l,\mathrm{re}}(\boldsymbol{k})^2 + \tilde{\delta}_{l,\mathrm{im}}(\boldsymbol{k})^2\right) \\ &\approx \frac{1}{N_{\mathrm{cubes}}} \frac{2l^3}{(2\pi)^3} \sum_{\substack{|\boldsymbol{k}| \in [k_{\min}, k_{\max}] \\ k_x > 0}} \left(\tilde{\delta}_{l,\mathrm{re}}(\boldsymbol{k})^2 + \tilde{\delta}_{l,\mathrm{im}}(\boldsymbol{k})^2\right) \ , \end{aligned} \tag{3.43}$$

where in the second line we have used the fact that $\tilde{\delta}_{l,\mathrm{re}}(\boldsymbol{k}) = \tilde{\delta}_{l,\mathrm{re}}(-\boldsymbol{k})$ and $\tilde{\delta}_{l,\mathrm{im}}(\boldsymbol{k}) = -\tilde{\delta}_{l,\mathrm{im}}(-\boldsymbol{k})$. The left hand side of Equation 3.43 is a sum over the squares of N_{cubes} independent random variables, each with the same, Gaussian distribution. If we factor out the

variances of these Gaussian variables as

$$\hat{P}[k_{\rm min}, k_{\rm max}] \approx \frac{1}{N_{\rm cubes}} P\left(\frac{k_{\rm min}+k_{\rm max}}{2}\right) \sum_{\substack{|\boldsymbol{k}|\in[k_{\rm min},k_{\rm max}] \\ k_x>0}} \frac{2l^3}{(2\pi)^3} \frac{\tilde{\delta}_{l,\rm re}(\boldsymbol{k})^2 + \tilde{\delta}_{l,\rm im}(\boldsymbol{k})^2}{P\left(\frac{k_{\rm min}+k_{\rm max}}{2}\right)} \ , \qquad (3.44)$$

the the sum on the left hand side is indeed over the squares of standard Gaussian random variables (i.e. Gaussian random variables with mean 0 and standard deviation 1). So $\hat{P}[k_{\rm min}, k_{\rm max}]$ is just a re-scaled version of a χ^2-distributed random variable with $N_{\rm cubes}$ degrees of freedom! So its PDF is given by

$$p\left(\hat{P}[k_{\rm min}, k_{\rm max}] = x\right) = \frac{N_{\rm cubes}}{P(k_{\rm mean})} \frac{\left(\frac{x\ N_{\rm cubes}}{P(k_{\rm mean})}\right)^{\frac{N_{\rm cubes}}{2}-1} \exp\left(-\frac{1}{2}\left(\frac{x\ N_{\rm cubes}}{P(k_{\rm mean})}\right)\right)}{2^{N_{\rm cubes}/2}\ \Gamma(N_{\rm cubes}/2)} \ , \qquad (3.45)$$

where we have introduced the notation $k_{\rm mean} = (k_{\rm min} + k_{\rm max})/2$ and where Γ is the gamma function.

Of course, for large $N_{\rm cubes}$ this PDF will tends to a Gaussian PDF itself, by virtue of the central limit theorem. In many cosmological applications this is indeed a good approximation, i.e. very often it can be assumed that measurements of the power spectrum have Gaussian errorbars. Equation 3.45 then becomes

$$p\left(\hat{P}[k_{\rm min}, k_{\rm max}] = x\right) \approx \frac{1}{\sqrt{2\pi {\rm Var}(\hat{P}[k_{\rm min}, k_{\rm max}])}} \exp\left(-\frac{1}{2}\frac{(x - P(k_{\rm mean}))^2}{{\rm Var}(\hat{P}[k_{\rm min}, k_{\rm max}])}\right) \ , \qquad (3.46)$$

where ${\rm Var}(\hat{P}[k_{\rm min}, k_{\rm max}])$ is given by Equation 3.41.

Exercise 36

Derive Equation 3.45. You are allowed to assume that the sum on the left hand side of Equation 3.44 is indeed a χ^2-distributed random variable with $N_{\rm cubes}$ degrees of freedom.

Exercise 37

Let's perform a simulated data analysis! Assume that $P(k) = Ak^{0.96}$, with an unknown amplitude A (your model parameter). The following steps will let you perform a mock analysis of a measurement of this power spectrum.

A) From the link https://cloud.physik.lmu.de/index.php/s/yegsefi4GwDiAjM *, download the file* `mock_Pk.txt`*. It contains a measurement of the above power spectrum in 8 different bins $[k_{\rm min}, k_{\rm max}]$, in a survey that has observed the cosmic density field in a finite volume $V = L^3$ with $L = 200$ (no units in this simple example). You find $k_{\rm min}$, $k_{\rm max}$ and $\hat{P}[k_{\rm min}, k_{\rm max}]$ given in the 3 columns of the file. You can load the file using the* `numpy` *command* `loadtxt`*.*

B) Plot this measurement (i.e. P vs. k) using the `matplotlib.pyplot.errorbar` *command in* `python`*. The errorbars should be given by the square roots of Equation 3.41. Hint: if you use the* `matplotlib.pyplot` *commands* `plt.xscale('log')` *and* `plt.yscale('log')` *your plot will look nicer.*

C) With the above information, perform a Bayesian likelihood analysis to calculate a posterior density function for the parameter A. Hint: use Equation 3.18, keeping in mind that

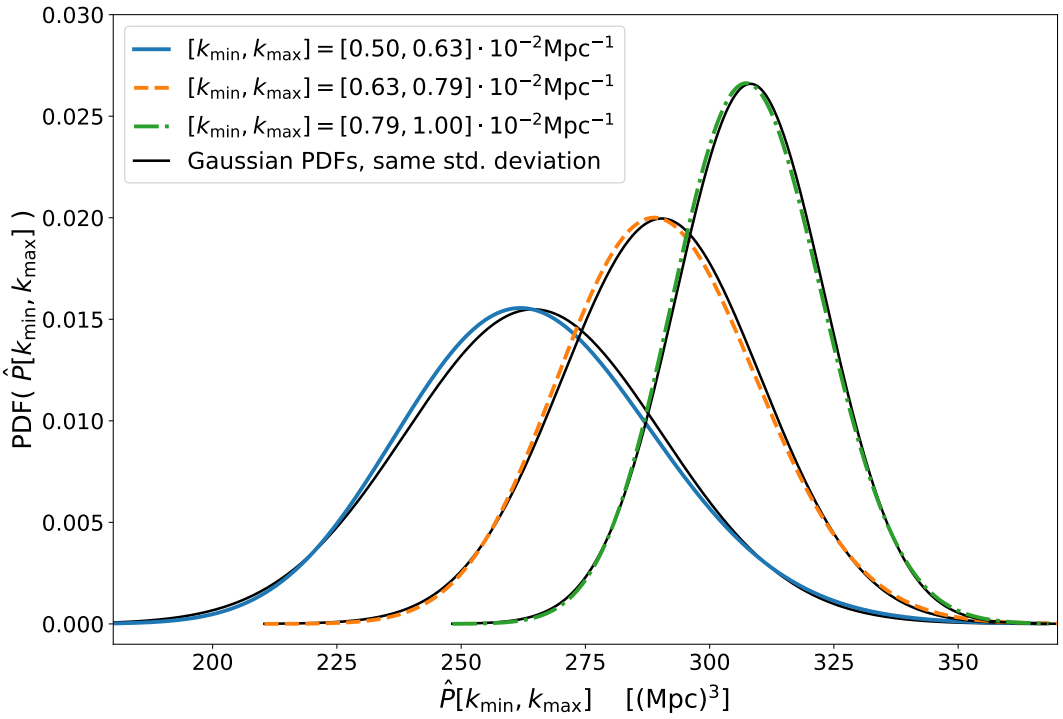

Figure 14. Using Equations 3.45, 3.39 and 3.40 to predict the PDF of the power spectrum estimator from Equation 3.44 for a number of different k-ranges and assuming that the cosmic density field was observed in a finite volume of 100Gpc^3. The thick, colored lines depict the χ^2-PDF of Equation 3.45 while the thin, black lines show a Gaussian approximation to the exact PDFs. The underlying, true power spectrum us assumed to be the lower most line of Figure 9, i.e. the linear power spectrum at a redshift of 20.

for power spectrum estimates the likelihood function $p(\hat{\boldsymbol{d}}|\boldsymbol{\pi})$ *is given by Equation 3.45. Since the ranges* $[k_{\rm min}, k_{\rm max}]$ *in* `mock_Pk.txt` *are not overlapping, you can assume each measurement to be independent, i.e. their joint PDF will be the product of the individual PDFs. You can assume a uniform prior distribution* $p(A)$ *within the range* $A \in [0.1, 10.0]$.

4 Predicting the PDF of matter density fluctuations in a sphere

We have seen in Section 2 that the probability distribution of a Gaussian random field is fully described by the field's power spectrum. At the same time, we have discussed that the late time cosmic density contrast is not a Gaussian random field (cf. Exercise 21), so the power spectrum may not be a complete description of its statistical properties. Figure 15 demonstrates that this is indeed the case. Both of the simulated density fields shown there have exactly the same power spectrum. But only one of the two is a physical density field, in the sense that is has evolved from Gaussian initial conditions according to the continuity and the Euler equation. Can you spot which one (without looking at Figure 3)? Or can you at least tell, that there is a difference in the statistical behaviour of the fluctuations in the two fields?

Exercise 38

What differences do you notice between the two density fields shown in Figure 15? Try to find at least two.

If you managed to answer Exercise 38 then you are already more effective in telling apart physical from unphysical density fields than our power spectrum estimator from Equation 3.36. The upper panel of Figure 16 shows measurements of that estimator in both density contrast fields of Figure 15, and they are indeed exactly identical. Recall our discussion from Exercise 7 about how observations of the cosmic density field can be used to test our understanding of gravitational collapse. The fact that the power spectrum does not fully characterise the density field has at least two implications for such a program:

- investigating the power spectrum of cosmic density fluctuations and its evolution with time, may not be sufficient to detect differences between cosmic structure formation and our current theoretical understanding thereof;
- and even if our cosmological standard model is correct - the power spectrum is only an incomplete piece of information about the cosmic density field. In particular, there may be other characteristics of the density field that would allow us to determine the parameters of the standard model to higher precision than with measurements of the power spectrum alone.

So what information about the density contrast field are we missing by looking at the power spectrum alone? To see that, let us work out a slightly different view on the power spectrum's information content. Let us consider a smoothed version of the density contrast $\delta(\boldsymbol{x})$ where we average the density fluctuations within a radius R around each location $\boldsymbol{x}$, i.e.

$$\delta_R(\boldsymbol{x}) \equiv \frac{3}{4\pi R^3} \int_{|\boldsymbol{x}'-\boldsymbol{x}|<R} \mathrm{d}^3x' \; \delta(\boldsymbol{x}) \; . \tag{4.1}$$

Let us calculate the variance $\langle \delta_R(\boldsymbol{x})^2 \rangle$ of this smoothed field. Since the cosmic density field is a homogeneous random field, this variance is independent of $\boldsymbol{x}$. One can show that it is given by

$$\langle \delta_R^2 \rangle = \frac{1}{2\pi^2} \int \mathrm{d}\ln k \; k^3 \; P(k) \; \tilde{W}_R(k)^2 \; , \tag{4.2}$$

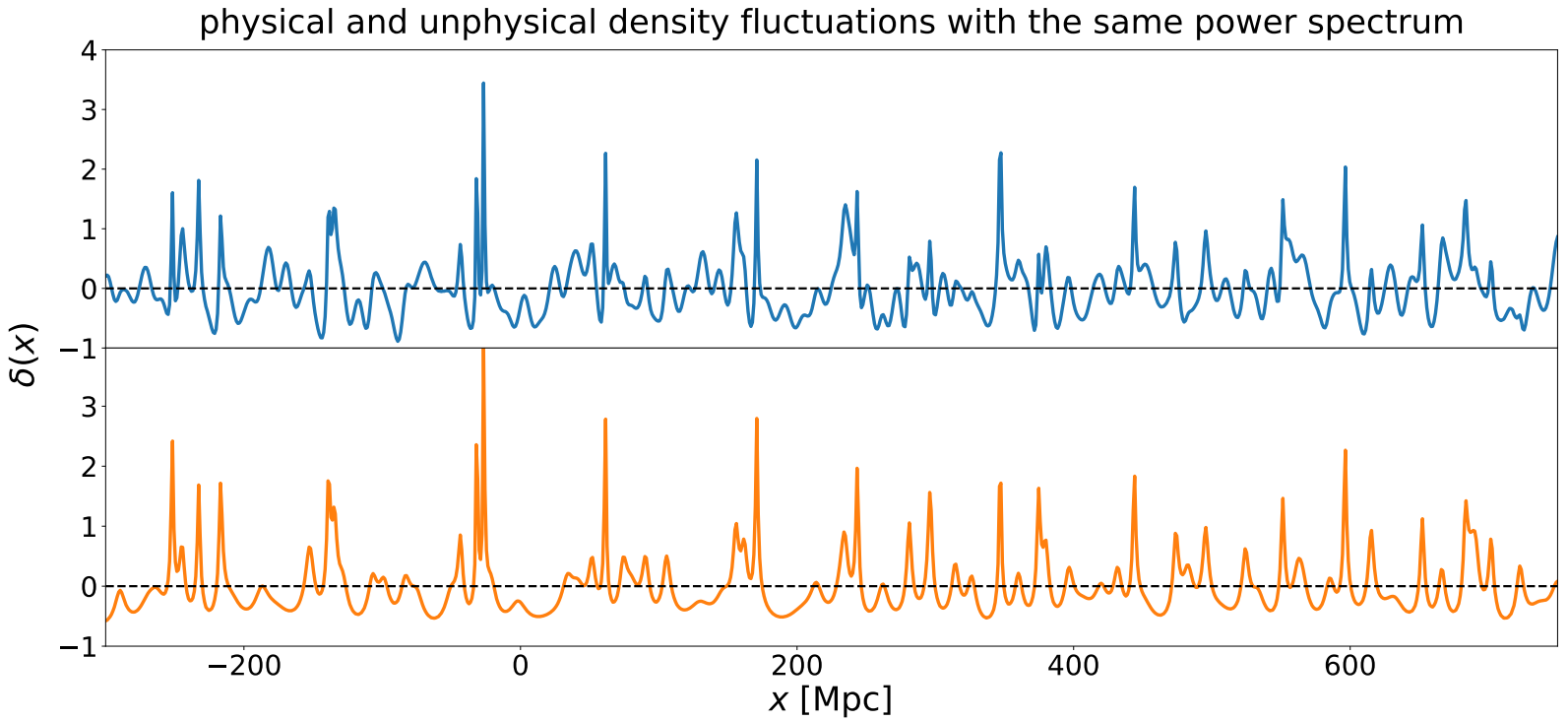

Figure 15. Two simulated density fields. One has been obtain by generating Gaussian initial fluctuations and evolving them according to the Euler and the continuity equation. The other has been obtain via an unphysical mechanism. Can you spot the physical density field?

where $\tilde{W}_R(k)$ is the Fourier transform of the top-hat widow function by which we have smoothed the field. It is given by

$$\tilde{W}_R(k) = 3\left(\frac{\sin(kR)}{(kR)^3} - \frac{\cos(kR)}{(kR)^2}\right) \ . \tag{4.3}$$

Exercise 39

Derive Equation 4.2.

For different values of the smoothing radius R the window function $\tilde{W}_R(k)$ will peak at a different value of k such that the integral in Equation 4.2 will be sensitive to a different part of the power spectrum. So the power spectrum determines how the variance of δ_R depends on the smoothing scale R. Hence, the information that is missing from the power spectrum is anything that goes beyond the variance of density fluctuations. It does e.g. know nothing about higher order moments of the density field such as $\langle\delta_R^3\rangle$.

Of course, at any given location $\boldsymbol{x}$ the smoothed density contrast $\delta_R(\boldsymbol{x})$ is a random variable. It will hence have a PDF $p(\delta_R)$, which because of homogeneity will not depend on $\boldsymbol{x}$ and which will contain information about all the moments $\langle\delta_R^n\rangle$. A measurement of this PDF may hence be able to distinguish between the two density fields of Figure 15. If we have observed $\delta_R(\boldsymbol{x})$ at many different locations, then we can obtain such a measurement by simply making a histogram of all the observed values $\delta_R(\boldsymbol{x})$. The histograms that result from the 1D fields of Figure 15 are shown in the lower panel Figure 16 - with the smoothing scale now simply being the resolution scale of our 1D simulation. The 1-point PDFs of the two fields clearly seems to be different, with the PDF of the physical density field (which is in fact the one in the lower panel of Figure 15) falling off much sharper in its low density tail, while also having a more extended high density tail. If we had a theoretical model for this PDF, we would have indeed been able figure out that the unphysical field does not conform to our paradigm of structure formation.

The is a rich amount of work on the theoretical modelling of the 1-point PDF of the cosmic density field - see e.g. [3–8, 10, 11, 14, 15, 18, 19, 26, 27, 29, 36, 49–55] for a selection. I will not attempt to present the insights from these papers in full detail here. Instead, I would like to explain to you the starting point of one of these modelling strategies (the one employed by [27, 29, 36, 53–55]) and how this relates back to the probability density functional of the initial density field we had explored in Section 2.

4.1 Laplace's method

As a preparation for some of the upcoming expressions, let us have a look at a technique for approximating integrals which is called *Laplace's method* and which has a wide range of applications. Consider an integral of the form

$$I = \int \mathrm{d}x \ e^{-f(x)} \ , \tag{4.4}$$

where we assume that the function f has a unique, global minimum x^* such that $f''(x^*) > 0$ (especially, it is not 0). We can then approximate f by its second order Taylor expansion

$$f(x) \approx f(x^*) + \frac{f''(x^*)}{2}\,(x - x^*)^2 \ , \tag{4.5}$$

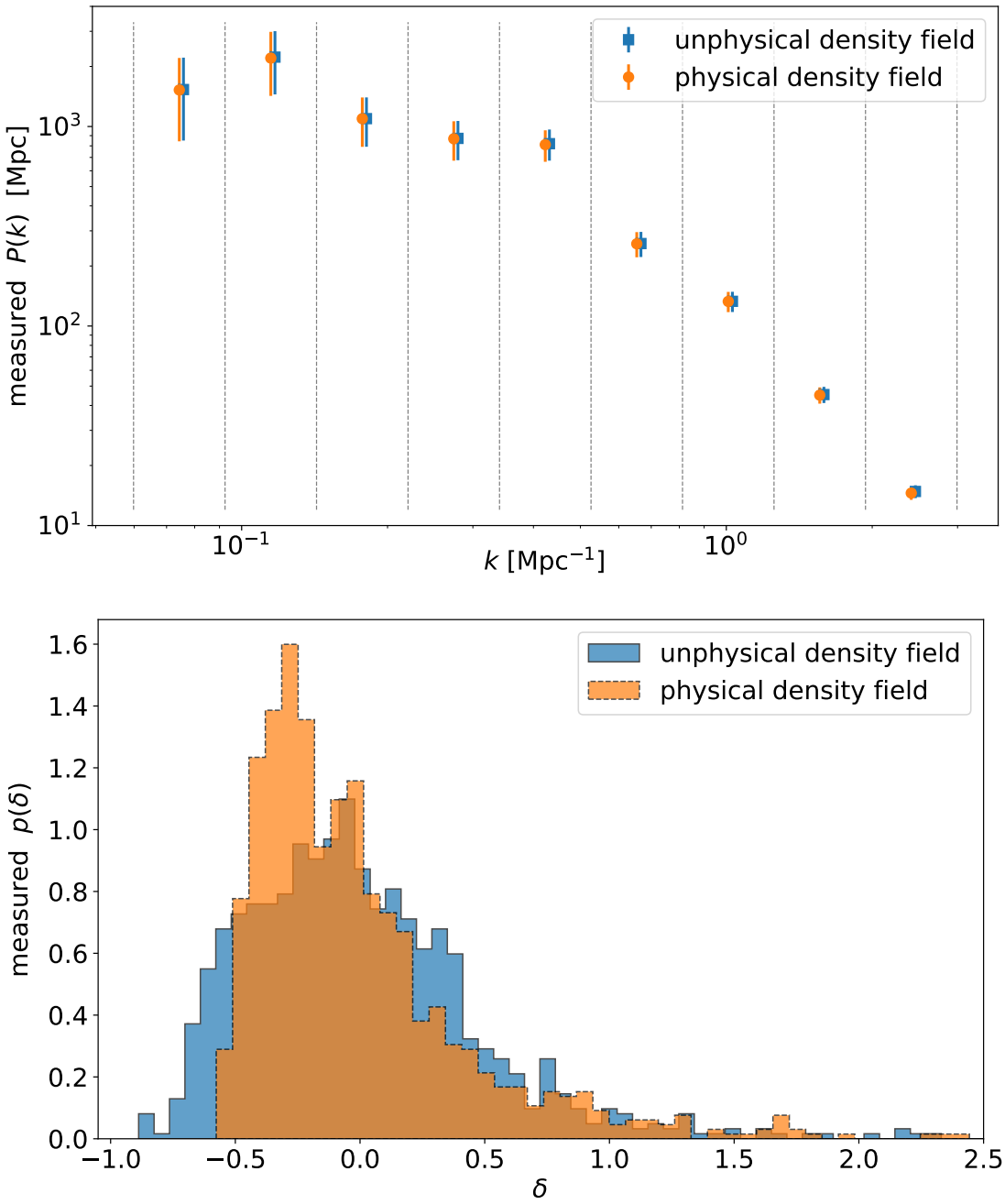

Figure 16. Upper panel: measurements of the power spectrum in both the physical and unphysical density field of Figure 15. Both fields turn out to have exactly the same power spectrum. Lower panel: measurements of the full 1-point PDF in both density fields. This PDF is indeed able to tell the two fields apart.

where the linear term of the expansion vanishes because x^* is a minimum. If we insert this approximation into our definition of the integral I we get

$$I \approx \exp\left(-f(x^*)\right) \int \mathrm{d}x \ \exp\left(-\frac{f''(x^*)}{2}\ (x-x^*)^2\right)$$
$$= \sqrt{\frac{2\pi}{f''(x^*)}}\ \exp\left(-f(x^*)\right)\ . \tag{4.6}$$

This approximation is usually used in contexts where the function f depends on a parameter m such that $f(x|m) = g(x)/m$. It can then be shown that Laplace's method becomes more and more accurate in the limit $m \to 0$.

Laplace's method can be extended to higher dimensions. Let the function $f = f(\boldsymbol{x})$ again have a global minimum $\boldsymbol{x}^*$ such that the Hesse matrix at that minimum

$$H_{f,ij}(\boldsymbol{x}^*) = \left.\frac{\partial^2 f}{\partial x_i \partial x_j}\right|_{\boldsymbol{x}^*} \tag{4.7}$$

is positive definite. We can then approximate f as

$$f(\boldsymbol{x}) \approx f(\boldsymbol{x}^*) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}^*)^T \boldsymbol{H}_f(\boldsymbol{x}^*)(\boldsymbol{x} - \boldsymbol{x}^*) \tag{4.8}$$

which leads to

$$\begin{aligned} I &= \int \mathrm{d}^n x \ \exp(-f(\boldsymbol{x})) \\ &\approx \exp(-f(\boldsymbol{x}^*)) \int \mathrm{d}^n x \ \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}^*)^T \boldsymbol{H}_f(\boldsymbol{x}^*)(\boldsymbol{x} - \boldsymbol{x}^*)\right) \\ &= \sqrt{\frac{(2\pi)^n}{\det(\boldsymbol{H}_f(\boldsymbol{x}^*))}} \exp(-f(\boldsymbol{x}^*)) \ . \end{aligned} \tag{4.9}$$

Again, this approximation is usually applicable in the limit $m \to 0$ when f depends on a parameter m as $f(\boldsymbol{x}|m) = g(\boldsymbol{x})/m$.

4.2 Path integral approach

In order to calculate the 1-point PDF of δ_R for a given smoothing radius R, let us recall that the PDF $p(\delta_R)$ is related to the cumulant generating function (CGF) through a Laplace transform as (cf. Equations 2.13 as well as 2.16 and 2.17)

$$\exp\left(\varphi_R(\lambda)\right) = \int \mathrm{d}\delta_R \ e^{\lambda \delta_R} \ p(\delta_R) \ . \tag{4.10}$$

So if we had a prediction for $\varphi_R(\lambda)$, then we could obtain from it a prediction for the PDF via the inverse Laplace transform

$$p(\delta_R) = \int \frac{\mathrm{d}\lambda}{2\pi} \ \exp\left(i\lambda\delta_R + \varphi_R(i\lambda)\right) \ . \tag{4.11}$$

At the same time, the CGF is given by the expecation value (cf. Equation 2.11)

$$\exp\left(\varphi_R(\lambda)\right) = \langle e^{\lambda \delta_R} \rangle \ . \tag{4.12}$$

We don't have to calculate this expectation value via Equation 4.11. We can also obtain it by averaging over all possible configurations of the initial density contrast field $\delta_i(\boldsymbol{x})$. This would result in the following functional integral:

$$\exp\left(\varphi_R(\lambda)\right) = \int \mathcal{D}\delta_i \ \mathcal{P}[\delta_i] \ \exp(\lambda \delta_R[\delta_i]) \ . \tag{4.13}$$

Here $\mathcal{P}[\delta_i]$ is the probability density functional of δ_i and the smoothed density contrast δ_R has been expressed as a functional of δ_i as well. Note that because of homogeneity it is again irrelevant at which location $\boldsymbol{x}$ we consider δ_R. Since the initial density fluctuations are Gaussian, their PDF was given by

$$\mathcal{P}[\delta_i] = \frac{1}{|2\pi C|^{1/2}} \exp\left\{-\frac{1}{2}\int \mathrm{d}^3x_1 \int \mathrm{d}^3x_2 \; \delta_i(\boldsymbol{x}_1) C^{-1}(\boldsymbol{x}_1, \boldsymbol{x}_2) \delta_i(\boldsymbol{x}_2)\right\} , \tag{4.14}$$

where $C^{-1}(\boldsymbol{x}_1, \boldsymbol{x}_2)$ was the (symbolic) inverse of the two-point function $C(\boldsymbol{x}_1, \boldsymbol{x}_2)$, and the latter can be interpreted as the "covariance matrix" of the field δ_i . It is somewhat difficult to deal with the inverse 2-point function, but we can work around it if we express $\mathcal{P}[\delta_i]$ via the cumulant generating functional of δ_i . In Exercise 16 you have shown that the CGF of a Gaussian random variable with mean μ and variance σ^2 is given by $\varphi_{\mathrm{Gauss}}(\lambda) = \mu\lambda + \frac{\sigma^2}{2}\lambda^2$. This is easily generalised to multivariate Gaussian random vectors $\boldsymbol{X}$ with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{C}$. They have the CGF

$$\begin{aligned}\varphi_{\mathrm{Gauss}}(\boldsymbol{\lambda}) &= \boldsymbol{\lambda}^T\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{\lambda}^T\boldsymbol{C}\boldsymbol{\lambda} \\ &= \sum_{i=1}^{n}\lambda_i\mu_i + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i C_{ij}\lambda_j \; .\end{aligned} \tag{4.15}$$

In complete analogy to this, the CGFunctional of the Gaussian random field δ_i is given by

$$\begin{aligned}\Phi[J] = &\int \mathrm{d}x \; J(\boldsymbol{x})\langle\delta_i(\boldsymbol{x})\rangle \\ &+ \frac{1}{2}\int \mathrm{d}^3x_1\mathrm{d}^3x_2 \; J(\boldsymbol{x}_1) \; \text{Covariance}(\delta_i(\boldsymbol{x}_1), \delta_i(\boldsymbol{x}_2)) \; J(\boldsymbol{x}_2) \\ = &\frac{1}{2}\int \mathrm{d}^3x_1\mathrm{d}^3x_2 \; J(\boldsymbol{x}_1) \; C(\boldsymbol{x}_1, \boldsymbol{x}_2) \; J(\boldsymbol{x}_2) \; .\end{aligned} \tag{4.16}$$

Now how can we use the CGFunctional of δ_i to express its PDFunctional? First note that in a straight forward generalisation of Equation 4.11 the PDF of an n-dimensional random vector $\boldsymbol{V}$ is given in terms of the CGF of $\boldsymbol{V}$ via

$$\begin{aligned}p(\boldsymbol{v}) = &\int \frac{\mathrm{d}^n\lambda}{(2\pi)^n} \exp\left(i\sum_{j=1}^{n}\lambda_j v_j + \varphi(i\boldsymbol{\lambda})\right) \\ = &\frac{1}{|2\pi\mathbf{1}_n|}\int \mathrm{d}^n\lambda \; \exp\left(i\boldsymbol{\lambda}^T\boldsymbol{v} + \varphi(i\boldsymbol{\lambda})\right) \; ,\end{aligned} \tag{4.17}$$

where in the last line we have expressed the factor $(2\pi)^n$ as the determinant of 2π times the n-dimensional unit matrix. In complete analogy to Equation 4.17, the PDFunctional of δ_i is given in terms of its CGFunctional by the inverse Laplace transform

$$\mathcal{P}[\delta_i] = \frac{1}{|2\pi\delta_D^{(3)}|}\int \mathcal{D}J \; \exp\left(i\int \mathrm{d}^3x \; J(\boldsymbol{x})\delta_i(\boldsymbol{x}) - \frac{1}{2}\int \mathrm{d}^3x_1\mathrm{d}^3x_2 \; J(\boldsymbol{x}_1) \; C(\boldsymbol{x}_1, \boldsymbol{x}_2) \; J(\boldsymbol{x}_2)\right) \; , \tag{4.18}$$

where $|2\pi\delta_D^{(3)}|$ is the functional determinant of 2π times the 3-dimensional delta function. So in summary, the CGF of δ_R is given by

$$\begin{aligned}
&\exp\left(\varphi_R(\lambda)\right) = \\
&\frac{1}{|2\pi\delta_D^{(3)}|}\int \mathcal{D}\delta_i\mathcal{D}J\ \exp\left(\lambda\delta_R[\delta_i] + i\int \mathrm{d}^3x\ J(\boldsymbol{x})\delta_i(\boldsymbol{x}) - \frac{1}{2}\int \mathrm{d}^3x_1\mathrm{d}^3x_2\ J(\boldsymbol{x}_1)\ C(\boldsymbol{x}_1,\boldsymbol{x}_2)\ J(\boldsymbol{x}_2)\right) \\
&\equiv \frac{1}{|2\pi\delta_D^{(3)}|}\int \mathcal{D}\delta_i\mathcal{D}J\ \exp\left(-S_\lambda[\delta_i, J]\right)\ , \qquad (4.19)
\end{aligned}$$

where in the last line we have defined the action

$$S_\lambda[\delta_i, J] = -\lambda\delta_R[\delta_i] - i\int \mathrm{d}^3x\ J(\boldsymbol{x})\delta_i(\boldsymbol{x}) + \frac{1}{2}\int \mathrm{d}^3x_1\mathrm{d}^3x_2\ J(\boldsymbol{x}_1)\ C(\boldsymbol{x}_1,\boldsymbol{x}_2)\ J(\boldsymbol{x}_2)\ . \qquad (4.20)$$

Equation 4.19 looks significantly more complicated than Equation 4.10! So what have we gained with this reformulation? The problem with Equation 4.10 is of course that we cannot evaluate it, since we do not know $p(\delta_R)$ - calculating this PDF is the very purpose of our derivations. On the other hand, we know everything on the right-hand side of Equation 4.19 at least in principle. The most troublesome ingredient in our definition of the action S_λ is the functional $\delta_R[\delta_i]$, which amounts to taking the initial density contrast field $\delta_i(\boldsymbol{x})$, evolving it to the time of interest via the Euler and the continuity equation, and then averaging the resulting density contrast field within a spherical aperture according to Equation 4.1. Let us assume that we can indeed evaluate this functional. Then we could attempt to solve the integral in the last line of Equation 4.19 via Laplace's method, which we had looked at in Section 4.1. This would require us to find the configurations $\delta_i^*(\boldsymbol{x})$ and $J^*(\boldsymbol{x})$ of the fields δ_i and J that minimize the action $S_\lambda[\delta_i, J]$. With these we could then approximate $\exp(\varphi_R(\lambda))$ as

$$\exp\left(\varphi_R(\lambda)\right) \approx \frac{1}{|2\pi\delta_D^{(3)}|}\frac{|2\pi\delta_D^{(3)}|}{|H_{S_\lambda}[\delta_i^*, J^*]|^{1/2}}\exp\left(-S_\lambda[\delta_i^*, J^*]\right)\ . \qquad (4.21)$$

Here $H_{S_\lambda}[\delta_i^*, J^*]$ is the Hessian of the functional S_λ, evaluated at the minimizing configurations δ_i^* and J^* . The prefactor $1/|2\pi\delta_D^{(3)}|$ in Equation 4.21 carries over from Equation 4.19, i.e. it has nothing to do with Laplace's method. The additional factor of $|2\pi\delta_D^{(3)}|$ is however part of Laplace's approximation - it corresponds to the factor

$$\sqrt{(2\pi)^n} = |2\pi\mathbf{1}_n|^{1/2} \qquad (4.22)$$

which appears in the last line of Equation 4.9, and which can again be expressed as the square root of the determinant of 2π times the n-dimensional unit matrix. But why is there no square root in Equation 4.21? Note that S_λ is a functional of two fields - δ_i and J. In a sense, this doubles the dimensionality of all the operators ("matrices") that appear in our functional version of Laplace's method, and the factor $|2\pi\delta_D^{(3)}|$ is in fact

$$|2\pi\delta_D^{(3)}| = \left|2\pi\begin{pmatrix}\delta_D^{(3)} & \mathbf{0}\\ \mathbf{0} & \delta_D^{(3)}\end{pmatrix}\right|^{1/2}\ . \qquad (4.23)$$

The CGF of the smoothed density contrast δ_R is then approximated as

$$\varphi_R(\lambda) \approx -S_\lambda[\delta_i^*, J^*] - \frac{1}{2}\ln\left(|H_{S_\lambda}[\delta_i^*, J^*]|\right)\ . \qquad (4.24)$$

The above derivations should show you that there is indeed nothing magical or overly complicated about the functional integrals and expressions of e.g. Equations 4.18, 4.19 or 4.20. They can be understood in complete analogy to more the familiar situation of n-dimensional integrals and random variables.

Another analogous operation we need to define is the functional derivative of a functional $F[f]$ wrt. its argument f at location $\boldsymbol{x}$. We will denote it as $\mathit{d}F/\mathit{d}f(\boldsymbol{x})$, and its analog for n-dimensional functions $f(\boldsymbol{x})$ would be $\partial f/\partial x_i$ - i.e. in the functional case the location $\boldsymbol{x}$ again plays the role of a label enumerating the different "entries" $f(\boldsymbol{x})$ in the "vector" f, just as the index i labels different entries x_i of the vector $\boldsymbol{x}$ in the standard case. Characteristic for the action of the functional derivative is the fact that

$$\frac{\mathit{d}f(\boldsymbol{x}_1)}{\mathit{d}f(\boldsymbol{x}_2)} = \delta_D(\boldsymbol{x}_2 - \boldsymbol{x}_1) \; . \tag{4.25}$$

This is completely analogous to the standard expression $\partial x_i/\partial x_j = \delta_{ij}$.

Exercise 40

Define the functionals F_1 and F_2 such that

$$F_1[f] = \int \mathrm{d}^3x' \; f(\boldsymbol{x}') \; e^{-|\boldsymbol{x}'|^2} \quad , \quad F_2[f] = \int \mathrm{d}^3x' \; f(\boldsymbol{x}')^2 \; .$$

Calculate the first order functional derivatives $\mathit{d}F_i/\mathit{d}f(\boldsymbol{x})$ and the second order functional derivatives $\mathit{d}^2F_i/(\mathit{d}f(\boldsymbol{x}_1)\mathit{d}f(\boldsymbol{x}_2))$.

To find the configurations δ_i^* and J^* that minimize the action S_λ, we have to solve the equations

$$\left.\frac{\mathit{d}S_\lambda}{\mathit{d}\delta_i(\boldsymbol{x})}\right|_{\delta_i^*, J^*} = 0 \tag{4.26}$$

$$\left.\frac{\mathit{d}S_\lambda}{\mathit{d}J(\boldsymbol{x})}\right|_{\delta_i^*, J^*} = 0 \; . \tag{4.27}$$

Taking into account the definition of S_λ this amounts to solving

$$\delta_i^*(\boldsymbol{x}) = \; - \int \mathrm{d}^3x' \; iJ^*(\boldsymbol{x}') \; C(\boldsymbol{x}', \boldsymbol{x}) \tag{4.28}$$

$$iJ^*(\boldsymbol{x}) = \; - \lambda \left.\frac{\mathit{d}\delta_R}{\mathit{d}\delta_i(\boldsymbol{x})}\right|_{\delta_i^*} \; . \tag{4.29}$$

Exercise 41

Derive Equations 4.28 and 4.29 from the definition of S_λ (Equation 4.20).

We will not attempt to solve Equations 4.28 and 4.29. Instead, I will refer you to [54] who has derived explicit solutions, and to [29] who have extended the above formalism to general non-Gaussian initial conditions. In the notation of [29] δ_i^* and J^* take the form

$$iJ^*(\boldsymbol{x}) = A_\lambda \cdot \begin{cases} 1 & \text{if } |\boldsymbol{x}| < R_i \\ 0 & \text{else} \end{cases} \tag{4.30}$$

$$\delta_i^*(\boldsymbol{x}) = A_\lambda \; \langle \delta_{i,R_i} \delta_{i,|\boldsymbol{x}|} \rangle \; . \tag{4.31}$$

Here A_λ is a prefactor that depends on the argument λ of the CGF, and which [29] calculate by means of an implicit equation (cf. their equation 38). Furthermore, R_i represents the *initial radius* of a spherically symmetric density perturbation which today has a density contrast of $\delta_R[\delta_i^*]$ within a radius of R. This can again only be calculated by means of an implicit equation. And finally, δ_{i,R_i} and $\delta_{i,|\boldsymbol{x}|}$ are spherical averages of the initial density contrast over radii of R_i and $|\boldsymbol{x}|$ respectively.

I would like you to take away two main conclusions from the above calculations. Firstly, complicated functional expressions like the one in Equation 4.19 can be broken down into easily understandable analogies with results from n-dimensional analysis, and it is indeed possible to derive concrete and calculable results from them. This is particularly surprising in the light of our discussion in Exercise 22, were we saw that the PDFunctional of δ_i is potentially ill defined. And secondly, you should note that the solutions in Equations 4.30 and 4.31 are spherically symmetric, which is a consequence of the fact that the random field δ_i is statistically isotropic and that the functional $\delta_R[\delta_i]$ is invariant under rotations. This rotational invariance can be used to significantly simplify Equations 4.28 and 4.29 and this is what allowed [54] and [29] to derive their solutions in the first place.

For completeness sake, let us also talk about the Hessian matrix appearing on the right-hand side of Equation 4.24. Since the functional S_λ depends on both fields δ_i and J the Hessian splits into four blocks as

$$
\begin{aligned}
H_{S_\lambda}[\delta_i, J] &= \begin{pmatrix} \frac{\mathrm{d}^2 S_\lambda}{\mathrm{d}\delta_i \mathrm{d}\delta_i} & \frac{\mathrm{d}^2 S_\lambda}{\mathrm{d}\delta_i \mathrm{d}J} \\ \\ \frac{\mathrm{d}^2 S_\lambda}{\mathrm{d}J \mathrm{d}\delta_i} & \frac{\mathrm{d}^2 S_\lambda}{\mathrm{d}J \mathrm{d}J} \end{pmatrix} \\
&= \begin{pmatrix} -\lambda \frac{\mathrm{d}^2 \delta_R}{\mathrm{d}\delta_i \mathrm{d}\delta_i} & -i\delta_D^{(3)} \\ \\ -i\delta_D^{(3)} & C \end{pmatrix} .
\end{aligned} \tag{4.32}
$$

You can think of each of these blocks as being a function of two positions $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$. E.g. the lower right block is given by the 2-point function $C(\boldsymbol{x}_1, \boldsymbol{x}_2)$, while the off-diagonal blocks are $-i\delta_D^{(3)}(\boldsymbol{x}_2 - \boldsymbol{x}_1)$. To fully evaluate Equation 4.24 we would have to compute the determinant of H_{S_λ}, evaluated at the minimizing configurations δ_i^* and J^*. This is a very involved calculation, which has been detailed by [36] with about 90 pages of derivations. As we will discuss in the remaining part of this section, real data analyses have so far circumvented this effort by means of a simple computational trick.

4.3 Comments on real-life applications

Let us touch on a three complications that arise when applying the above theoretical calculations to real observational data.

Firstly, while the functional determinant on the right-hand side of Equation 4.24 can indeed be calculated (e.g. along the lines of [36]), this calculation is very computationally expensive. This is a problem because the parameter space of typical cosmological analyses is very high-dimensional, and in order to sufficiently resolve a posterior distribution on such high-dimensional spaces one typically needs Millions of model evaluations. Thus, even if our PDF calculation took only, say, a minute, it would become a bottle neck for any realistic data analysis.

There is a workaround for this problem, which consists in considering the so-called *reduced cumulant generating function* $\tilde{\varphi}_R$ as opposed to the regular CGF φ_R of δ_R. The reduced CGF is given in terms of the CGF and in terms of the variance $\sigma_R^2 = \langle \delta_R^2 \rangle$ of δ_R as

$$\begin{aligned} \tilde{\varphi}_R(\lambda) &= \sigma_R^2 \, \varphi_R\left(\frac{\lambda}{\sigma_R^2}\right) \\ &= \sum_{n=1}^{\infty} \frac{\langle \delta_R^n \rangle_c}{\sigma_R^{2(n-1)}} \, \frac{\lambda^n}{n!} \\ &\equiv \sum_{n=1}^{\infty} S_n \, \frac{\lambda^n}{n!} \, , \end{aligned} \tag{4.33}$$

where the last line serves as a definition of the so called *reduced cumulants*.

Exercise 42

Show that the second line of Equation 4.33 follows from the first.

From the reduced CGF the PDF can be calculated as

$$p(\delta_R) = \int \frac{\mathrm{d}\lambda}{2\pi\sigma_R^2} \, \exp\left(\frac{i\lambda\delta_R + \tilde{\varphi}_R(i\lambda)}{\sigma_R^2}\right) \, . \tag{4.34}$$

Exercise 43

Show that Equation 4.34 follows from Equation 4.11 and Equation 4.33.

What is the benefit of working with $\tilde{\varphi}$ instead of φ? It can be shown that $\tilde{\varphi}$ only has a very weak time dependence. This allows us to calculate it at early times t in the evolution of the universe, where functional determinant term on the right-hand side of Equation 4.24 can be shown to be negligible. We can then use this early-time reduced CGF as a good approximation to the reduced CGF at later times. Then all that is left in order to evaluate Equation 4.34 for the PDF is to compute the late-time variance σ_R^2. If we have a good model for the late-time power spectrum, then this can be done via Equation 4.2.

A second complication in real data analysis is the fact that our primary observable of the cosmic large-scale structure is the galaxy density field, and not the matter density field. So in many situations, we would actually like to compute the PDF of the galaxy density contrast $\delta_{g,R}$ as opposed to the matter density contrast δ_R. We can think of those two as a pair random variables. According to Bayes' theorem, their joint PDF can be expressed as

$$p(\delta_{g,R}, \delta_R) = p(\delta_{g,R}|\delta_R) p(\delta_R) \, , \tag{4.35}$$

where $p(\delta_R)$ can be calculated along the lines explained above, and $p(\delta_{g,R}|\delta_R)$ is the conditional PDF for finding a galaxy density contrast $\delta_{g,R}$ inside an aperture of radius R, if the matter density contrast in that aperture is δ_R . In this course we will not look into how this conditional PDF can be modelled, but it is an important ingredient in real data analyses.

The final real-life complication we will touch on is the fact that real observations do not tell us about the density contrast field $\delta(\boldsymbol{x})$ at one individual time t. Instead, all cosmological observations are done along our backward light cone and the cosmic time at which we can see $\delta(\boldsymbol{x})$ depends on how far the location $\boldsymbol{x}$ is away from us. In addition to that, we often only have limited information about how far away objects like galaxies are from us. So in many cases we only observe a version of the density field that is smeared out along the lines-of-sight

in our light cone. We won't look into this problem in any detail, but it is something that needs to be taken into account in theoretical predictions for cosmological observables.

Having mentioned the above caveats, let us now turn to the 1-point PDF in actual observations. In Figure 17 it is sketched how one would measure the PDF of $\delta_{g,R}$ in observed density fields like the one resulting from the Dark Energy Survey (DES) we talked about in Section 1. In a sense, this is just a graphical representation of Equation 4.1, but applied to the galaxy density field. Figure 18 shows a PDF measurement obtain from real DES data, obtain along those lines (blue points) [30], and compares it to a best-fitting theoretical prediction, obtained according to our calculations in Section 4.2 [26]. You can see, that the agreement is excellent. You may of course suspect, that this is just because our model has so many free parameters. But parameters required to model the conditional PDF $p(\delta_{g,R}|\delta_R)$ in this prediction have indeed been fixed by means of independent gravitational lensing observations. And the cosmological parameters that went into fitting the remaining part of the prediction are in fact consistent with power spectrum measurements in the same data set. Figure 18 hence provides a powerful confirmation of the cosmological standard model's paradigm of structure formation (i.e. starting from Gaussian initial conditions and then evolving the density field according to the continuity equation and the Euler equation).

5 How likely is my Universe?

In Section 3 we have talked about the power spectrum $P(|\boldsymbol{k}|)$ of the cosmic density fluctuations and in Section 4 we have considered the 1-point PDF $p(\delta_R)$ of smoothed density fluctuations δ_R. Both of these quantities can be estimated from observations of the cosmic density contrast field $\delta(\boldsymbol{x})$ in a finite cosmic volume - the power spectrum by employing Equation 3.36 for a number of different ranges $[k_{\rm min}, k_{\rm max}]$ and the PDF by smoothing $\delta(\boldsymbol{x})$ according to Equation 4.1 and then measuring a histogram of the resulting field δ_R using a number of different bins $[\delta_{R,\rm min}, \delta_{R,\rm max}]$. We could arrange all of these measurements into one data vector $\hat{\boldsymbol{d}}$, and this data vector would constitute a quite powerful characterisation of the statistical properties of the cosmic density field - or in other words: of the PDFunctional $\mathcal{P}[\delta]$.

But this vector $\hat{\boldsymbol{d}}$ would certainly not be an exhaustive characterisation of $\delta(\boldsymbol{x})$. Many other statistics of cosmic density fluctuations have been considered in the cosmological literature: higher order correlation function such as the 3-point function $\langle\delta(\boldsymbol{x}_1)\delta(\boldsymbol{x}_2)\delta(\boldsymbol{x}_3)\rangle$; the abundance of density peaks as a function of the peak height; the abundance and profile of cosmic voids; characteristics of the topology of the cosmic web and many more. **(OF: References!)** In principle we could measure all of these and put them into one giant data vector $\hat{\boldsymbol{d}}$ in an attempt to obtain a complete amount of information about the cosmic density field and its evolution. We could then analyse this data according to the Bayesian update rule

$$p(\boldsymbol{\pi}|\hat{\boldsymbol{d}}) = \frac{p(\hat{\boldsymbol{d}}|\boldsymbol{\pi})p(\boldsymbol{\pi})}{p(\hat{\boldsymbol{d}})} \tag{5.1}$$

in order to obtain a very precise measurement of the unknown parameters $\boldsymbol{\pi}$ of the cosmological standard model (cf. the list at the end of Section 3.1). Such a program requires that we know the likelihood function $p(\hat{\boldsymbol{d}}|\boldsymbol{\pi})$, i.e. the PDF that quantifies the statistical uncertainties of our measurement $\hat{\boldsymbol{d}}$. In Section 3.3 we have calculated this PDF for measurements of the power spectrum, and even within our simplifying assumption of Gaussianity of the density field this was a non-trivial derivation. In Section 4 we haven't even talked about

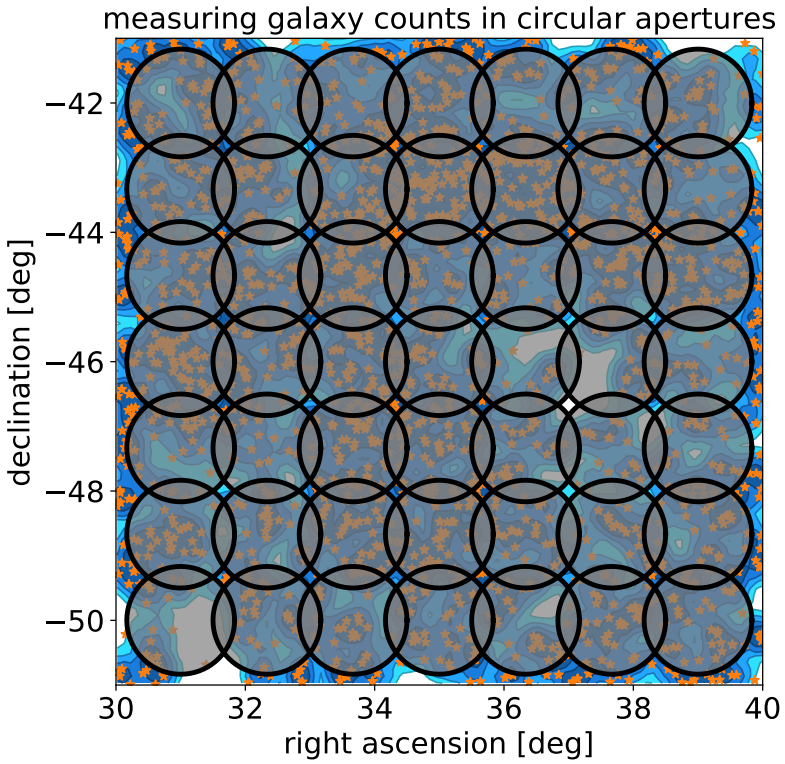

Figure 17. Sketch on how to measure the PDF of galaxy density contrast in circular apertures. Step 1: cover the survey with a set of circular apertures and count the number of galaxies N_g in each aperture. Step 2: calculate the galaxy density contrast in each apertures as $\delta_g = N_g/\bar{N}_g - 1$, where $\bar{N}_g$ is the average value of N_g among all apertures. Step 3: generate a histogram of the values of δ_g in all the apertures. This histogram is an estimate of the 1-point PDF of galaxy density contrast smoothed by the aperture. You can see the result of such a PDF measurement in real data in Figure 18.

how statistical uncertainties in a PDF measurement may be distributed, and determining this distribution is indeed a very difficult task. If we were to include all of the higher order statistics mentioned about into one data vector, then we wouldn't just have to model the statistical uncertainties for each of them separately, but we would have to understand their full, joint distribution including correlations between the different probes. Doing this analytically becomes an unfeasible task very quickly.

So instead one often uses simulated observations in order to estimate the likelihood function $p(\hat{\boldsymbol{d}}|\boldsymbol{\pi})$ needed for cosmological analyses. In this section, we will go through a very brief overview of standard methods to do so.

5.1 Gaussian likelihood assumption and covariance estimation

Very often in cosmological analyses it is simply assumed (and demonstrated by more or less stringent means) that the likelihood function $p(\hat{\boldsymbol{d}}|\boldsymbol{\pi})$ of a given data vector $\hat{\boldsymbol{d}}$ is Gaussian. The likelihood is then fully determined by specification of a mean vector $\boldsymbol{\mu}$ with elements

$$\mu_i = \int \mathrm{d}^n d \; p(\hat{\boldsymbol{d}}|\boldsymbol{\pi}) \; d_i \tag{5.2}$$

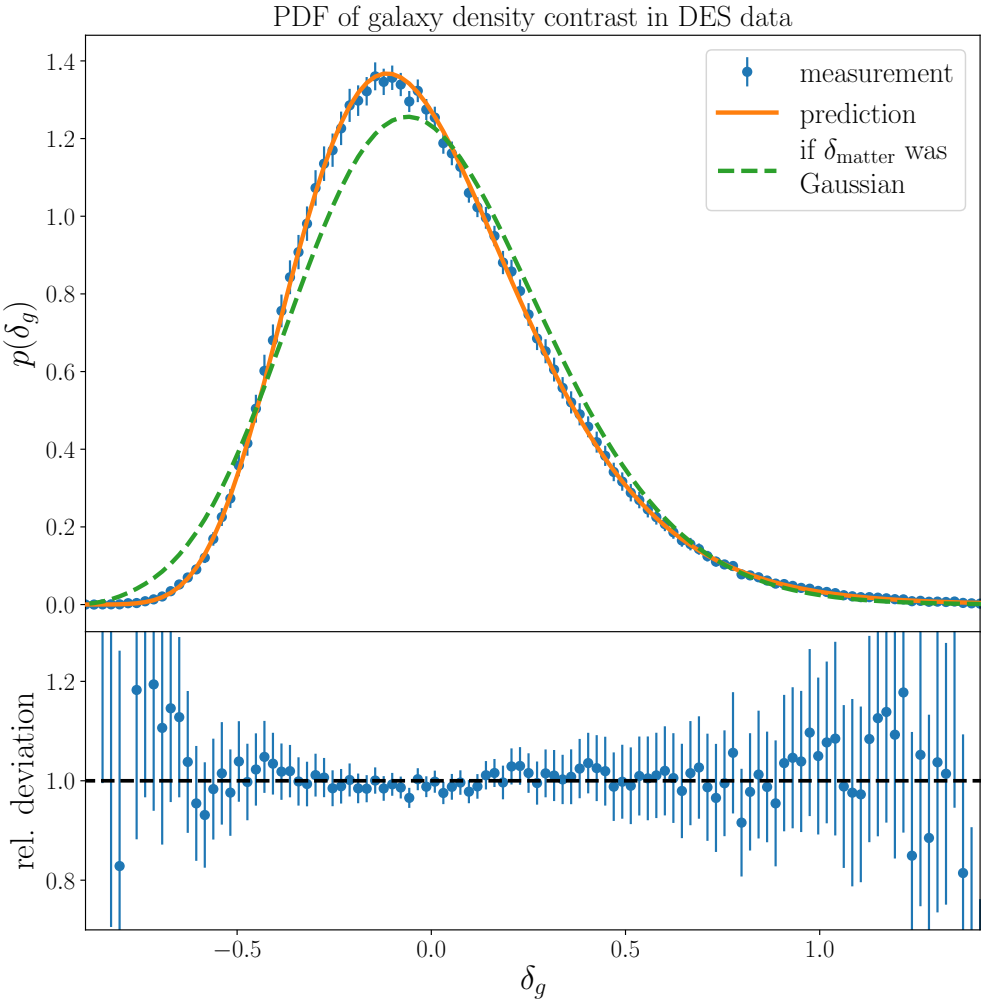

Figure 18. Blue points: observation of the 1-point PDF of galaxy density contrast $\delta_{g,R}$ within a smoothing scale of ~ 10 Mpc (but cf. Section 4.3 for caveat) within observations from the Dark Energy Survey [30]. Orange line: besti fitting theoretical prediction, calculated along the lines of Section 4.2 [26]. Green, dashed line: Same prediction, but assuming that the PDF of matter density contrast δ_R is Gaussian.

and a covariance matrix $\boldsymbol{C}$ with elements

$$C_{ij} = \int \mathrm{d}^n d \; p(\hat{\boldsymbol{d}}|\boldsymbol{\pi}) \; (d_i - \mu_i)(d_j - \mu_j) \; . \tag{5.3}$$

Both $\boldsymbol{\mu}$ and $\boldsymbol{C}$ will depend on the cosmological parameters $\boldsymbol{\pi}$. But at least for the covariance this dependence is often ignored. This may seem like a strange thing to do - we have e.g. seen in Equation 3.41 that the variance of power spectrum measurements is proportional to the square of μ, so it should be similarly sensitive to the parameters as the expectation value of those measurements. In real observations there are however additional components of noise which we haven't talked about in Section 3.3, such as shot-noise which results from the fact that we observe the cosmic density field with the help of discrete tracers (e.g. the galaxies in Figure 1). These additional components are largely independent of the cosmological model,

so computing $\boldsymbol{C}$ at some fiducial set of parameters $\boldsymbol{\pi}_{\rm fid}$ and then treating it as constant is often a very good approximation (cf. [25] for a demonstration of that in the context of 2-point function analyses).

It will often be the case that we have a theoretical model for the expectation value $\boldsymbol{\mu}[\boldsymbol{\pi}]$ but not for the covariance matrix - e.g. in the case of the power spectrum $\boldsymbol{\mu}$ corresponds to two-point statistics of the density field while $\boldsymbol{C}$ corresponds to 4-point statistics, which can be significantly harder to model for non-Gaussian fields. So instead one often estimates $\boldsymbol{C}$ from simulated observations. Let us e.g. assume that for the fiducial parameters $\boldsymbol{\pi}_{\rm fid}$ we can simulate the cosmos and obtain measurements of $\hat{\boldsymbol{d}}$ from such a simulated Universe. If we do this N times, we will obtain N data vectors $\hat{\boldsymbol{d}}_1$, ... , $\hat{\boldsymbol{d}}_N$. Assuming that the initial density field of the Universe was Gaussian, we can draw the initial conditions of these N simulations from the appropriate Gaussian PDFunctional, and this will propagate into appropriate statistical fluctuations amongs the simulated measurements $\hat{\boldsymbol{d}}_i$. A standard estimator for the covariance matrix is then the sample covariance of these measurement, which has the elements

$$\hat{C}_{ij} = \frac{1}{N-1}\sum_{i=1}^{N}(\hat{d}_i - \bar{d}_i)(\hat{d}_j - \bar{d}_j) \ . \tag{5.4}$$

Here $\bar{\boldsymbol{d}}$ is the mean of all data vectors, i.e.

$$\bar{\boldsymbol{d}} = \frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{d}}_i \ . \tag{5.5}$$

Let us assume that our simulations are accurate in the sense that we have drawn initial conditions from the correct distribution and that we have solved the subsequent gravitational evolution correctly. In that case Equation 5.4 is indeed an unbiased estimator of the true covariance matrix, i.e.

$$\langle\hat{\boldsymbol{C}}\rangle = \boldsymbol{C}[\boldsymbol{\pi}_{\rm fid}] \ , \tag{5.6}$$

where the expectation value $\langle\cdot\rangle$ is taken wrt. many sets of simulated Universes. For any finite set of simulated Universes the sample covariance $\hat{\boldsymbol{C}}$ will however be different from $\boldsymbol{C}[\boldsymbol{\pi}_{\rm fid}]$. This is often coined into the phrase "$\hat{\boldsymbol{C}}$ is an unbiased, but noisy estimate of $\boldsymbol{C}[\boldsymbol{\pi}_{\rm fid}]$". This has two important consequences. Firstly, if we want to approximate the likelihood function $p(\hat{\boldsymbol{d}}|\boldsymbol{\pi})$ as a Gaussian PDF (cf. Equation 2.24) then we need to know the inverse covariance matrix. But matrix inversion is a non-linear operation, so the noise in $\hat{\boldsymbol{C}}$ will actually produce a bias for $\hat{\boldsymbol{C}}^{-1}$,

$$\langle\hat{\boldsymbol{C}}^{-1}\rangle \neq \boldsymbol{C}[\boldsymbol{\pi}_{\rm fid}]^{-1} \ . \tag{5.7}$$

By a very convenient miracle of nature, this bias is in fact characterised by a single multiplicative factor for all matrix elements - the so called *Kaufman-Hartlap factor* [32, 38],

$$\langle\hat{\boldsymbol{C}}^{-1}\rangle = \frac{N_{\rm sim}-1}{N_{\rm sim}-N_{\rm data}-2}\,\boldsymbol{C}[\boldsymbol{\pi}_{\rm fid}]^{-1} \ . \tag{5.8}$$

Here $N_{\rm sim}$ is the number of simulations used to estimate $\hat{\boldsymbol{C}}$ and $N_{\rm data}$ is the number of data points in $\hat{\boldsymbol{d}}$. So we can simply divide $\hat{\boldsymbol{C}}^{-1}$ before we insert it into our Gaussian approximation for $p(\hat{\boldsymbol{d}}|\boldsymbol{\pi})$.

Another, more severe but often overlooked problem, is the fact that the noise in the estimator $\hat{\boldsymbol{C}}$ also scrambles around the location of parameter contour such as the ones shown in Figure 10 (while the Kaufman-Hartlap bias only affects the width of parameter contours).

5.2 Simulation based posterior estimation (also known as 'likelihood free inference')

In Section 5.1 we had assumed that the data vector $\hat{\boldsymbol{d}}$ has a Gaussian distribution, and we used simulated observations $\hat{\boldsymbol{d}}_i$ to estimate the covariance matrix of that distribution. But what if $\hat{\boldsymbol{d}}$ contains statistics for which this assumption of Gaussianity is inaccurate? How can we use simulated observations to estimate a general likelihood function $p(\hat{\boldsymbol{d}}|\boldsymbol{\pi})$, or rather: the resulting posterior $p(\boldsymbol{\pi}|\hat{\boldsymbol{d}})$? To achieve this, let us start with the observation that

$$p(\boldsymbol{\pi}|\hat{\boldsymbol{d}}) \propto p(\hat{\boldsymbol{d}}|\boldsymbol{\pi})p(\boldsymbol{\pi}) \equiv p(\hat{\boldsymbol{d}}, \boldsymbol{\pi}) \ , \tag{5.9}$$

where the second equality is due to Bayes' theorem. So to calculate the posterior, we need to know the *joint distribution* of the parameters and the data. How can we use cosmological simulations to estimate this joint distribution? A very primitive way to do so, which nevertheless captures the essence of more sophisticated methods, would be the following:

Goal: obtain independent random draws $(\hat{\boldsymbol{d}}_i, \boldsymbol{\pi}_i)$, $i = 1, \ \ldots \ , N_{\rm sim}$ from the joint distribution $p(\hat{\boldsymbol{d}}, \boldsymbol{\pi})$. Then measure a (high dimensional) histogram from these draw to obtain an estimate of $p(\hat{\boldsymbol{d}}, \boldsymbol{\pi})$.

Step 1: Draw parameter values $\boldsymbol{\pi}_i$, $i = 1, \ \ldots \ , N_{\rm sim}$ from the prior density $p(\boldsymbol{\pi})$. (We can always to this, because we ourselves choose, and hence know, the prior density).

Step 2: For every drawn value $\boldsymbol{\pi}$ of the cosmological parameters, draw Gaussian initial density fluctuations for a simulated universe. The power spectrum needed to characterise this Gaussian distribution is determined by the parameter values $\boldsymbol{\pi}$.

Step 3: Now evolve this initial density field forward in time, using e.g. N-body simulation techniques. This evolution will again depend on the parameter values $\boldsymbol{\pi}$.

Step 4: In your evolved, simulated universe, measure your desired data vector $\hat{\boldsymbol{d}}_i$.

Step 5: From the draws $(\hat{\boldsymbol{d}}_i, \boldsymbol{\pi}_i)$ you can now measure a histogram that estimated the joint distribution $p(\hat{\boldsymbol{d}}, \boldsymbol{\pi})$.

If $N_{\rm param}$ is the number of parameters we consider, and $N_{\rm data}$ is the dimension of our data vector $\hat{\boldsymbol{d}}$, then the joint distribution $p(\hat{\boldsymbol{d}}, \boldsymbol{\pi})$ is defined on an $N_{\rm param} + N_{\rm data}$ dimensional space. In usual cosmological analyses, this dimension will be of the order of several hundreds! In such a situation, the histogram estimator employed above would require a gigantic number $N_{\rm sim}$ of simulations. This will usually make the above approach unfeasible. In the following, we will go over three ways to circumvent, or at least ease this requirement for large numbers of simulations.

5.3 Data compression

5.4 Kernel density estimation

5.5 Parametric density models

One possible way to do that is *approximate Bayesian computation* (ABC). This starts from the observation that the postertior can - up to an overall normalisation factor - be expressed

as

$$\begin{aligned} p(\boldsymbol{\pi}|\hat{\boldsymbol{d}}) &\propto p(\hat{\boldsymbol{d}}|\boldsymbol{\pi})p(\boldsymbol{\pi}) \\ &= p(\boldsymbol{\pi}) \int \mathrm{d}^n d\ \delta_D^{(n)}(\hat{\boldsymbol{d}} - \boldsymbol{d})\ p(\boldsymbol{d}|\boldsymbol{\pi}) \\ &\approx p(\boldsymbol{\pi}) \int \mathrm{d}^n d\ \mathrm{ker}_\epsilon(\hat{\boldsymbol{d}} - \boldsymbol{d})\ p(\boldsymbol{d}|\boldsymbol{\pi}) \ . \end{aligned} \tag{5.10}$$

Here, the second line is a trivial re-formulation of the first, while in the third line we approximated the Dirac delta function $\delta_D^{(n)}(\hat{\boldsymbol{d}} - \boldsymbol{d})$ by some kernel $\mathrm{ker}_\epsilon(\hat{\boldsymbol{d}} - \boldsymbol{d})$. This could e.g. be a Gaussian kernel, such as

$$\mathrm{ker}_\epsilon(\hat{\boldsymbol{d}} - \boldsymbol{d}) = \frac{1}{(2\pi\epsilon)^{n/2}} \exp\left(-\frac{1}{2}\frac{|\hat{\boldsymbol{d}} - \boldsymbol{d}|^2}{\epsilon}\right) \ , \tag{5.11}$$

or a simple top-hat kernel

$$\mathrm{ker}_\epsilon(\hat{\boldsymbol{d}} - \boldsymbol{d}) = \frac{\theta(|\hat{\boldsymbol{d}} - \boldsymbol{d}| - \epsilon)}{\pi^{n/2}\epsilon^n/\Gamma(n/2+1)} \ , \tag{5.12}$$

which vanishes outside of a sphere of radius ϵ and which is $1/V_\epsilon$ inside that sphere, where $V_\epsilon = \pi^{n/2}\epsilon^n/\Gamma(n/2+1)$ is the volume of that sphere. In the limit $\epsilon \to 0$ both of these kernels recover the exact posterior. Of course, so far we haven't gained much, since our - unknown - likelihood function $p(\cdot|\boldsymbol{\pi})$ still appears on the left-hand side of Equation 5.10. But assume that we can simulate observations $\hat{\boldsymbol{d}}_i$ for all relevant values of the parameter vector $\boldsymbol{\pi}$, in the same manner as was assumed for covariance estimation in Section 5.1. We can then approximate the integral in the last line of Equation 5.10 as

$$\int \mathrm{d}^n d\ \mathrm{ker}_\epsilon(\hat{\boldsymbol{d}} - \boldsymbol{d})\ p(\boldsymbol{d}|\boldsymbol{\pi}) \approx \frac{1}{N_\mathrm{sim}} \sum_{i=1}^{N_\mathrm{sim}} \mathrm{ker}_\epsilon(\hat{\boldsymbol{d}} - \hat{\boldsymbol{d}}_i) \ . \tag{5.13}$$

The dependence of the left-hand side of this equation on the parameters $\boldsymbol{\pi}$ is hidden in the fact that out procedure to generate the simulated observations $\hat{\boldsymbol{d}}_i$ depends on $\boldsymbol{\pi}$. If we want to calculate the posterior $p(\boldsymbol{\pi}|\hat{\boldsymbol{d}})$ for grid of parameter vectors $\boldsymbol{\pi}_j$, $j = 1 \ \ldots \ N_\mathrm{grid}$, then we could generate simulated measurements $\hat{\boldsymbol{d}}_{i,j}$, $i = 1 \ \ldots \ N_\mathrm{sim}$, at each grid point $\boldsymbol{\pi}_j$ and then approximate the posterior as

$$p_\mathrm{ABC}(\boldsymbol{\pi}_j|\hat{\boldsymbol{d}}) \propto \frac{1}{N_\mathrm{sim}} \sum_{i=1}^{N_\mathrm{sim}} \mathrm{ker}_\epsilon(\hat{\boldsymbol{d}} - \hat{\boldsymbol{d}}_{i,j})\ p(\boldsymbol{\pi}_j) \ . \tag{5.14}$$

This is the essence of ABC. And it can be shown that for $\epsilon \to 0$ and $N_\mathrm{sim} \to \infty$ it indeed recovers the exact posterior at each grid point $\boldsymbol{\pi}_j$. But for any finite number of simulations N_sim it is only a noise estimate of the posterior, and for any finite width $\epsilon > 0$ of the kernel it is a biased estimate of the posterior. In particular, the ABC posterior will on average be wider than the exact posterior because the last line of Equation 5.10 smears out the exact likelihood function $p(\cdot|\boldsymbol{\pi})$ with ker_ϵ .

There is however a much more severe problem with ABC: it is extremely inefficient with the simulated data. Say e.g. that ker_ϵ is the top-hat kernel of Equation 5.12. Then Equation 5.13 means that in order to estimate the posterior we throw away all simulated measurements that are more than ϵ away from the true data $\hat{\boldsymbol{d}}$. Since we are interested in the limit $\epsilon \to 0$, this will be almost all simulated measurements! In other words, only a small fraction of our simulated data actually enters the ABC estimator of the posterior.

5.6 Direct analyses of the large-scale density field

It is the holy grail of large-scale structure cosmology to use the full information content of the cosmic density field for parameter inference. To achieve this, one would optimally not compress the density field $\delta(\cdot)$ into any summary statistics (such as e.g. the power spectrum), but directly consider the likelihood functional $\mathcal{P}[\delta(\cdot)|\boldsymbol{\pi}]$, which quantifies how likely the observed density field $\delta(\cdot)$ was if $\boldsymbol{\pi}$ were the true values of the cosmological parameters.

If we observe the density field as some late time of cosmic evolution, finding this likelihood functional is very non-trivial, because the late time density field is not a Gaussian random field anymore. However, assuming that there is a deterministic connection between the late time and early time density field, we can make use of the fact that the initial density contrast field was a Gaussian random field. We can then write the likelihood functional of the late field as

$$\mathcal{P}[\tilde{\delta}_{\rm late}(\cdot)|\boldsymbol{\pi}] = \frac{1}{|2\pi\tilde{C}_i(\boldsymbol{\pi})|^{1/2}}\ \exp\left\{-\frac{1}{2}\int\frac{{\rm d}^3k}{(2\pi)^3}\ \frac{|\tilde{\delta}_i[\tilde{\delta}_{\rm late}\,,\boldsymbol{k}]|^2}{P(k,\boldsymbol{\pi})}\right\}\ \left|\frac{\mathcal{D}\delta_i}{\mathcal{D}\delta_{\rm late}}\right|. \tag{5.15}$$

The left-handside of this equation consists of the Gaussian PDFunctional of the initial density contrast field δ_i and of the Jacobian determinant $|\mathcal{D}\delta_i/\mathcal{D}\delta_{\rm late}|$ of the mapping that connects the late time field to the initial field. Evaluating this mapping (and its Jacobi determinant) analytically is of course unfeasible. But it can be numerically evaluated with the help of cosmological simulations. This way one can indeed derive

WILL COME AS A SUPPLEMENT LATER IN THE COURSE

6 Discussion: What do we mean by probability?

At different points in the previous sections we have come in contact with questions of how to interpret probabilities. In particular, we have encountered two schools of thought: interpretations of probability based on frequencies of outcomes in repeated experiments and subjective interpretations (which are sometimes referred to as Bayesian interpretations). Let us briefly collect what we have learned about these two.

6.1 Frequentist interpretations

Let the possible outcomes E_1 , ... , E_N of a random process $\mathcal{R}$ have the probabilities P_1 , ... , P_N . What do these probabilities mean? A frequentist interpretation would be to say that in many independent but identical realisations $\mathcal{R}_1$, ... , $\mathcal{R}_K$ of the process $\mathcal{R}$ the outcome E_i would approximately appear with a frequency P_i. And that approximation would become better and better as the number K of independent realisations of $\mathcal{R}$ is taken to be bigger and bigger. In fact, based on the probabilities P_i frequentism can even predict a probability distribution for the possible number of occurrences of the even E_i - it should be given by the binomial distribution

$$P(\#E_i = n) = P_i^n\ (1-P_i)^{K-n}\ \binom{K}{n} \tag{6.1}$$

and this distribution peaks more and more sharply around $\#E_i = P_iK$ as $K\to\infty$.

Of course, we had already said that there is a cyclicity in this definition of probability: to define the meaning of probabilities of the outcomes of the random process $\mathcal{R}$ we refer to the probabilities of the outcome of another random process (repeated realisations of $\mathcal{R}$). We

had also seen, that this cyclicity is not easily overcome by considering infinite realisations of $\mathcal{R}$, because there are no well defined frequencies in infinite sequences - at least if we consider two sequences of possible outcomes to be equivalent as long as they are re-shuffled versions of each other.

6.2 Bayesian interpretations

Bayesianists would take the probabilities P_1 , ... , P_N to quantify the degrees of believe of some agent (i.e. some person) in whether or not the random process $\mathcal{R}$ will result in the events E_1 , ... , E_N . What are these degrees of believe? One can attempt to answer this question with a decision theoretic approach - see e.g. [45, 59] (though we do not strictly follow their more detailed line of thought). Consider the following situation: The agent has to choose between N different actions A_1 , ... , A_N . After they have performed one of these actions the random process $\mathcal{R}$ takes place. Now assume that each potential outcome E_i of $\mathcal{R}$ holds some reward for the agent in the form of a utility U_i (you can e.g. think of U_i as some amount of money the agent gets paid if the event E_i occurs). But the twist is: the agent only gets the reward U_i if they have also taken the action A_i . If they have instead taken an action A_j with $j \neq i$ then the reward is 0 . Based on this information, which action will the agent choose?

The idea behind the above setup is **not** that the probabilities P_1 , ... , P_N provide an answer to the above question. Instead, if we know the answer to the above question for all possible utility assignments, then we can **infer** what the probability assigments P_1 , ... , P_N of the agent are. To gain a better understanding of this, let us consider the case where $N = 2$, and let us assume that the agent choosed the action A_1 whenever

$$\frac{U_1}{U_2} > \frac{1-p}{p} \tag{6.2}$$

for a number $p \in [0, 1]$. Then the decision theoretic perspective is that the subjective probability assignments of the agent are

$$P_1 = p \ , \ P_2 = 1 - p \ . \tag{6.3}$$

Note that in order for this assignment to be well defined, the utility needs to satisfy a linearity property: The agents decision between action A_1 and action A_2 has to remain the same if the utilities U_1 and U_2 are exchanged with new utilities $\tilde{U}_1 = \alpha U_1$ and $\tilde{U}_2 = \alpha U_2$, where α is some real number.

This approach for defining the meaning of probability smells a bit like frequentism, because if p and $1 - p$ where frequentist probabilities, then pU_1 and $(1 - p)U_2$ would be the average utility that is expected upon many repeatitions of the above situation. And then the decision criterion 6.2 would maximise the average payout for the agent. But the agent is of course not forced to assign frequentist probabilities to the outcomes of $\mathcal{R}$ - maybe the circularity of the frequentist interpretation makes it impossible for them to even think in a frequentist way. And even if they did assign such probabilities, they are not forced to act such that these probabilities agree with the ones through which the above decision theoretic ansatz would characterize their behaviour. Within the above interpretation probabilities are simply a way to describe **how** an agent will act, but the question of **why** they act this way is left unanswered.

Interpreting probabilities as degrees of believe has the advantage(?) that we can apply probabilistic calculus even to situations one may not consider to be outcomes of truly random

processes. For example, you may not think that the values of certain natural constants are the result of random draws. But you may nevertheless have certain degrees of believe on what the likeli values for those constants are. And with every new piece of information you may update these degrees of believe - e.g. along the line of the Bayesian update procedure of Equation 3.18 . While this procedure is handy and easily implemented in practice, we have seen in **example 2** of Section 3.2 that it can lead to non-intuitive consequences. The true Bayesians may try to defend this by pointing out that the prior density in Equation 3.18 is of crucial importance in the Bayesian program, and that the prior chosen in **example 2** was simply wrong. But there is a prior no way of knowing which prior will lead Bayesian inference to have the desired properties upon many realisations of an experiment. So the more honest reaction would be to say: Bayesian inference is not about success frequencies in repeated experiment. It is just a consistent way of updating ones personal degrees of believe (and hence decision strategy) upon the arrival of new data.

So the frequentist interpretation only gives a circular definition of probability, and the subjective interpretation reduces probabilities to mere descriptions of how a person will act without worrying whether or not a certain action strategy make sense in the context of the real world. Of course, if you follow the update procedure of Equation 3.18, then your degrees of believe will eventually be informed by the frequencies of real life events. Quantum mechanics tells us that many of these events are indeed outcomes of truly random processes. But in a sense, this only throws us back at our initial question: What are probabilities? In particular, what are the probabilities that quantum mechanics talks about?

Exercise 44

(During lecture, no homework) *Your lecturer has 8 coins: one 2-Euro coin and seven 50-cent coins. You now have the opportunity to win one of these coins! First, let us go to the webpage* `https://qrng.anu.edu.au/dice-throw/` *, which provides truly random numbers by observing quantum vacuum fluctuations in the electro-magnetic field (see [33] for details). Your lecturer will use this service to draw a uniform random number* $1 \leq n \leq 8$ *. If* $n = 1$*, they will put the 2-Euro coin into an envelope. If* $n > 1$ *then they will put one of the 50-cent coins into an envelope. Now guess which type of coin is in the envelope! If you guess correctly, the you will obtain the coin in the envelope (or an equivalent coin - I have enough for everyone).*

Before we actually play out this situation, let's form groups of 3-4 people and discuss the following points within these groups.

A) Who is the agent in this scenario? What are the actions A_i *and what are the outcomes* E_i*?*

B) What action A_i *do you take?*

C) Would your answer to B) change if the rewards - either 2 Euros or 50 cents - were multiplied by some number? E.g. what if the rewards where 200 Euros and 50 Euros? What if they were 20 Mio Euros and 5 Mio Euros?

D) Would your answer change if the rewards were 2 Euros and 10 cents? What if they were 200 Mio Euros and 10 Mio Euros?

E) Based on your answers to C) and D), is utility linear with money for you? If not, estimate which function $f(\text{money})$ *is your utility.*

F) What probabilities P_i do you assign to the different outcomes E_i?

G) Is your assignment in F) motivated by a Frequentist or by a Bayesian interpretation of probability (or by another interpretation)? Give an argument for why you find either side more convincing. What do you say to the criticism of either method that was voiced above?

H) Are there questions, to which we can in principle never know the answer, but for which you nevertheless have certain degrees of believe about its possible answers? What does this mean for both the Frequentist and the decision theoretic interpretation of probability.

Part II:
Quantum Probabilities and the early-time Cosmic Density Field

7 Journey to Quantum Field Theory

To understand how the initial density fluctuations of the Universe have been generated, we will need to tread the waters of quantum mechanics and, ultimately, quantum field theory. This is in fact quite a challenge, because as of today (September 26, 2023) it is not unreasonable to believe that nobody on earth has understood Quantum mechanics. Throughout the upcoming sections I will do my very best to make clear, that I do not understand it either. But if there are steps in the following derivations and discussions which you are confused about despite the fact that I make them seem obvious - please note that I have most likely also not understood those steps.

7.1 Re-casting discrete random variables

In Section 2 we started our discussion of probability theory by looking at discrete random variables. We will slightly modify our language from that section in an attempt to discuss Quantum and classical probabilities in a unified manner. Let E_1 , ... , E_N be the possible outcomes of a discrete random process and let us call these outcomes the *elementary events* of the random process. Furthermore, we will call a PDF that assigns probabilities $P(E_i)$ to each elementary event E_i *a state*.

If we are given a number of different states P_1 , ... , P_K and a set of real numbers c_1 , ... , c_K with $c_j \geq 0$ and $c_1 + \ldots + c_K = 1$ then the mixture

$$P = \sum_{j=1}^{K} c_j P_j \tag{7.1}$$

is also a state (i.e. it is also a PDF over the events E_1 , ... , E_N). In fact, we can express each state P as a mixture of other states P_j , e.g. if we set $K = N$, $P_j(E_i) = \delta_{ij}$ (the Kronecker δ) and $c_j = P(E_j)$.

Exercise 45

Find other examples of how to express a generic state (a generic PDF) as a mixture of two other states.

We will call a state P a *pure state* if it cannot be expressed as a non-trivial mixture of other states. This means that for any decomposition of P as in Equation 7.1 we have either $c_j = 0$ or $P_j = P$. All states that can be expressed as non-trivial mixtures of other states will be called *mixed states*.

Exercise 46

Show that the only pure states are the "Kronecker δs", i.e. the states with $P(E_i) = \delta_{ij}$ for some j.

7.2 Step 1: finite dimensional Hilbert spaces

Let $\mathcal{H}$ be a complex vector space with an inner product

$$\begin{aligned} \langle \cdot, \cdot \rangle : \mathcal{H} \oplus \mathcal{H} &\to \mathbb{C} \\ (\boldsymbol{v}, \boldsymbol{w}) &\mapsto \langle \boldsymbol{v}, \boldsymbol{w} \rangle \end{aligned}$$

such that for all vectors $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in \mathcal{H}$ and all complex numbers $\alpha \in \mathbb{C}$

$$\langle \boldsymbol{u} + \boldsymbol{v}, \boldsymbol{w} \rangle = \langle \boldsymbol{u}, \boldsymbol{w} \rangle + \langle \boldsymbol{v}, \boldsymbol{w} \rangle \tag{7.2}$$

$$\langle \boldsymbol{v}, \alpha \boldsymbol{w} \rangle = \alpha \langle \boldsymbol{w}, \boldsymbol{v} \rangle \tag{7.3}$$

$$\langle \boldsymbol{v}, \boldsymbol{w} \rangle = \langle \boldsymbol{w}, \boldsymbol{v} \rangle^* \tag{7.4}$$

$$0 < \langle \boldsymbol{v}, \boldsymbol{v} \rangle \quad \text{for all } \boldsymbol{v} \neq 0 \ , \tag{7.5}$$

where * is complex conjugation. The inner product defines a norm

$$||\boldsymbol{v}|| \equiv \sqrt{\langle \boldsymbol{v}, \boldsymbol{v} \rangle} \tag{7.6}$$

on $\mathcal{H}$, and $\mathcal{H}$ is called a *Hilbert space* if it is complete wrt. the distance metric

$$d(\boldsymbol{v}, \boldsymbol{w}) \equiv ||\boldsymbol{v} - \boldsymbol{w}|| \ . \tag{7.7}$$

Let us for now only consider a Hilbert space of finite dimension N. Then we can always find a basis of $\mathcal{H}$ such that the inner product of two vectors $\boldsymbol{v} = (v_1 \ , \ \ldots \ , \ v_N)^T$ and $\boldsymbol{w} = (w_1 \ , \ \ldots \ , \ w_N)^T$ is given by

$$\langle \boldsymbol{v}, \boldsymbol{w} \rangle = \boldsymbol{v}^\dagger \cdot \boldsymbol{w} = v_1^* w_1 \ + \ \ldots \ + \ v_N^* w_N \ , \tag{7.8}$$

where $\boldsymbol{v}^\dagger$ is the transposed and complex conjugate of the vector $\boldsymbol{v}$.

Each vector $\boldsymbol{v}$ in $\mathcal{H}$ spans an entire, 1-dimensional sub-vector space

$$\text{span}(\boldsymbol{v}) \equiv \{\alpha \boldsymbol{v} | \alpha \in \mathbb{C}\} \subset \mathcal{H} \ . \tag{7.9}$$

In analogy to the previous subsection, let us **call these 1-dimensional subspaces the elementary events of a quantum theory** based on the Hilbert space $\mathcal{H}$. Why do we specifically choose 1D sub-spaces to represent these events, and not, say, circles in Hilbert space? A cheap answer would be: this is just how we found nature to work. But we will look at a more elaborate answer in Section 9.

On each Hilbert space there will be linear operators operators

$$\begin{aligned} \hat{O} : \mathcal{H} &\to \mathcal{H} \\ \boldsymbol{v} &\mapsto \hat{O}\boldsymbol{v} \ , \end{aligned}$$

which in finite dimensions N you can always think of as square matrices. The adjoint $\hat{O}^\dagger$ of an operator $\hat{O}$ is defined such that

$$\langle \boldsymbol{v}, \hat{O}\boldsymbol{w} \rangle = \langle \hat{O}^\dagger \boldsymbol{v}, \boldsymbol{w} \rangle \tag{7.10}$$

for all $\boldsymbol{v}$ and $\boldsymbol{w}$. In the finite dimensional case, and viewing $\hat{O}$ as a matrix, $\hat{O}^\dagger$ is the transposed and complex conjugate matrix of $\hat{O}$. In quantum mechanics, observables are represented by operators $\hat{A}$ that are selfadjoint, i.e. which satisfy $\hat{A} = \hat{A}^\dagger$. Such operators are also called *Hermitian operators*. It is both a standard and central result of linear algebra that Hermitian matrices admit an orthonormal basis (an ONB) of *eigenvectors*. This means that for each Hermitian $\hat{A}$ (on our finite dimensional Hilbert space) there is a set of vectors $\boldsymbol{a}_1 \ , \ \ldots \ , \ \boldsymbol{a}_N$ such that

$$\langle \boldsymbol{a}_i, \boldsymbol{a}_j \rangle = \delta_{ij} \tag{7.11}$$

$$\hat{A}\boldsymbol{a}_i = \lambda_i \boldsymbol{a}_i \ , \tag{7.12}$$

where the λ_i are real numbers and are called the *eigenvalues* of $\hat{A}$. If $\hat{A}$ represents a physical observable then λ_i are the possible values of that observable.

Exercise 47

Show that if $\hat{A}$ is Hermitian and $\boldsymbol{v}$ is any vector, then $\langle \boldsymbol{v}, \hat{A}\boldsymbol{v} \rangle$ is always a real number.

Given any ONB $\boldsymbol{v}_1$, ... , $\boldsymbol{v}_N$ we can define the *trace* of an operator $\hat{O}$ to be

$$\mathrm{tr}(\hat{O}) \equiv \sum_{i=1}^{N} \langle \boldsymbol{v}_i, \hat{O}\boldsymbol{v}_i \rangle \ . \tag{7.13}$$

It can be shown that this definition is indeed independent of the choice of the ONB. If an operator $\hat{\rho}$ satisfies the properties

$$\mathrm{tr}(\hat{\rho}) = 1 \tag{7.14}$$

$$\hat{\rho} = \hat{\rho}^\dagger \quad \text{(i.e. } \hat{\rho} \text{ is Hermitian)} \tag{7.15}$$

$$\langle \boldsymbol{v}, \hat{\rho}\ \boldsymbol{v} \rangle \geq 0 \quad \text{for all } \boldsymbol{v} \tag{7.16}$$

then we call $\hat{\rho}$ a state. Why do we call these operators our states, and not literally anything else? The cheap answer would again be that this is just how nature works. But we will see a much better answer in Section 9.

In analogy to the previous subsection, if we are given a number of different states $\hat{\rho}_1$, ... , $\hat{\rho}_K$ and a set of real numbers c_1 , ... , c_K with $c_j \geq 0$ and $c_1 + \ldots + c_K = 1$ then the mixture

$$\hat{\rho} = \sum_{i=1}^{K} c_i \hat{\rho}_j \tag{7.17}$$

is also a state. We will **call $\hat{\rho}$ a pure state if it can only be expresses as mixtures where either $c_j = 0$ or $\hat{\rho}_j = \hat{\rho}$** . And we will call $\hat{\rho}$ a mixed state if it is not a pure state.

We had previously associated states with PDFs that assign probabilities to elementary events. And in the quantum context we defined elementary events to be the 1-dimensional subspaces of the Hilbert space $\mathcal{H}$. So is there a sense in which a quantum state $\hat{\rho}$ defines "probabilities" of such events? We can attempt such a definition as

$$P_{\hat{\rho}}\left(\mathrm{span}(\boldsymbol{v})\right) = \frac{\langle \boldsymbol{v}, \hat{\rho}\ \boldsymbol{v} \rangle}{\langle \boldsymbol{v}, \boldsymbol{v} \rangle} \ . \tag{7.18}$$

Because we are normalising by $\langle \boldsymbol{v}, \boldsymbol{v} \rangle$ this is indeed well defined, i.e. it does not depend on which vector $\boldsymbol{v}$ we use to represent $\mathrm{span}(\boldsymbol{v})$. To simplify our notation, we will in the following simply write $P_{\hat{\rho}}(\boldsymbol{v})$ instead of $P_{\hat{\rho}}(\mathrm{span}(\boldsymbol{v}))$.

Exercise 48

Show explicitly that two vectors $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ for which $\mathrm{span}(\boldsymbol{v}_1) = \mathrm{span}(\boldsymbol{v}_2)$ give the same value for $P_{\hat{\rho}}(\mathrm{span}(\boldsymbol{v}_i))$.

But in what sense are these $P_{\hat{\rho}}(\boldsymbol{v})$ probabilities? Consider an ONB $\boldsymbol{v}_1$, ... , $\boldsymbol{v}_N$. Then the values

$$P_{\hat{\rho}}(\boldsymbol{v}_i) = \frac{\langle \boldsymbol{v}_i, \hat{\rho}\ \boldsymbol{v}_i \rangle}{\langle \boldsymbol{v}_i, \boldsymbol{v}_i \rangle} = \langle \boldsymbol{v}_i, \hat{\rho}\ \boldsymbol{v}_i \rangle \tag{7.19}$$

sum up to 1 by virtue of the trace condition in Equation 7.14 and they are all ≥ 0 by virtue of Equation 7.16 . So for any set of mutually orthogonal events $E_1 = \mathrm{span}(\boldsymbol{v}_1)$, ... , $E_N =$

span($\boldsymbol{v}_N$) a state $\hat{\rho}$ indeed defines a PDF in the spirit of Section 7.1 . This gains physical meaning when measuring an observable $\hat{A}$. If $\boldsymbol{a}_1$, ... , $\boldsymbol{a}_N$ is an ONB of eigenvectors of $\hat{A}$ with eigenvalues λ_1 , ... , λ_N , then

$$P_{\hat{\rho}}(\boldsymbol{a}_i) = \langle \boldsymbol{a}_i, \hat{\rho}\ \boldsymbol{a}_i \rangle \tag{7.20}$$

are the probabilities of obtaining the measurement outcome λ_i . But despite these formal analogies between our quantum definition of events and states and the corresponding concepts for discrete random variables, there are a number of severe differences.

Exercise 49

Find and discuss at least two differences between the events and states we defined in Section 7.1 and the quantum events and states we discussed above.

Since a state $\hat{\rho}$ is also a Hermitian operator, we can find an ONB $\boldsymbol{r}_1$, ... , $\boldsymbol{r}_N$ of eigenvectors of $\hat{\rho}$. If λ_1 , ... , λ_N are the corresponding eigenvalues, then we can express the state as

$$\hat{\rho} = \sum_{i=1}^{N} \lambda_i\ \boldsymbol{r}_i \boldsymbol{r}_i^\dagger\ , \tag{7.21}$$

where $\boldsymbol{v}\boldsymbol{v}^\dagger$ is the dyadic product of a vector $\boldsymbol{v}$ with itself, i.e. the matrix whose element $m-n$ is given by $v_m v_n^*$.

Exercise 50

Demonstrate that the decomposition of Equation 7.21 is indeed correct. Hint: it is sufficient to show that the vectors $\boldsymbol{r}_i$ are eigenvectors of the right-hand side of that equation with eigenvalue λ_i. This uniquely characterizes any Hermitian matrix.

Exercise 51

Show that for any vector $\boldsymbol{v}$ with $\langle \boldsymbol{v}, \boldsymbol{v} \rangle = 1$ the dyadic product $\boldsymbol{v}\boldsymbol{v}^\dagger$ is a state. Use this, together with Equation 7.21 to show that any pure state is of the form $\hat{\rho} = \boldsymbol{v}\boldsymbol{v}^\dagger$ for a normalised vector $\boldsymbol{v}$.

Because of our result from Exercise 51 one often identifies pure states $\boldsymbol{v}\boldsymbol{v}^\dagger$ with the normalised vector $\boldsymbol{v}$. And a state $\hat{\rho}$ which is decomposed as in Equation 7.21 is often interpreted as saying "we know with (classical) probabilities λ_i that the quantum system is in the state $\boldsymbol{r}_i \boldsymbol{r}_i^\dagger$ ". This implies a distinction between classical and quantum uncertainties. But in reality the line between those two types of uncertainty is somewhat fluent. Consider e.g. two pure states represented by two vectors $\boldsymbol{v}_1 \neq 0 \neq \boldsymbol{v}_2$ which are **not** orthogonal to each other but which still span different 1D subspaces, span($\boldsymbol{v}_1$) $\neq$ span($\boldsymbol{v}_2$) . Now assume that we know with probability p that a quantum system is in the pure state $\boldsymbol{v}_1 \boldsymbol{v}_1^\dagger$ and with a probability $(1-p)$ that it is in the pure state $\boldsymbol{v}_2 \boldsymbol{v}_2^\dagger$. The mixed state representing our knowledge about the system would then be

$$\hat{\rho} = p\ \boldsymbol{v}_1 \boldsymbol{v}_1^\dagger + (1-p)\ \boldsymbol{v}_2 \boldsymbol{v}_2^\dagger\ . \tag{7.22}$$

This $\hat{\rho}$ is a matrix of rank 2, and it will thus have two non-zeros eigenvalues λ_1 and λ_2 and corresponding eigenvectors $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$. Hence, we can decompose $\hat{\rho}$ as

$$\hat{\rho} = \lambda\ \boldsymbol{r}_1 \boldsymbol{r}_1^\dagger + (1-\lambda)\ \boldsymbol{r}_2 \boldsymbol{r}_2^\dagger\ . \tag{7.23}$$

The eigenvectors $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ **will be orthogonal**! So Equation 7.23 decomposes $\hat{\rho}$ into fundamentally different states than Equation 7.22 . If we are only given $\hat{\rho}$, how should we interpret

it? As being in the states $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ with classical probabilities p and $(1-p)$ or as being in the states $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ with classical probabilities λ and $(1-\lambda)$? There is no intrinsic way to answer this question. But in Section 10 we will attempt an answer within the concept of decoherence.

7.3 Re-casting continuous random variables

On the classical side of things, let us now move to random processes with continuos sets of outcomes. In particular, let us assume that elementary events are represented by points in the 2-dimensional plane $\mathbb{R}^2$. With physical applications in mind, we will take one axes of this plane to represent the (generalised[4]) position q of a physical system, and the other axes to represent the (generalised) momentum p of that system. We call any real function ρ on $\mathbb{R}^2$ with $p \geq 0$ and

$$\int \mathrm{d}q\mathrm{d}p \ \rho(p,q) = 1 \tag{7.24}$$

a state. Within the language of Section 2 these states can be considered as PDFs of 2-dimensional random vectors. We can define the notion of pure and mixed states in complete analogy to Section 7.1 and it can be shown that the only pure states are the "Dirac δ's"

$$\rho_{q_0,p_0}(q,p) = \delta_D(q-q_0) \ \delta_D(p-p_0) \ . \tag{7.25}$$

Any state can be decomposed into such pure states as

$$\rho(p,q) = \int \mathrm{d}q_0\mathrm{d}p_0 \ \rho(p_0,q_0) \ \rho_{q_0,p_0}(q,p) \ . \tag{7.26}$$

If the evolution of our system is govern by a Hamilton function $H(q,p,t)$ then the state ρ will change with time, and according to Liouville's theorem, this change is governed by the equation

$$\frac{\partial \rho}{\partial t} = -\{\rho, H\} \equiv -\left(\frac{\partial \rho}{\partial q}\frac{\partial H}{\partial p} - \frac{\partial \rho}{\partial p}\frac{\partial H}{\partial q}\right) \ . \tag{7.27}$$

Here the last equality defines the so called *Poisson braket* $\{\cdot,\cdot\}$.

7.4 Step 2: Quantum mechanics

In a finite dimensional Hilbert space, all Hermitian operators have a discrete spectrum. So to mirror the step (discrete) $\to$ (continuous) from Section 7.3 we need to move to infinite dimensional Hilbert space. To be on the safe side, let's assume that the dimension of $\mathcal{H}$ is only countably infinite, i.e. that $\mathcal{H}$ is a *separable Hilbert space*. This is also handy because all complex, separable Hilbert spaces are isomorphic to each other, or in other words: they are all exactly the same. This is why you will sometimes see people refer to $\mathcal{H}$ as *the* Hilbert space.

To move closer to usual notation in quantum mechanics, let us now denote vectors in $\mathcal{H}$ by $|\psi\rangle$ and the corresponding adjoint vectors by $\langle\psi|$. The inner product between two vectors $|\psi_1\rangle$ and $|\psi_2\rangle$ is then written as $\langle\psi_1|\psi_2\rangle$. One representation for $\mathcal{H}$ is the space of functions

$$\psi : \mathbb{R} \to \mathbb{C}$$

[4]Here "generalised" means that q does not have to literally be the position of some point particle.

which are square integrable, i.e. which satisfy

$$\int \mathrm{d}q \; |\psi(q)|^2 < \infty \; . \tag{7.28}$$

To really turn this function space into a Hilbert space we need to identify functions ψ_1 and ψ_2 which differ only on a null-set of points, but let us keep that technicality aside. The inner product of two vectors ψ_1 and ψ_2 is then defined as

$$\langle \psi_1 | \psi_2 \rangle = \int \mathrm{d}q \; \psi_1(q)^* \; \psi_2(q) \; . \tag{7.29}$$

On this Hilbert space $\mathcal{H}$ we can still define linear operators, and it is still true, that a Hermitian operator $\hat{A}$ (i.e. an operator with $\hat{A} = \hat{A}^\dagger$) admits a orthonormal basis $\{|a\rangle_i\}$ of eigenvectors for which

$$\hat{A} \, |a\rangle_i = a_i \, |a\rangle_i \tag{7.30}$$

and where the eigenvalues a_i are still real numbers. The only difference is that now this eigenbasis contains countably infinite elements and that there are potentially countably infinite eigenvalues associated to them.

The fact that a Hermitian operator defined on $\mathcal{H}$ can have (at most) a countably infinite number of eigenvalues seems to be a problem for our program to mimic continuous random processes within our quantum language, since apparently every observable can only have a discrete (albeit infinite) set of values! But this is only true for operators defined on the entirety of the Hilbert space. And one curiosity of infinite dimensional Hilbert spaces is that there can be operators which are well defined only on part of $\mathcal{H}$. In the function space representation of above two common examples for this are the operators

$$\begin{aligned} \hat{Q} : (\text{part of } \mathcal{H}) &\to \mathcal{H} \\ \psi(q) &\mapsto q\psi(q) \end{aligned}$$

and

$$\begin{aligned} \hat{P} : (\text{part of } \mathcal{H}) &\to \mathcal{H} \\ \psi(q) &\mapsto \; -i\frac{\partial \psi(q)}{\partial q} \; , \end{aligned}$$

which you will recognise as the *position* and *momentum operators*. Let us explore in the following exercise why they are not defined on the entirety of $\mathcal{H}$.

Exercise 52

Consider the functions

$$\psi_1(q) = \begin{cases} 1 & \text{for } q \in [-1,1] \\ \\ 1/q^{\frac{3}{2}} & \text{else} \end{cases} \; . \tag{7.31}$$

and

$$\psi_2(q) = \begin{cases} 1/q^{\frac{1}{3}} & \text{for } q \in [-1,1] \\ \\ 0 & \text{else} \end{cases} \; . \tag{7.32}$$

i) Show that $\psi_1(q)$ and $\psi_2(q)$ are square integrable.

ii) Show that $q\ \psi_1(q)$ is not square integrable, i.e. that the operator $\hat{Q}$ takes $|\psi_1\rangle$ outside of the Hilbert space $\mathcal{H}$.

iii) Show that $-i\partial\psi_2(q)/\partial q$ is not square integrable, i.e. that the operator $\hat{P}$ takes $|\psi_2\rangle$ outside of the Hilbert space $\mathcal{H}$.

On the subset of states for which $\hat{Q}$ and $\hat{P}$ are defined they are indeed also Hermitian operators. This leads the two operators to have eigenvectors $|q_0\rangle$ and $|p_0\rangle$ with

$$\hat{Q}\,|q_0\rangle = q_0\ |q_0\rangle \quad , \quad \hat{P}\,|p_0\rangle = p\ |p_0\rangle \ . \tag{7.33}$$

E.g. the eigenvectors of $\hat{Q}$ would be the Dirac delta functions $q_0(q) = \delta_D(q - q_0)$. Of course, these are not part of the Hilbert space $\mathcal{H}$ itself, because they are not square integrable functions. But on the other hand, any actual element $|\psi\rangle \in \mathcal{H}$ can be decomposed into the $|q_0\rangle$ as

$$\psi(q) = \int \mathrm{d}q_0\ \psi(q_0)\ \delta_D(q - q_0) \ , \tag{7.34}$$

so we can nevertheless think of the $|q_0\rangle$ as a basis of $\mathcal{H}$. You can think of this situation as follows: we allow the operators $\hat{Q}$ and $\hat{P}$ to be ill-defined for some vectors in $\mathcal{H}$, and in return they can provide us with a continuous set of eigenvectors with a continuous spectrum of eigenvalues.[5]

In our new, infinite dimensional situation we will still consider 1-dimensional subspaces span($|\psi\rangle$) to represent elementary events. And states are again operators $\hat{\rho}$ that satisfy Equations 7.14, 7.15 and 7.16 . Given an ONB $\{|\psi_i\rangle\}$ of the Hilbert space $\mathcal{H}$ we can still think of such an operator as a matrix, and its element i-j would be given by

$$\rho_{ij} = \langle\psi_i|\,\hat{\rho}\,|\psi_j\rangle \ . \tag{7.35}$$

In particular, in the basis $|q\rangle$ of $\hat{Q}$ you can think of $\hat{\rho}$ as a function of two coordinates q_1 and q_2 as

$$\rho(q_1, q_2) \equiv \langle q_1|\,\hat{\rho}\,|q_2\rangle \ . \tag{7.36}$$

Since $\hat{\rho}$ is a state, this function will satisfy the properties

$$\rho(q_1, q_2) = \rho(q_2, q_1)^* \tag{7.37}$$

$$\int \mathrm{d}q_1\mathrm{d}q_2\ \psi(q_1)^*\rho(q_1, q_1)\psi(q_2) \geq 0 \quad \text{for all} \ \ |\psi\rangle \in \mathcal{H} \tag{7.38}$$

$$\int \mathrm{d}q\ \rho(q, q) = 1 \ . \tag{7.39}$$

Extending Equation 7.38 to the states $|q\rangle$ we can see that $\rho(q, q) \geq 0$. Together with Equation 7.39 this means that each state $\hat{\rho}$ defines a probability density function over the eigenvalue spectrum of $\hat{Q}$ via

$$p(q) = \rho(q, q) \ . \tag{7.40}$$

If $\hat{\rho}$ is a pure state, i.e. if it can be written as

$$\hat{\rho} = |\psi\rangle\,\langle\psi| \ , \tag{7.41}$$

[5] Furthermore, the set of states for which they are well defined is in fact dense in $\mathcal{H}$.

then this probability density is given by

$$p(q_0) = \langle q_0|\psi\rangle \langle\psi|q_0\rangle = |\psi(q_0)|^2 \ . \tag{7.42}$$

If instead $\hat{\rho}$ is a mixture of pure states $|\psi_1\rangle\langle\psi_1|$, ... , $|\psi_K\rangle\langle\psi_K|$ with mixture probabilities c_1 , ... , c_K then $p(q_0)$ is given by

$$p(q_0) = \sum_{i=1}^{K} c_i \ |\psi_i(q_0)|^2 \ . \tag{7.43}$$

To turn the concepts we have considered above into a description of quantum mechanics we still need to introduce a notion of time evolution. Given the Hamilton operator $\hat{H}$, i.e. the operator that represents the observable *energy* in our quantum system, any state $\hat{\rho}$ evolves according to the von-Neumann equation

$$\frac{\partial\hat{\rho}}{\partial t} = i\left[\hat{\rho}, \hat{H}\right] \ , \tag{7.44}$$

where $[\hat{A}, \hat{B}] = \hat{A}\hat{B} - \hat{B}\hat{A}$ is the commutator of two operators $\hat{A}$ and $\hat{B}$. For pure states, this equation is equivalent to the Schrödinger equation you may be more familiar with.

7.5 Superpositions and mixtures

Before we make our final step towards quantum field theory, let us reflect on an important aspect of quantum theory. Given an ONB $\{|\psi_i\rangle\}$ of $\mathcal{H}$, and a number of mixture probabilities $\{c_i\}$ we can form the state

$$\hat{\rho}_A = \sum_i c_i \ |\psi_i\rangle\langle\psi_i| \tag{7.45}$$

which is a mixture of the pure states $|\psi_i\rangle\langle\psi_i|$. We could also combine the vectors $\{|\psi_i\rangle\}$ in a different way, by first forming their superposition

$$|\psi\rangle = \sum_i \sqrt{c_i} \ |\psi_i\rangle \tag{7.46}$$

and then using that to form the (pure) state

$$\begin{aligned} \hat{\rho}_B = & \ |\psi\rangle\langle\psi| \\ = & \sum_{ij} \sqrt{c_i c_j}\, |\psi_i\rangle\langle\psi_j| \ . \end{aligned} \tag{7.47}$$

In the basis $\{|\psi_i\rangle\}$ both of these states - when considered as matrices - have exactly the same diagonal. In particular, the probability assigned to the events span($|\psi_i\rangle$) is c_i in both states. But it is of course a standard result of quantum mechanics that $\hat{\rho}_A$ and $\hat{\rho}_B$ describe very different situations!

To understand that, let us look at a simple example. Consider two vectors $|\psi_1\rangle$ and $|\psi_2\rangle$, which we take to represent two Gaussian wave packages with the same width and centered around two different locations q_1 and q_2 - cf. the blue lines in the lower two panels of Figure 19 for a graphical depiction. From those two vectors we can form the two states

$$\hat{\rho}_A = \frac{1}{2}\,|\psi_1\rangle\langle\psi_1| + \frac{1}{2}\,|\psi_2\rangle\langle\psi_2| \tag{7.48}$$

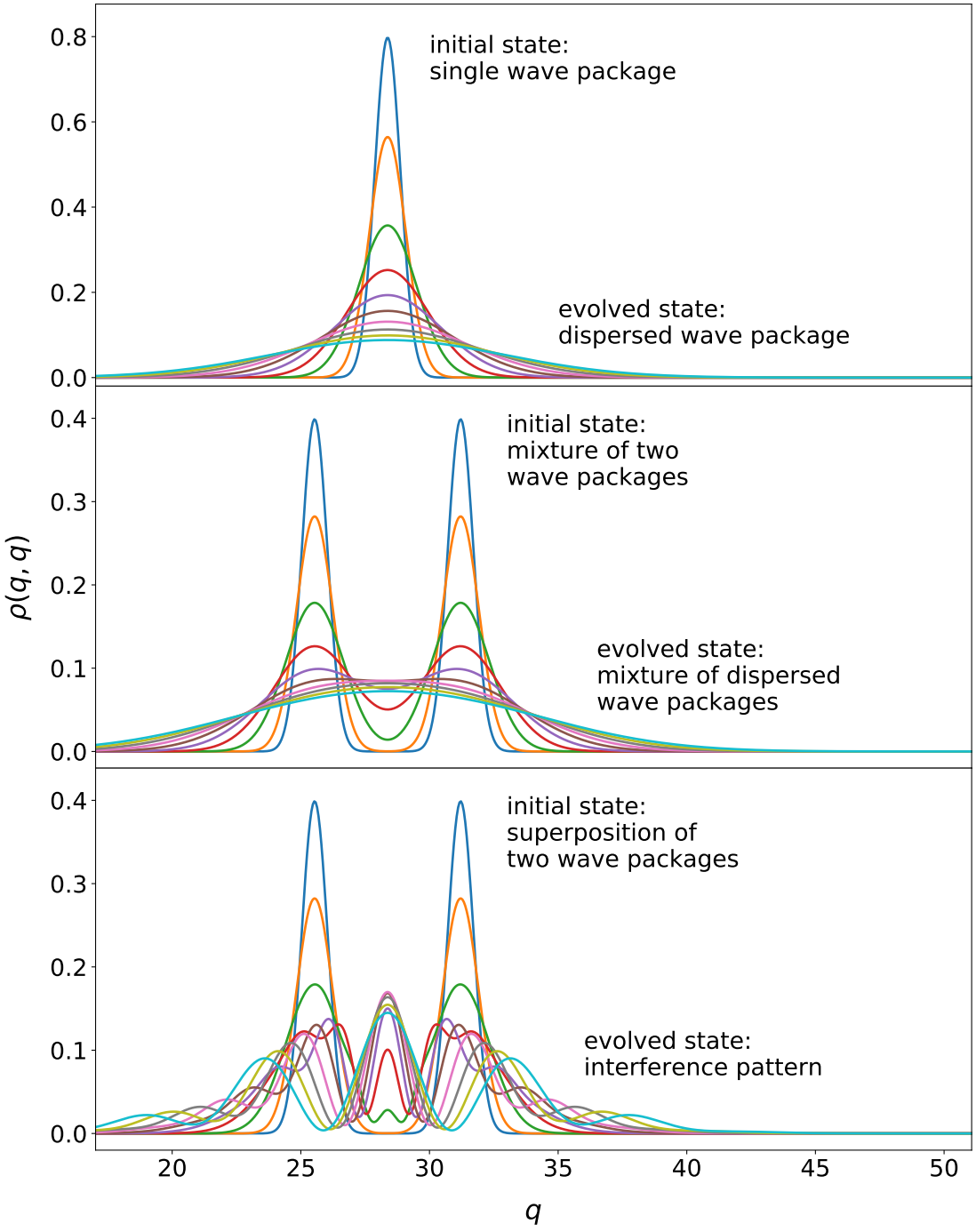

Figure 19. Time evolution of the PDF $\rho(q, q)$ for three different quantum states. The middle panel starts with the state $\hat{\rho}_A$ that was discussed in Section 7.5 and the lower panel starts with the state $\hat{\rho}_B$. The upper panel shows the evolution of a state corresponding to a single wave package.

and

$$\hat{\rho}_B = |\psi\rangle \langle\psi| \tag{7.49}$$

with

$$|\psi\rangle = \frac{1}{\sqrt{2}}|\psi_1\rangle + \frac{1}{\sqrt{2}}|\psi_2\rangle \ . \tag{7.50}$$

If the two wave packages are so far from each other that their overlap is negligible, then

$$\langle\psi_1|\hat{\rho}_A|\psi_1\rangle \approx \langle\psi_1|\hat{\rho}_B|\psi_1\rangle \approx \frac{1}{2} \tag{7.51}$$

and

$$\langle\psi_2|\hat{\rho}_A|\psi_2\rangle \approx \langle\psi_2|\hat{\rho}_B|\psi_2\rangle \approx \frac{1}{2} \ , \tag{7.52}$$

i.e. both states assign the same probability to the two wave packages. Let us now consider how our two states evolve with time. If our quantum theory describes a free particle with mass m, then the Hamiltonian which enters Equation 7.44 is

$$\hat{H} = \frac{1}{2m}\ \hat{P}^2 \ , \tag{7.53}$$

where $\hat{P}$ is the momentum operator. Let us assume that initially our wave packages don't have any momentum, i.e. $\langle\psi_1|\hat{P}|\psi_1\rangle = \langle\psi_2|\hat{P}|\psi_2\rangle = 0$. Then any individual wave package will simply disperse over time, i.e. it will become wider but stay centered around the same location. This is demonstrated by the differently coloured lines in the upper panel of Figure 19 which show the evolution of the PDF $\rho(q,q)$ for a state $\hat{\rho}$ representing an individual wave package.

The different lines in the middle panel of Figure 19 show the evolution of $\rho_A(q,q)$. For this mixture of two wave packages the evolution is very similar to that of the individual wave package. The two packages simply disperse, independently of each other. It is easy to understand from Equation 7.44 why this is the case: let us denote the two states $|\psi_1\rangle\langle\psi_1|$ and $|\psi_2\rangle\langle\psi_2|$ as $\hat{\rho}_1$ and $\hat{\rho}_2$. Then the von-Neumann equation for $\hat{\rho}_A$ - since it is linear in the state - can be written as

$$\frac{\partial\hat{\rho}_1}{\partial t} + \frac{\partial\hat{\rho}_2}{\partial t} = i[\hat{\rho}_1, H] + i[\hat{\rho}_2, H] \ . \tag{7.54}$$

This equation is solved by letting both wave packages $\hat{\rho}_1$ and $\hat{\rho}_2$ evolve independently according to their own von-Neumann equations. This is why a state like $\hat{\rho}_A$ is often said to exist of two independent *branches*. And an increasing portion of physicist takes this notion quite literally as part of the so called *many-worlds interpretation* of Quantum mechanics (see e.g. the poll reported by [47]). Of course, we have seen at the end of Section 7.2 that such a decomposition of a state into branches is not unique, and an additional criterion will be needed to determine which branches really constitute "independent worlds". We will examine this and other questions surrounding the many-worlds interpretation later in the course.

The evolution of $\hat{\rho}_B$ on the other hand is very different. You can see this in the lower panel of Figure 19 where we show how $\rho_B(q,q)$ changes over time. Now the two wave packages *do* notice each other and produce a strong interference pattern upon their contact. So in a purely quantum superposition the two wave packages can interact with each other and we cannot consider them as independent branches of the quantum state.

7.6 Step 3: Quantum field theory

In Section 7.4 we had defined elementary events to be 1-dimensional subspaces of a countably infinite dimensional Hilbert space $\mathcal{H}$ and states to be linear operators on $\mathcal{H}$ that satisfy Equations 7.14, 7.15 and 7.16 . Since all such Hilbert spaces are the same (they are all isomorphic

to each other) we could use the space of square integrable functions as a representation for $\mathcal{H}$ (modulo the technicality of identifying functions that differ only on null-sets). In that space, the position operator $\hat{Q}$ has a continuous eigenbasis $|q\rangle$. Within that basis each vector $|\psi\rangle \in \mathcal{H}$ can be represented be the function

$$\psi(q) = \langle q|\psi\rangle \tag{7.55}$$

and each state $\hat{\rho}$ can be represented by

$$\rho(q_1, q_2) = \langle q_1|\hat{\rho}|q_2\rangle \tag{7.56}$$

which you can think of as the element q_1-q_2 of the matrix $\hat{\rho}$, as expressed in the basis $|q\rangle$.

From here it is only a small step towards quantum field theory, which in spirit is very similar to the step from continuous random variables to random fields that we undertook in Section 2.2 . Sticking with real scalar fields for now, let $q(\boldsymbol{x})$ be a real function and let Ψ be a functional, such that

$$\Psi[q] \in \mathbb{C} \ .$$

We could attempt to define a notion of square integrability as

$$\int \mathcal{D}q \ |\Psi[q]|^2 < \infty \tag{7.57}$$

but as in our discussion of the PDFunctional of a Gaussian random field, we would encounter problems of properly normalising such integrals (cf. Exercise 22). Assuming that these technical problems can be solved, we can consider the Hilbert space $\mathcal{H}$ consisting of such functionals Ψ (which in a basis-free way we may denote as $|\Psi\rangle$). On that Hilbert space the field of operators $\hat{Q}(\boldsymbol{x})$ defined as

$$\left(\hat{Q}(\boldsymbol{x})\Psi\right)[q] = q(\boldsymbol{x})\Psi[q] \tag{7.58}$$

would play the role of "position" observables. In analogy to Section 7.4 the operator field $\hat{Q}$ will have a continuous set of eigenvectors $|q_0\rangle$ which we can think of as the "delta functionals"

$$q_0[q] = \prod_{\boldsymbol{x}} \delta_D(q(\boldsymbol{x}) - q_0(\boldsymbol{x})) \ . \tag{7.59}$$

As in Section 7.4 we could now take the sub-spaces span($|\Psi\rangle$) to represent elementary events, and states would be operators $\hat{\rho}$ satisfying appropriately modified versions of Equations 7.14, 7.15 and 7.16 . In the eigenbasis of $\hat{Q}$ we can again think of these operators as matrices with elements

$$\rho[q_1, q_2] = \langle q_1|\hat{\rho}|q_2\rangle \ . \tag{7.60}$$

And in terms of this basis Equations 7.14, 7.15 and 7.16 become

$$\rho[q_1, q_2] = \rho[q_2, q_1]^* \tag{7.61}$$

$$\int \mathcal{D}q_1\mathcal{D}q_2 \ \Psi[q_1]^*\rho[q_1, q_2]\Psi(q_2) \geq 0 \ \text{ for all } \ |\Psi\rangle \in \mathcal{H} \tag{7.62}$$

$$\int \mathcal{D}q \ \rho[q, q] = 1 \ . \tag{7.63}$$

As before, it follows from Equation 7.62 that $\rho[q,q] \geq 0$ and together with Equation 7.63 this means that the functional

$$\mathcal{P}[q] = \rho[q,q] \tag{7.64}$$

defines a PDFunctional over the functions $q(\boldsymbol{x})$.

How will all of this connect back to the first part of this course? As we will see in upcomming lectures, the density contrast field δ turns into a quantum field in the early Universe. Instead of a PDFunctional $\mathcal{P}[\delta]$ we will then need a state matrix $\rho[\delta_1, \delta_2]$ to describe its statistical properties. Of course, the diagonal $\rho[\delta, \delta]$ will again define a PDFunctional for δ and in a - yet to be explained - sense this diagonal turns into the classical PDFunctional of δ which describes the late-time cosmic density field. This transition from quantum to classical probabilities will be related to our discussion of independent branches of quantum states in Section 7.5 - somehow our quantum state $\hat{\rho}$ will evolve into a state that can approximately be seen as a mixture of states $|\delta\rangle \langle\delta|$ in which different configurations of δ do not interfere with each other any more.

Let us close this section by drawing another parallel with our previous treatment of classical random variables. In Section 2.2 we said that random fields $f(\boldsymbol{x})$ can be seen as collections of continuous random variables - one for each $\boldsymbol{x}$. We can understand the above situation in a similar way: the quantum system described by the Hilbert space $\mathcal{H}$ of functionals Ψ is a collection of quantum systems - one for each $\boldsymbol{x}$ with Hilbert space $\mathcal{H}_{\boldsymbol{x}}$. To combine a set of quantum systems, one needs to consider the tensor product of the Hilbert spaces of the individual systems. Hence, you can think of the Hilbert space $\mathcal{H}$ as the tensor product of all the Hilbert spaces $\mathcal{H}_{\boldsymbol{x}}$ (cf. Exercise 53 for a further exploration of this). This of course makes it sound as if the step from quantum mechanics to quantum field theory is trivial. But as a (fatal) philosopher once said - sometimes a higher level of quantity also leads to a different level of quality. The dimension of the Hilbert space $\mathcal{H}$ is uncountably infinite. This makes quantum field theory a very powerful tool to describe reality, and it can also bring with it a number of technical difficulties that can only be addressed by highly advanced methodology. At the same time, there are actually reasons to believe that the Hilbert space of the entire observable Universe is rendered countable (and even finite) by quantum gravitational effects [see e.g. 2, 17, 28]. But this is beyond the scope of this course.

Exercise 53

In Section 2.2 we had made an intermediate step when moving from random variables to random fields: random vectors. Let us try something similar in the quantum context. In Section 7.4 we have only considered square integrable functions ψ of one variable q, and their corresponding Hilbert space $\mathcal{H}$. Show that the space of square integrable functions $\psi(q_1, q_2)$ of two variables, which we shall denote by $\mathcal{H}^2$, can be seen as the tensor product of two copies of the one-variable Hilbert space $\mathcal{H}$, i.e.

$$\mathcal{H}^2 = \mathcal{H} \otimes \mathcal{H} \ . \tag{7.65}$$

Hint: Let the functions $\{\psi_i\}$ be an ONB for $\mathcal{H}$. It is sufficient to show that the functions

$$\phi_{ij}(q_1, q_2) = \psi_i(q_1)\psi_j(q_2) \tag{7.66}$$

are an ONB for $\mathcal{H}^2$.

8 The early-time density contrast as a quantum field

8.1 A weird early Universe: $\bar{p} \approx -\bar{\rho}$ & $\delta p \approx +\delta\rho$

To investigate the quantum origin of cosmic density perturbations we have to go back in time until shortly after the big bang. We said in Section 1.3 that at such early times the evolution of the density contrast field δ is well described by the linearized versions of the continuity equation and the Euler equation. This even lead to a single, closed form equation for the evolution of δ, cf. Equation 1.28. One of the terms we ignored in order to arrive at that equation was the pressure term in the Euler equation (Equation 1.9). But to encounter quantum effects in the density field we have to go back to such early times that this is not a good approximation. In fact, in the very early Universe the cosmic energy density was completely dominated by a plasma of highly relativistic particles. In the limit where all particles move close to the speed of light we can think of this plasma as a gas of radiation (i.e. of mass-less particles like photons). For such a gas the density ρ and pressure p are connected via an equation of state of the form

$$p(\boldsymbol{x}, t) = \frac{1}{3}\, \rho(\boldsymbol{x}, t) \ . \tag{8.1}$$

Note that because of $E = mc^2$ we can think of $\hat{\rho}$ both as a matter density and an energy density, so Equation 8.1 is a relation between the total energy density of the relativistic gas and its pressure.

But to meet the first, quantum density fluctuations in the Universe, we have to move to even early times. According to the current standard paradigm of cosmology there was a phase before the radiation-dominated epoch of the Universe in which the cosmic density field displayed a rather peculiar behaviour. If we parameterize both pressure and energy density in terms of background quantities and perturbations as

$$p(\boldsymbol{x}, t) = \bar{p}(t) + \delta p(\boldsymbol{x}, t) \tag{8.2}$$

$$\rho(\boldsymbol{x}, t) = \bar{\rho}(t) + \delta\rho(\boldsymbol{x}, t) \tag{8.3}$$

then the background pressure and background density where related via the equation of state

$$\bar{p}(t) \approx -\bar{\rho}(t) \ . \tag{8.4}$$

This is already very strange, because it means that the fluid that filled this very early Universe had a negative pressure! But the curiosities don't stop here, because in contrast to the background quantities the pressure and density perturbations were related by

$$\delta p(\boldsymbol{x}, t) \approx +\delta\rho(\boldsymbol{x}, t) \ . \tag{8.5}$$

The theory from which these relations follow is the paradigm of the so called *cosmic inflation*. In principle, Equations 8.4 and 8.5 are everything we will need in order to characterise the quantum fluctuations in the early Universe. But let us nevertheless look a bit behind the scenes to see where these equations are coming from.

According to the simplest versions of inflation, the energy density of the early Universe is mostly constituted by a scalar field $\phi(\boldsymbol{x}, t)$ - the so called inflaton field. If g_{ab} are the coefficients of the spacetime metric (with g^{ab} being the coefficients of the inverse metric), and defining the auxiliary variable

$$X \equiv \frac{1}{2} g^{ab} \partial_a \phi \partial_b \phi \equiv \frac{1}{2} \sum_{a,b=0}^{3} g^{ab} \frac{\partial \phi}{\partial x^a} \frac{\partial \phi}{\partial x^b} \ , \tag{8.6}$$

then it can be shown that the inflaton field acts as a fluid with pressure

$$p(\boldsymbol{x},t) = X(\boldsymbol{x},t) - V(\phi(\boldsymbol{x},t)) \tag{8.7}$$

and energy density

$$\rho(\boldsymbol{x},t) = X(\boldsymbol{x},t) + V(\phi(\boldsymbol{x},t)) \ . \tag{8.8}$$

Here $V(\phi)$ is a function quantifying the potential energy of the inflaton field (and you can think of X as its kinetic energy). Obviously, to achieve $p \approx -\rho$ we need to have that

$$V(\phi) \gg X \ . \tag{8.9}$$

So let us assume that 8.9 indeed holds. But how can we at the same time achieve $\delta p \approx +\delta\rho$? To see this, let us also split the scalar field ϕ and the auxiliary variable X into a background contributions and perturbations as

$$\phi(\boldsymbol{x},t) = \bar{\phi}(t) + \delta\phi(\boldsymbol{x},t) \tag{8.10}$$

$$X(\boldsymbol{x},t) = \bar{X}(t) + \delta X(\boldsymbol{x},t) \ . \tag{8.11}$$

For small perturbations we then have

$$\delta p \approx \delta X - \left.\frac{\partial V}{\partial \phi}\right|_{\bar{\phi}} \delta\phi \tag{8.12}$$

$$\delta \rho \approx \delta X + \left.\frac{\partial V}{\partial \phi}\right|_{\bar{\phi}} \delta\phi \ . \tag{8.13}$$

For for the pressure and density perturbations to equal we need that

$$\delta X \gg \left.\frac{\partial V}{\partial \phi}\right|_{\bar{\phi}} \delta\phi \ . \tag{8.14}$$

Of course, both δX and $\delta\phi$ will be (draws from) random fields, so they can take all sorts of values and it will be hard to satisfy Equation 8.14 in full generality. But it will be satisfied for most of the time if we demand that

$$\left.\frac{\partial V}{\partial \phi}\right|_{\bar{\phi}} \ll \frac{\sqrt{\langle \delta\phi^2 \rangle}}{\sqrt{\langle \delta X^2 \rangle}} \ . \tag{8.15}$$

Equation 8.9 and Equation 8.15 are a (somewhat unusual) version of what is called the *slow roll conditions* in the inflationary paradigm. This name comes from the fact that we can interpret those conditions as demanding that the scalar field "rolls" down a very high potential (Equation 8.9) which at the same time is very shallow (Equation 8.15) such that the rolling proceeds in a very slow manner. In the end, these conditions are just an elegant way to achieve $\bar{p} \approx -\bar{\rho}$ and $\delta p \approx +\delta\rho$, and we will see later, why these conditions are crucial for the early density fluctuations.

Before moving on, let us still investigate the impact of $\bar{p} \approx -\bar{\rho}$ on the background expansion of the Universe. In Section 1 we had derived the second Friedmann equation in the absence of pressure (cf. Equation 1.7). Let us quote here without further derivations that the Friedmann equations in a spatially flat Universe with pressure (and setting $c = 1 = G$) are

$$H^2 = \frac{8\pi}{3}\,\bar{\rho} + \frac{\Lambda}{3} \tag{8.16}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi}{3}\,(\bar{\rho} + 3\bar{p}) + \frac{\Lambda}{3} \ , \tag{8.17}$$

where $H = \dot{a}/a$ is the expansion rate of the Universe (the *Hubble rate*) and Λ is the cosmological constant.

Exercise 54

Show that

$$\dot{H} = \frac{\ddot{a}}{a} - H^2 \ . \tag{8.18}$$

Using that result together with the Friedmann equations, show that

$$\dot{H} = -4\pi(\bar{\rho} + \bar{p}) \ . \tag{8.19}$$

From the Friedmann equations and from your result in Exercise 54 we can draw the following conclusions for the situation when $\bar{p} \approx -\bar{\rho}$:

- The expansion rate H will approximately stay constant in time. In particular, the Hubble horizon which is given by $1/H$ will be constant.
- Because of the first Friedmann equation (Equation 8.16) a constant H will also lead to a constant density ρ .
- The expansion of the Universe will be accelerating, i.e. $\ddot{a} > 0$, because the right-hand side of Equation 8.17 will be positive.

Especially the first of these conclusions will become important in the following.

8.2 An action for the density contrast field - first attempt

So how are we going to build a quantum theory of the early cosmic density fluctuations? The simple answer is that we will make the density field look like a collection of harmonic oscillators. And then we will replace each of these oscillators by a quantum harmonic oscillator. As a starting point in that program, let us consider the simple situation when the Universe is filled with dust (i.e. when the pressure p vanishes) and where the cosmological constant is 0. Let us furthermore assume that the cosmic velocity field is a potential flow. This means that there is a scalar field $\theta(\boldsymbol{r}, t)$, such that $\mathbf{v}(\boldsymbol{r}, t)$ is given by

$$\mathbf{v} = \boldsymbol{\nabla}_{\boldsymbol{r}}\theta \ . \tag{8.20}$$

This is a very good approximation in the early Universe, because the rapid expansion during the epoch of inflation dilutes all velocity contributions which can not be expressed as the divergence of a scalar.

In this simplified scenario the continuity equation, the Euler equation and the Poisson equation can be shown to follow from an action principle, with the action

$$S[\rho, \theta, \phi] = -\int \mathrm{d}t\mathrm{d}^3r \ \left\{ \rho\dot{\theta} + \frac{\rho}{2}(\boldsymbol{\nabla}_{\boldsymbol{r}}\theta)^2 + \rho\phi + \frac{1}{8\pi}(\boldsymbol{\nabla}_{\boldsymbol{r}}\phi)^2 \right\} \ . \tag{8.21}$$

Varying this action wrt. the gravitational potential ϕ yields the Poisson equation (cf. Equation 1.14)

$$\Delta_{\boldsymbol{r}}\phi = 4\pi G\rho \tag{8.22}$$

which we can symbolically invert as

$$\phi = \frac{4\pi G}{\Delta}\rho \ . \tag{8.23}$$

Using this, we can reduce the above action to

$$S[\rho,\theta] = -\int \mathrm{d}t\mathrm{d}^3r \ \left\{\rho\dot{\theta} + \frac{\rho}{2}(\boldsymbol{\nabla}\theta)^2 + \frac{1}{2}\rho\frac{4\pi G}{\Delta}\rho\right\} \ . \tag{8.24}$$

Exercise 55

Show that the continuity equation (Equation 1.8) follows from the above action, by demanding that

$$\frac{\delta S}{\delta\theta} = 0 \tag{8.25}$$

for physical fields ρ and θ .

Exercise 56

Not graded: *Show that the Euler equation (Equation 1.9) follows from demanding that*

$$\frac{\delta S}{\delta\rho} = 0 \tag{8.26}$$

for physical fields ρ and θ . Hint: you can read the integral

$$\int \mathrm{d}^3r \ \rho\left(\frac{4\pi G}{\Delta}\rho\right) \tag{8.27}$$

as a product of the form $\boldsymbol{w}^T(\boldsymbol{M}\boldsymbol{w})$ with ρ playing the role of the vector $\boldsymbol{w}$ and $\frac{4\pi G}{\Delta}$ playing the role of the matrix $\boldsymbol{M}$. In particular, this matrix is symmetric. If you feel like it, you can analytically write down what the inverse of the operator Δ is (e.g. using the Fourier transform) to make this analogy more apparent.

S is the action for the total density field ρ and the total velocity potential θ . But we only want to quantize the perturbations of these fields wrt. the background Hubble flow. So let us split the two fields into background quantities and perturbations as

$$\rho(\boldsymbol{r},t) = \bar{\rho}(t) + \delta\rho(\boldsymbol{r},t) \tag{8.28}$$
$$\theta(\boldsymbol{r},t) = \bar{\theta}(\boldsymbol{r},t) + \vartheta(\boldsymbol{r},t) \ . \tag{8.29}$$

Exercise 57

Show that $\bar{\theta}(\boldsymbol{r},t) = H(t)\boldsymbol{r}^2/2$, which is a scalar version of Hubble's law.

Of course, the background quantities $\bar{\rho}$ and $\bar{\theta}$ already satisfy the continuity equation and Euler equation. So they must be at least a local minimum of the action S . Because of that,

in the Taylor expansion

$$
\begin{aligned}
S[\rho,\theta] \approx S[\bar{\rho},\bar{\theta}] &+ \int \mathrm{d}t\mathrm{d}^3r \left(\frac{\not{d} S}{\not{d}\rho(t,\boldsymbol{r})}\bigg|_{\bar{\rho},\bar{\theta}} \delta\rho(t,\boldsymbol{r}) \right) + \int \mathrm{d}t\mathrm{d}^3r \left(\frac{\not{d} S}{\not{d}\theta(t,\boldsymbol{r})}\bigg|_{\bar{\rho},\bar{\theta}} \vartheta(t,\boldsymbol{r}) \right) \\
&+ \frac{1}{2}\int \mathrm{d}t_1\mathrm{d}t_2\mathrm{d}^3r_1\mathrm{d}^3r_2 \left(\delta\rho(t_1,\boldsymbol{r}_1) \frac{\not{d}^2 S}{\not{d}\rho(t_1,\boldsymbol{r}_1)\not{d}\rho(t_2,\boldsymbol{r}_2)}\bigg|_{\bar{\rho},\bar{\theta}} \delta\rho(t_2,\boldsymbol{r}_2) \right) \\
&+ \frac{1}{2}\int \mathrm{d}t_1\mathrm{d}t_2\mathrm{d}^3r_1\mathrm{d}^3r_2 \left(\vartheta(t_1,\boldsymbol{r}_1) \frac{\not{d}^2 S}{\not{d}\theta(t_1,\boldsymbol{r}_1)\not{d}\theta(t_2,\boldsymbol{r}_2)}\bigg|_{\bar{\rho},\bar{\theta}} \vartheta(t_2,\boldsymbol{r}_2) \right) \\
&+ \int \mathrm{d}t_1\mathrm{d}t_2\mathrm{d}^3r_1\mathrm{d}^3r_2 \left(\delta\rho(t_1,\boldsymbol{r}_1) \frac{\not{d}^2 S}{\not{d}\rho(t_1,\boldsymbol{r}_1)\not{d}\theta(t_2,\boldsymbol{r}_2)}\bigg|_{\bar{\rho},\bar{\theta}} \vartheta(t_2,\boldsymbol{r}_2) \right) \qquad (8.30)
\end{aligned}
$$

the terms that are linear in the perturbations $\delta\rho$ and ϑ must vanish. Also, as far as the perturbations are concerned, the background contribution $S[\bar{\rho},\bar{\theta}]$ is just an irrelevant constant term in the action. So we can consider the quadratic terms in Equation 8.30 as an action for the perturbations - the so called *quadratic action*, which we will denote with $S_2[\delta\rho,\vartheta]$.

Exercise 58

Assume that the evolution of some fields is governed by an action that is quadratic in those fields. What can you conclude from this for the equations of motion of those fields?

To evaluate the quadratic action, we need to compute the second order functional derivatives of S at the background fields. These derivatives are given by

$$
\begin{aligned}
\frac{\not{d}^2 S}{\not{d}\rho(\boldsymbol{r}_1,t_1)\not{d}\rho(\boldsymbol{r}_2,t_2)}\bigg|_{\bar{\rho},\bar{\theta}} &= -\delta_D(t_1-t_2)\frac{4\pi G}{\Delta}(\boldsymbol{r}_1,\boldsymbol{r}_2) \qquad (8.31) \\
\frac{\not{d}^2 S}{\not{d}\theta(\boldsymbol{r}_1,t_1)\not{d}\theta(\boldsymbol{r}_2,t_2)}\bigg|_{\bar{\rho},\bar{\theta}} &= \delta_D(t_1-t_2)\delta_D^3(\boldsymbol{r}_1-\boldsymbol{r}_2)\bar{\rho}(t_1)\Delta \qquad (8.32) \\
\frac{\not{d}^2 S}{\not{d}\rho(\boldsymbol{r}_1,t_1)\not{d}\theta(\boldsymbol{r}_2,t_2)}\bigg|_{\bar{\rho},\bar{\theta}} &= \delta_D(t_1-t_2)\delta_D^3(\boldsymbol{r}_1-\boldsymbol{r}_2)\left\{\frac{\partial}{\partial t} - (\boldsymbol{\nabla}\bar{\theta})\boldsymbol{\nabla}\right\} \\
&= \delta_D(t_1-t_2)\delta_D^3(\boldsymbol{r}_1-\boldsymbol{r}_2)\left\{\frac{\partial}{\partial t} + H\boldsymbol{r}\boldsymbol{\nabla}\right\} . \qquad (8.33)
\end{aligned}
$$

Here we have again viewed $4\pi G/\Delta$ as a matrix. Also, all differential operators are supposed to act on $\boldsymbol{r}_2$.

With the above derivatives the quadratic action becomes

$$
\Rightarrow S_2[\delta\rho,\vartheta] = -\int \mathrm{d}t\mathrm{d}^3r \left\{ \delta\rho\dot{\vartheta} + \delta\rho H\boldsymbol{r}\boldsymbol{\nabla}_{\boldsymbol{r}}\vartheta + \frac{\bar{\rho}}{2}(\boldsymbol{\nabla}_{\boldsymbol{r}}\vartheta)^2 + \frac{1}{2}\delta\rho\frac{4\pi G}{\Delta_{\boldsymbol{r}}}\delta\rho \right\} . \qquad (8.34)
$$

It will be convenient for us to switch from physical coordinates $\boldsymbol{r}$ to co-moving coordinates $\boldsymbol{x} = \boldsymbol{r}/a$ now, because of the result of following exercise.

Exercise 59

Show that

$$
\frac{\partial}{\partial t}\bigg|_{\boldsymbol{x}=\text{const.}} = \frac{\partial}{\partial t}\bigg|_{\boldsymbol{r}=\text{const.}} + H\boldsymbol{r}\boldsymbol{\nabla}_{\boldsymbol{r}} . \qquad (8.35)
$$

(This is very similar to what you already calculated in Exercise 5.)

So switching to $\boldsymbol{x}$ and denoting with $\dot{\vartheta}$ from now on the time derivative of $\vartheta(\boldsymbol{x}, t)$ for constant $\boldsymbol{x}$, the quadratic action becomes

$$S_2[\delta\rho, \vartheta] = -\int \mathrm{d}t\mathrm{d}^3x\ a^3\ \left\{\delta\rho\dot{\vartheta} + \frac{\bar{\rho}}{2a^2}(\boldsymbol{\nabla}_{\boldsymbol{x}}\vartheta)^2 + \frac{1}{2}\delta\rho\frac{4\pi G a^2}{\Delta_{\boldsymbol{x}}}\delta\rho\right\}\ . \tag{8.36}$$

Let us now relabel the fields $\delta\rho$ and ϑ in a somewhat suggestive manner as

$$Q \equiv a^3\delta\rho\ \ ,\ \ \ P \equiv \vartheta \tag{8.37}$$

such that the action becomes

$$\begin{aligned} S_2[\delta\rho, \vartheta] &= -\int \mathrm{d}t\mathrm{d}^3x\ \left\{Q\dot{P} + \frac{a\bar{\rho}}{2}(\boldsymbol{\nabla}_{\boldsymbol{x}}P)^2 + \frac{1}{2}Q\frac{4\pi G}{a\Delta_{\boldsymbol{x}}}Q\right\} \\ &= \int \mathrm{d}t\mathrm{d}^3x\ \left\{\dot{Q}P - \frac{a\bar{\rho}}{2}(\boldsymbol{\nabla}_{\boldsymbol{x}}P)^2 - \frac{1}{2}Q\frac{4\pi G}{a\Delta_{\boldsymbol{x}}}Q\right\}\ . \end{aligned} \tag{8.38}$$

From this action we can read of the Lagrangian of the new fields Q and P as

$$L[Q, \partial Q, P, \partial P, t] = \int \mathrm{d}^3x\ \left\{\dot{Q}P - \frac{a\bar{\rho}}{2}(\boldsymbol{\nabla}_{\boldsymbol{x}}P)^2 - \frac{1}{2}Q\frac{4\pi G}{a\Delta_{\boldsymbol{x}}}Q\right\}\ . \tag{8.39}$$

From the fact that

$$P = \frac{\mathrm{d}L}{\mathrm{d}\dot{Q}} \tag{8.40}$$

it is clear that P is the canonically conjugate momentum of Q . So the Hamiltonian that corresponds to the above Lagrangian is given by

$$\begin{aligned} H[Q, P, t] &= \int \mathrm{d}^3x\ \left(\dot{Q}\frac{\mathrm{d}L}{\mathrm{d}\dot{Q}} + \dot{P}\frac{\mathrm{d}L}{\mathrm{d}\dot{P}}\right) - L \\ &= \int \mathrm{d}^3x\ \left\{\frac{a\bar{\rho}}{2}(\boldsymbol{\nabla}_{\boldsymbol{x}}P)^2 + \frac{1}{2}Q\frac{4\pi G}{a\Delta_{\boldsymbol{x}}}Q\right\}\ . \end{aligned} \tag{8.41}$$

Exercise 60

Show that the second line of Equation 8.41 is correct.

Does this Hamiltonian look like a sum of Hamiltonians of harmonic oscillators? At least we are summing over terms that are quadratic in both the "position" Q and the "momentum" P, which is exactly the structure the the Hamiltonian of harmonic oscillators. But to get a more definite answer let us transform the above expression to Fourier space, where

$$H[\tilde{Q}, \tilde{P}, t] = \int \frac{\mathrm{d}^3k}{(2\pi)^3}\ \left\{\frac{a\bar{\rho}k^2}{2}|\tilde{P}|^2 - \frac{1}{2}\frac{4\pi G}{ak^2}|\tilde{Q}|^2\right\}\ . \tag{8.42}$$

Comparing this to the standard Hamiltonian of a harmonic oscillator,

$$H_{\mathrm{osc}} = \frac{1}{2M}P^2 + \frac{M\Omega^2}{2}Q^2\ , \tag{8.43}$$

we could attempt to identify

$$M_{\boldsymbol{k}}(t) = \frac{1}{a(t)\bar{\rho}(t)k^2} \tag{8.44}$$

$$\Omega(t)^2 = -4\pi G\bar{\rho}(t)\ . \tag{8.45}$$

But this would of course mean that Ω^2 is negative for our oscillators! In particular, our Hamiltonian is not bounded from below, so in a quantum version of our oscillators there would be no stable ground state and the density field would constantly decay into lower and lower energy states. There is actually some physical significance to this. The gravitational potential is indeed not bounded from below, such that there is - in the absence of pressure - indeed not stable, gravitationally bound object. This is for example the reason why globular star clusters only have a finite (albeit very long) life time[6].

8.3 An action for the density field: inside the Hubble horizon

So our first attempt to obtain a version of hydrodynamics that we can quantize has failed. But this may of course be because we have neglected pressure in the above. To see how the presence of a non-zero pressure p modifies our derivations we need to slightly adjust our previous versions of the Euler equation and the continuity equation. We had previously written the continuity equation as (cf. Equation 1.8)

$$\left.\frac{\partial\rho}{\partial t}\right|_{\boldsymbol{r}=\text{const.}} + \boldsymbol{\nabla}_{\boldsymbol{r}}(\rho\mathbf{v}) = 0 \ , \tag{8.46}$$

and if we switch to co-moving coordinates $\boldsymbol{x}$ and take time derivatives at constant $\boldsymbol{x}$ this becomes

$$\left.\frac{\partial\rho}{\partial t}\right|_{\boldsymbol{x}=\text{const.}} - H\boldsymbol{x}\boldsymbol{\nabla}_{\boldsymbol{x}}\rho + \frac{1}{a}\boldsymbol{\nabla}_{\boldsymbol{x}}(\rho\mathbf{v}) = 0 \ . \tag{8.47}$$

At linear order in the perturbations $\delta\rho$ and ϑ this becomes

$$\dot{\delta\rho} + 3H\delta\rho + \frac{\bar{\rho}}{a^2}\Delta_{\boldsymbol{x}}\vartheta = 0 \ . \tag{8.48}$$

In Section 1 we had interpreted Equation 8.47 as an equation of mass conservation: the amount by which the density changes over time at some location has to be equal to (-1) times the outflow of mass from an infinitesimal volume around that location. However, in the present section we have started to interpret ρ as the total energy density of a fluid with non-negligible pressure p.

Exercise 61

If ρ represents the total energy density of a fluid with non-negligible pressure (i.e. if it is the sum of heat density and c^2 times mass density), why does Equation 8.47 need to be modified?

If we think of the Universe as an expanding gas cloud, then the answer is: In the presence of pressure the expansion of the gas is performing work, and this will reduce the heat energy in the gas (it will cool the gas). To consistently workout how this changes the continuity equation we would need to resort to a fully relativistic treatment. In particular, we would have to consider the energy-momentum tensor $T^{\mu\nu}$ of the fluid within the equation

$$\nabla_\mu T^{\mu 0} = 0 \ , \tag{8.49}$$

[6] In rare, close encounters of three stars of such a cluster, two of the stars will form a close binary system. Because of energy conservation, this may leave the third star with so much kinetic energy that it gets shot out of the cluster. Because the gravitational potential is unbound from below, having binary systems become closer and closer will release more and more kinetic energy and eventually evaporate the star cluster.

where ∇_μ denotes covariant derivative wrt. the spacetime metric. We will not attempt such a derivation here, and instead I will just tell you, that the above equation at linear order in the perturbations $\delta\rho$ and ϑ becomes

$$\dot{\delta\rho} + 3(1+\bar{w})H\delta\rho + \frac{(1+\bar{w})\bar{\rho}}{a^2}\Delta_{\boldsymbol{x}}\vartheta = 0 \ . \tag{8.50}$$

So all that changes wrt. Equation 8.48 is the appearance of an additional $\bar{p}$, and you can roughly think of this new term as the work performed by our fluid during the expansion of the Universe. Of course, we had said in Section 8.1 that $\bar{p} \approx \bar{\rho}$, so the last term in the above equation seems to vanish. But this is not entirely true, as you will explore in the following exercise.

Exercise 62

Let us define the equation-of-state parameter $\bar{w}$ of the background density field as

$$\bar{p} = \bar{w}\bar{\rho} \ . \tag{8.51}$$

Can $\bar{w}$ in the simple model of Section 8.1 ever be smaller than (-1)? If it was exactly (-1) , could the epoch of inflation ever end? Hint: recall the bullet points below Exercise 54.

So let us in the following assume that $\bar{w} \gtrsim -1$. And as a further simplification, let as assume that

$$\frac{\partial\bar{w}}{\partial t} \approx 0 \ , \tag{8.52}$$

which will certainly be satisfied if inflation is sufficiently "slow-roll".

The presence of a non-negligible pressure will not just modify the continuity equation but also the Euler equation. At linear order in the perturbations, and using physical time t as opposed to conformal time, our previous version of the Euler equation (Equation 1.27) becomes

$$\frac{\partial\boldsymbol{v}}{\partial t} + H\boldsymbol{v} + \frac{1}{a}\boldsymbol{\nabla}_{\boldsymbol{x}}\varphi = 0 \ . \tag{8.53}$$

To workout how to modify this in the presence of pressure we would again have to resort to a fully relativist treatment and consider the equation

$$\nabla_\mu T^{\mu i} = 0 \ . \tag{8.54}$$

We will not attempt this, and I will instead simply quote the result which is

$$\frac{\partial\boldsymbol{v}}{\partial t} + (1-3\bar{w})H\boldsymbol{v} + \frac{1}{a}\boldsymbol{\nabla}_{\boldsymbol{x}}\varphi + \frac{1}{a}\frac{\boldsymbol{\nabla}\delta p}{(1+\bar{w})\bar{\rho}} = 0 \ . \tag{8.55}$$

The last term on the left-hand side of this equation simply states that the pressure gradient contributes to the force acting on a fluid element (cf. the full Euler Equation 1.9). And the presence of $\bar{w}$ in this equations is due to relativistic effects.

The coefficient $\bar{w}$ characterises the relation between the background quantities $\bar{\rho}$ and $\bar{p}$. In Section 8.1 we allowed the relation between the perturbations $\delta\rho$ and δp to be different from that. So let's introduce a separate equation-of-state parameter for the perturbations as

$$\delta p = w_\delta \ \delta\rho \tag{8.56}$$

such that the perturbed Euler equation becomes

$$\frac{\partial \boldsymbol{v}}{\partial t} + (1-3\bar{w})H\boldsymbol{v} + \frac{1}{a}\boldsymbol{\nabla}_{\boldsymbol{x}}\varphi + \frac{1}{a}\frac{w_\delta \boldsymbol{\nabla}\delta\rho}{(1+\bar{w})\bar{\rho}} = 0 \ . \tag{8.57}$$

If we further more take into account that the velocity perturbation $\boldsymbol{v}$ is related to the velocity potential ϑ via

$$\boldsymbol{v} = \frac{1}{a}\boldsymbol{\nabla}\vartheta \tag{8.58}$$

then we can integrate Equation 8.57 to obtain

$$\frac{\partial \vartheta}{\partial t} - 3\bar{w}H\vartheta + \varphi + \frac{w_\delta \ \delta\rho}{(1+\bar{w})\bar{\rho}} = \text{const.} \ . \tag{8.59}$$

Equation 8.59 and Equation 8.50 are our new, relativistic versions of the Euler and the continuity equation. **Note however, that in Equation 8.50 we have ignored some terms that are only relevant on distance scales comparable to the Hubble horizon $1/H$ of the Universe** (hence the name of this subsection). We will consider these corrections in the next subsection, and our current derivations will be valid only for scales $k \gg H$.

Both Equation 8.59 and Equation 8.50 are reproduced by the action

$$S_2[\delta\rho,\vartheta] = -\int \mathrm{d}t\mathrm{d}^3x \ a^3 \left\{ \delta\rho\dot{\vartheta} - 3\bar{w}H\vartheta\delta\rho + \frac{(1+\bar{w})\bar{\rho}}{2a^2}(\boldsymbol{\nabla}_{\boldsymbol{x}}\vartheta)^2 + \frac{1}{2}\delta\rho\frac{4\pi a^2}{\Delta_{\boldsymbol{X}}}\delta\rho + \frac{1}{2}\frac{w_\delta}{1+\bar{w}}\frac{\delta\rho^2}{\bar{\rho}} \right\} \ . \tag{8.60}$$

Exercise 63

Not graded: *Building on your results from Exercise 56 show that the above, new version of S_2 indeed generates our new, modified continuity and Euler equations.*

Changing our definition of the variables Q and P to

$$Q \equiv a^{3(1+\bar{w})}\delta\rho \tag{8.61}$$

$$P \equiv a^{-3\bar{w}}\vartheta \tag{8.62}$$

this action becomes

$$\begin{aligned} &S_2[Q,\partial Q,P,\partial P] \\ &= \int \mathrm{d}t\mathrm{d}^3x \ \left\{ \dot{Q}P - \frac{a^{1+6\bar{w}}(1+\bar{w})\bar{\rho}}{2}(\boldsymbol{\nabla}_{\boldsymbol{x}}P)^2 - \frac{1}{2}Q\frac{4\pi}{a^{1+6\bar{w}}\Delta_{\boldsymbol{x}}}Q - \frac{1}{2}\frac{w_\delta}{1+\bar{w}}\frac{Q^2}{a^{3+6\bar{w}}\bar{\rho}} \right\} \ . \end{aligned} \tag{8.63}$$

Exercise 64

Show that Equation 8.63 indeed follows from Equation 8.60 via our new definitions of Q and P. Hint: start by writing Equation 8.60 in terms of Q and ϑ and then perform integration by parts.

From the action S_2 we can read of the Lagrangian to be

$$\begin{aligned} &L[Q,\partial Q,P,\partial P,t] \\ &= \int \mathrm{d}^3x \ \left\{ \dot{Q}P - \frac{a^{1+6\bar{w}}(1+\bar{w})\bar{\rho}}{2}(\boldsymbol{\nabla}_{\boldsymbol{x}}P)^2 - \frac{1}{2}Q\frac{4\pi}{a^{1+6\bar{w}}\Delta_{\boldsymbol{x}}}Q - \frac{1}{2}\frac{w_\delta}{1+\bar{w}}\frac{Q^2}{a^{3+6\bar{w}}\bar{\rho}} \right\} \ , \end{aligned} \tag{8.64}$$

and as before the fact that

$$P = \frac{\partial L}{\partial \dot{Q}} \tag{8.65}$$

means that P is indeed the canonically conjugate momentum of Q . The Hamiltonian then becomes

$$H[Q,P,t] = \int \mathrm{d}^3x \ \left\{ \frac{a^{1+6\bar{w}}(1+\bar{w})\bar{\rho}}{2}(\boldsymbol{\nabla}_{\boldsymbol{x}}P)^2 + \frac{1}{2}Q\frac{4\pi}{a^{1+6\bar{w}}\Delta_{\boldsymbol{x}}}Q + \frac{1}{2}\frac{w_\delta}{1+\bar{w}}\frac{Q^2}{a^{3+6\bar{w}}\bar{\rho}} \right\} \ , \tag{8.66}$$

which in Fourier space is

$$H[\tilde{Q},\tilde{P},t] = \int \frac{\mathrm{d}^3k}{(2\pi)^3} \ \left\{ \frac{a^{1+6\bar{w}}(1+\bar{w})\bar{\rho}k^2}{2}|\tilde{P}|^2 + \frac{1}{2}\left(\frac{w_\delta}{1+\bar{w}}\frac{1}{a^{3+6\bar{w}}\bar{\rho}} - \frac{4\pi}{a^{1+6\bar{w}}k^2} \right)|\tilde{Q}|^2 \right\} \ . \tag{8.67}$$

Exercise 65

Show that Equation 8.67 can be written as

$$H[\tilde{Q},\tilde{P},t] = \int \frac{\mathrm{d}^3k}{(2\pi)^3} \ \left\{ \frac{1}{2M_k}|\tilde{P}|^2 + \frac{M_k\Omega_k^2}{2}|\tilde{Q}|^2 \right\} \tag{8.68}$$

with the definitions

$$M_k \equiv \frac{1}{a^{1+6\bar{w}}(1+\bar{w})\bar{\rho}k^2} \tag{8.69}$$

$$\Omega_k^2 \equiv \frac{w_\delta k^2}{a^2} - 4\pi(1+\bar{w})\bar{\rho} \ . \tag{8.70}$$

From Equations 8.69 and 8.70 it is clear that **at any given moment in time** the Hamiltonian of Equation 8.68 will represent a collection of harmonic oscillators for modes with

$$w_\delta k^2 > 4\pi(1+\bar{w})a^2\bar{\rho} \ . \tag{8.71}$$

This can be achieved for any positive w_δ . However, the more important question is: If we fix k, will Ω_k^2 always becomes positive if we go sufficiently far back into the past? If this is the case, then **every mode k will have behaved like a harmonic oscillators** in the past. And this would allow us to build a quantum theory of Q and P which had a well defined ground state in the past.

Exercise 66

From the first and the "third" Friedmann equations (Equations 8.16 and 8.18), show that

$$\bar{\rho} \propto a^{-3(1+\bar{w})} \ . \tag{8.72}$$

Hint: take the time derivative of Equation 8.16 and insert Equation 8.18 into this. Then express the logarithmic derivative $\mathrm{d}\log\bar{\rho}/\mathrm{d}t$ *in terms of* $H = \mathrm{d}\log a/\mathrm{d}t$.

With you result from Exercise 66 Ω_k^2 becomes

$$\Omega_k^2 = \frac{w_\delta k^2}{a^2} - 4\pi(1+\bar{w})\bar{\rho}_0\left(\frac{a_0}{a}\right)^{3(1+\bar{w})} \ , \tag{8.73}$$

where $\bar{\rho}_0$ is the background density at some arbitrarily chosen time t_0 and $a_0 = a(t_0)$.

Exercise 67

If you require that $\Omega_k^2 > 0$ in the infinite past, what can you conclude for the value of $\bar{w}$?

As you saw in Exercise 67 , for $\bar{w} \approx -1$ (in fact, for even higher values) the first term in Equation 8.73 will always dominate if we go back for enough into the past, i.e.

$$\Omega_k^2 \approx \frac{w_\delta k^2}{a^2} \quad \text{for} \quad a \to 0 \; . \tag{8.74}$$

Also, for $w_\delta > 0$, this term is positive. So at an early enough point in time the mode Q_k indeed behaves like a harmonic oscillator. So to arrive at a quantum theory of the early-time density fluctuations, let us replace these oscillators by quantum harmonic oscillators, and let us assume that these oscillators started in their ground state at some (early) moment in time. It is a standard result of quantum mechanics, that the ground state of a harmonic oscillators is given by a Gaussian wave function of the form

$$\psi_0(Q_k) = \frac{1}{(2\pi\sigma_k^2)^{1/4}} \; \exp\left(-\frac{1}{4}\frac{|Q_k|^2}{\sigma_k^2}\right) \; . \tag{8.75}$$

In our language of Section 7 we would say that we are in a pure state of the form

$$\begin{aligned} \rho(Q_{k,1}, Q_{k,2}) &= \psi_0(Q_{k,1})^* \psi_0(Q_{k,2}) \\ &= \frac{1}{(2\pi\sigma_k^2)^{1/2}} \; \exp\left(-\frac{1}{4}\frac{|Q_{k,1}|^2 + |Q_{k,2}|^2}{\sigma_k^2}\right) \; , \end{aligned} \tag{8.76}$$

which on its diagonal leads to the Gaussian PDF

$$p(Q_k) \equiv \rho(Q_k, Q_k) = \frac{1}{(2\pi\sigma_k^2)^{1/2}} \; \exp\left(-\frac{1}{2}\frac{|Q_k|^2}{\sigma_k^2}\right) \; . \tag{8.77}$$

Note that Q_k here is a complex variable, and we can read ψ_0 as a joint wave function of both the real and imaginary part of Q_k . Both of those parts will have the same variance σ_k^2 which is given by

$$\sigma_k^2 = \frac{1}{2M_k\Omega_k} = \frac{a^{2(1+3\bar{w})}(1+\bar{w})\bar{\rho}k}{\sqrt{w_\delta}} \; . \tag{8.78}$$

This variance is the power spectrum of Q_k ! So we have shown that at sufficiently early times (or equivalently for sufficiently large k) the power spectrum of Q is proportional to k .

Exercise 68

Remember that we had defined $Q \equiv a^{3(1+\bar{w})}\delta\rho$. We had also said that $\bar{\rho} \propto a^{-3(1+\bar{w})}$. What can you conclude from that for the power spectrum of the relative density contrast $\delta = \delta\rho/\bar{\rho}$?

The assumption that the oscillators Q_k start in their ground state at some early moment in time indeed leads to the Gaussian initial conditions we had assumed from the very beginning of this course. But the above derivations have a major flaw: the mass and frequency of the oscillator Q_k are time dependent! This means in particular, that the ground state of the oscillator will change with time. So it seems that our construction crucially depends on the moment in time at which we choose to start in the ground state. Fortunately, this is not a problem. It turns our that at early enough times the mass M_k and the frequency Ω_k change adiabatically. This means that they change so slowly, that the wave function has time to "catch up" with their evolution and can manage to stay in the (changing) ground state. SEE VIDEO DURING LECTURE.

8.4 Spectrum of perturbations at super-Horizon scales

Very early in this course - in Section 3.1 - we had said that the power spectrum $P_i(k)$ of the initial density contrast field scales as k^{n_s} with a power law index n_s that is close to 1. Now we have seen in the previous subsection that the field Q is a constant multiple of the density contrast δ (cf. Exercise 68) and we have also seen that - at least on scales below the Hubble radius during inflation - the power spectrum of Q is proportional to k . This seems to be close to where we want to get. And maybe there is just some inaccuracy in our above derivations which, once we fix it, changes the power spectrum scaling to k^{n_s} ?

This is in fact **NOT THE CASE** ! And to truly understand where the k^{n_s} scaling is coming from we need to consider the behaviour of $\tilde{Q}_{\boldsymbol{k}}$ on super-horizon scales, i.e. when the physical length scale $\lambda_{\rm ph} \sim a/k$ becomes larger than the size of the Hubble horizon $1/H$. At this scale our above derivations break down in the following two respects.

- The Poisson equation and the continuity equation pick up general relativistic corrections which we ignored in Equation 8.48 and in Equation 8.57 (note that we had used the Poisson equation to express the gravitational potential in terms of $\delta(\rho)$.
- The perturbation δX of the auxiliary quantity X we defined in our discussion of the inflaton field is suppressed at super horizon scales. As a consequence, our assumption that the equation-of-state parameter of the perturbations w_δ is approximately equal to 1 does not hold anymore.

A fully relativistic discussion of super-horizon effects is beyond the scope of this course (though see e.g. the text books [41] and [42] for a complete treatment). To nevertheless arrive at quantitative statements, we will simply make the following assumption.

- **At super-horizon scales**: $w_\delta \approx \bar{w} \equiv w$.

In Appendix A we show that this leads to the following equation of motion for $\tilde{Q}_{\boldsymbol{k}}$ when $k^2/a^2 \ll H^2$:

$$\ddot{\tilde{Q}}_{\boldsymbol{k}} + (4+3w)H\dot{\tilde{Q}}_{\boldsymbol{k}} = 0 \ . \tag{8.79}$$

Exercise 69

Show that for a constant equation-of-state parameter w the Hubble rate is given by

$$H(t) = \frac{2}{3(1+w)}\frac{1}{t} \ . \tag{8.80}$$

Hint: use the Friedmann equations together with the fact that $\bar{\rho} \propto a^{-3(1+w)}$.

Exercise 70

Using your result from Exercise 69, show that Equation 8.79 has the two solutions

$$\tilde{Q}_{\boldsymbol{k}}(t) = \text{const.} \tag{8.81}$$

and

$$\tilde{Q}_{\boldsymbol{k}}(t) \propto t^{-1-\frac{2}{3}(4+3w)/(1+w)} \ . \tag{8.82}$$

Any general solution of Equation 8.79 will be a linear superposition of the two solutions you derived in Exercise 70 . But the second solution decays very quickly with time. So soon after

the physical size a/k of the wave vector $\boldsymbol{k}$ has crossed the horizon scale, the mode $\tilde{Q}_{\boldsymbol{k}}$ will freeze at a constant value. What does this mean for the average amplitude of $\tilde{Q}_{\boldsymbol{k}}$? We had seen in the previous section that on **sub-horizon scales** the time dependence and the scale dependence of the variance of $Q_{\boldsymbol{k}}$ is given by

$$\begin{aligned}\sigma_k^2 &\propto a^{2(1+3\bar{w})}\bar{\rho}k \\ &\propto a^{2(1+3\bar{w})}H^2k \;\approx\; a^{-4}H^2k \;,\end{aligned} \tag{8.83}$$

where in the first line we have used the fact that $\bar{w} \approx -1$. Now there will be some moment in time $t_{k,\mathrm{exit}}$ where

$$k/a(t_{k,\mathrm{exit}}) = H(t_{k,\mathrm{exit}}) \;, \tag{8.84}$$

i.e. when the mode exits the horizon. At that moment the time dependence of σ_k^2 freezes at a value

$$\sigma_k^2(t > t_{k,\mathrm{exit}}) \propto a^{-4}(t_{k,\mathrm{exit}})H^2(t_{k,\mathrm{exit}})k \;=\; H^6(t_{k,\mathrm{exit}}) \;/\; k^3 \;. \tag{8.85}$$

Remember that for $w \approx -1$ the expansion rate H stays almost constant in time. So the above factor of $H^6(t_{k,\mathrm{exit}})$ will only very weakly depend on k and the power spectrum of Q (and hence of δ!) will roughly scale as k^{-3} .

This seems to bring us farther away from the k^{n_s} scaling that was mentioned in Section 3.1 ! But there is one more step in the evolution of the initial density perturbations we need to consider - the end of inflation. For simplicity, we will assume that at some moment in time the inflaton field decays into the particles of the standard model. At this point the Universe becomes filled by a hot, highly relativistic plasma, i.e. it enters its radiation dominated phase. Note that on super-horizon scales $k/a \ll H$ Equation 8.79 will still be valid in this phase - just with a changed equation-of-state parameter $w = 1/3$. In particular, the equation still has a solution that is constant in time, and our modes $\tilde{Q}_{\boldsymbol{k}}$ stay frozen (though the transition from $w = -1$ to $w = 1/3$ causes a k-independent change in the overall amplitude of $\tilde{Q}_{\boldsymbol{k}}$, cf. [41]).

But in the radiation dominated phase the expansion rate H decreases faster than $1/a$, as you will show in the following exercise.

Exercise 71

Using the first Friedmann equation as well as your result from Exercise 66 , show that in a radiation dominated universe H scales as $\propto 1/a^2$.

This means that at some point k/a will again be larger that H and the Fourier modes of Q will re-enter the horizon. This will happen at a time $t_{k,\mathrm{enter}}$ when

$$k/a(t_{k,\mathrm{enter}}) = H(t_{k,\mathrm{enter}}) \tag{8.86}$$

$$\Rightarrow \quad a(t_{k,\mathrm{enter}}) \propto 1/k \;. \tag{8.87}$$

There is now a time window in which the mode $\tilde{Q}_k$ starts to feel gravity, but where its physical wavelength is still too large to be sensitive to the pressure of the relativistic plasma. Within that window, the mode $\tilde{Q}_k$ will evolve according to

$$\tilde{Q}_k(t) = \tilde{Q}_k(t_{k,\mathrm{enter}}) \; \frac{D(t)}{D(t_{k,\mathrm{enter}})} \;, \tag{8.88}$$

where D is the k-independent, linear growth factor we already encountered in Section 1 . Correspondingly, the power spectrum will evolve as

$$\sigma_k^2(t > t_{k,\text{enter}}) \propto \frac{H^6(t_{k,\text{exit}})}{k^3} \, \frac{D^2(t)}{D^2(t_{k,\text{enter}})} \; . \tag{8.89}$$

Our final piece of the puzzle is now given by the result of the following exercise.

Exercise 72

Not graded: *Consider the Hamiltonian we had derived in Equation 8.67 , but set $w_\delta = \bar{w} = 1/3$. Show that when approximating the squared frequency as $\Omega_k^2 \approx -4\pi(1+w)\bar{\rho}$ the resulting equation of motion for $\tilde{Q}_{\boldsymbol{k}}$ is solved by the ansatz $\tilde{Q}_{\boldsymbol{k}}(t) \propto D(t)$ with the linear growth factor*

$$D(t) \propto a(t)^2 \; . \tag{8.90}$$

Also, show that the other independent solution to the equation of motion is again decaying.

With the result of the previous exercise we can conclude that

$$\sigma_k^2(t > t_{k,\text{enter}}) \propto \frac{H^6(t_{k,\text{exit}})}{k^3} \, \frac{a^4(t)}{a^4(t_{k,\text{enter}})} \;=\; H^6(t_{k,\text{exit}}) \; a^4(t) \; k \; . \tag{8.91}$$

So we again end up with a power spectrum that seems to be linear in k . But remember that during inflation the Hubble rate was only approximately constant. There is in fact a slight decrease of H with time during that epoch. Since modes with larger k left the horizon at a later time $t_{k,\text{exit}}$, this means that $H^6(t_{k,\text{exit}})$ slightly decreases as k increases. This leads to an effective scaling of σ_k^2 as k^{n_s} with n_s slightly smaller than 1 .

Exercise 73

Go through Section 8.4 carefully again and find at least three points in our derivation, where things could go wrong and where we would need to think harder in order to make sure that our final result is correct.

Exercise 74

Summarise for yourself all the approximations and simplifications we made throughout Section 8 .

Exercise 75

What conceptual difference do you notice between how we treated the perturbations $\tilde{Q}_{\boldsymbol{k}}$ in Section 8.4 compared to how we treated them at the end of Section 8.3 ?

Exercise 75 is aiming for an issue that will concern us in more detail at the end of this course (cf. Sections 10 and 11).

9 Exkurs: What is a quantum PDF?

Before we continue, we will do a little exkurs into the mathematical subject of Quantum logic. The purpose of this is twofold: Quantum logics provides a unified framework for formulate both classical and quantum probability theory in (cf. [1, 12, 44, 56] and references therein). And it will widen our vocabulary when it comes to discussing how to interpret quantum mechanics.

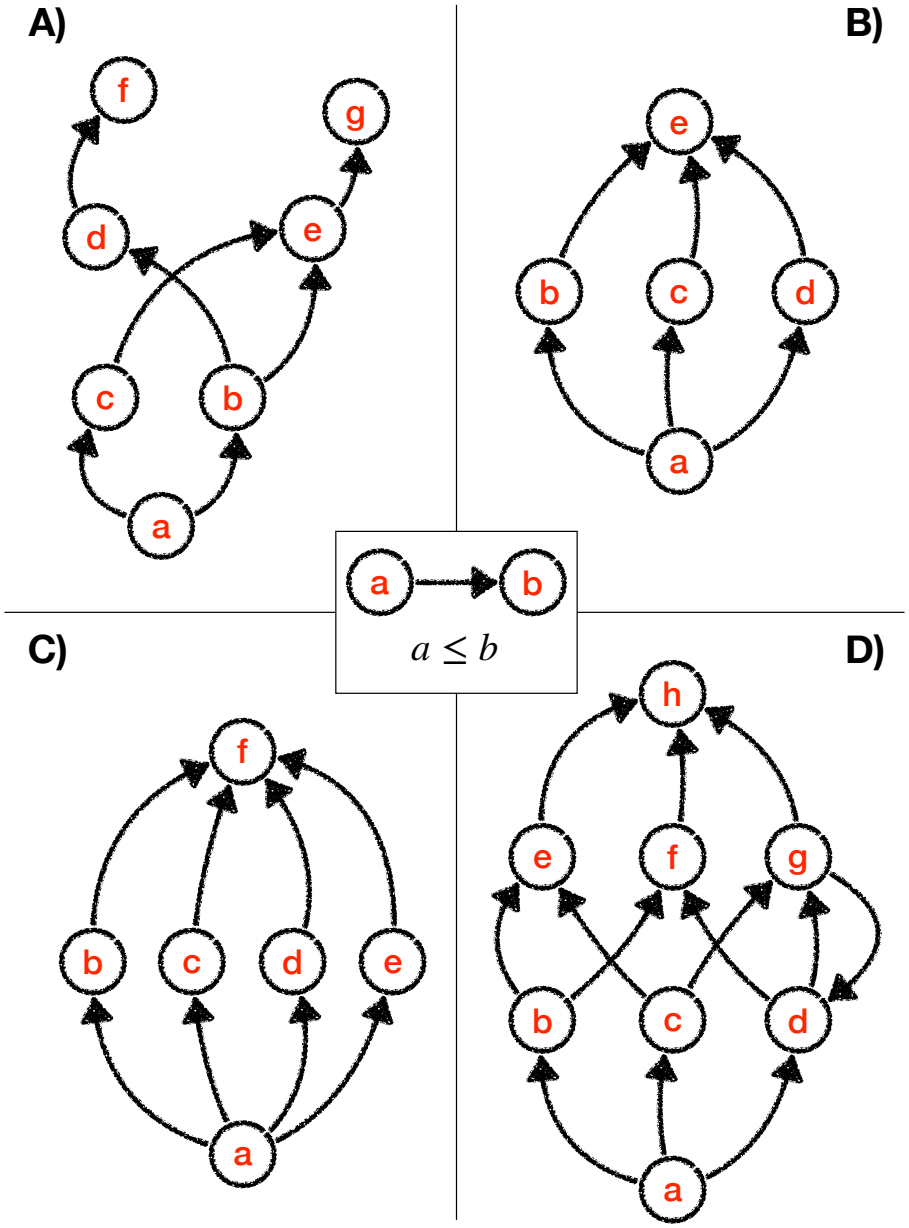

Figure 20. Graphs representing sets of statements together with an implication relation between them. Which of the cases could represent a mathematical logic (as it was defined in Section 9.1). Figure referred to in Exercise 76 .

9.1 From lattices to logics

In quantum logics (or QL for short) a physical system is described by a set of possible experimental statements about this system together with an implication relation $\leq$ between some of those statements. Let us denote statements by lowercase letters a, b, c etc. We can then read sentences like

$$a \leq b \tag{9.1}$$

as "the statement a implies the statement b ". What are natural properties we would expect from this implication relations? For starters, every statement should imply itself, i.e.

$$a \leq a \ . \tag{9.2}$$

Also, if a implies b but b also implies a then the two statements a and b are logically equivalent, so we can take them to be identical:

$$((a \leq b) \text{ and } (b \leq a)) \ \Rightarrow (a = b) \ . \tag{9.3}$$

And thirdly, if a implies b and b implies c then we would expect that also a implies c . This means that

$$((a \leq b) \text{ and } (b \leq c)) \ \Rightarrow (a \leq c) \ . \tag{9.4}$$

If we denote the set of all possible statements as $\mathcal{L}$ then we can group Equations 9.2, 9.3 and 9.4 into the following assertion.

A) **The relation $\leq$ is a *partial ordering* on $\mathcal{L}$.**

When we describe a physical system, then there are typically two special types of statements: *tautologies*, i.e. statements that are always true regardless of the state of the system, and *contradictions*, i.e. statements that are always false. Since all tautologies are logically equivalent, we will collectively refer to them as a single statement $\mathbf{1}$. Similarly, we will collectively refer to all contradictions as the single statement $\mathbf{0}$. Any statement implies the tautology, so we have

$$a \leq \mathbf{1} \tag{9.5}$$

for each statement a . Also, the contradiction implies every statement, so we have

$$\mathbf{0} \leq a \tag{9.6}$$

for all statements a . The existence of $\mathbf{0}$ and $\mathbf{1}$ implies that $\mathcal{L}$ together with $\leq$ is a *bounded partial ordering*.

Given any two statements a and b in $\mathcal{L}$ there should be a smallest possible statement which is implied by both a and b. We will denote this statement by $a \wedge b$ (in standard Boolean logic you would refer to this as "a **and** b"). Similarly, there should be a largest possible statement which implies both a and b, and we will refer to this as $a \vee b$ (in standard Boolean logic you would refer to this as "a **or** b"). The existence of $a \wedge b$ and $a \vee b$ for all statements a and b as well as the existence of the elements $\mathbf{0}$ and $\mathbf{1}$ can be summarised in the following assertion.

B) **The set of all statements $\mathcal{L}$ together with the relation $\leq$ constitutes a *lattice.***

To each statement a it should also be able to make the opposite statement $a^{\perp}$. Obviously, a is then also the opposite of $a^{\perp}$, so we have

$$a^{\perp\perp} = a \ . \tag{9.7}$$

Assume that a statement a implies another statement b. Then the opposite of b ("not b") should imply the opposite of a ("not a"), i.e.

$$(a \leq b) \implies (b^{\perp} \leq a^{\perp}) \ . \tag{9.8}$$

Obviously, "a or not a" is a tautology, so

$$a \vee a^{\perp} = \mathbf{1} \ , \tag{9.9}$$

while "a and not a" is a contradiction, so

$$a \wedge a^{\perp} = \mathbf{0} \ . \tag{9.10}$$

Equations 9.7 to 9.10 can be summarised as follows.

C) **The mapping $\perp$: $\mathcal{L} \to \mathcal{L}$ is an *orthocomplementation* on $\mathcal{L}$.**

We need two more items to finish our list of the properties we would expect from the set of experimental statements that can be made about any physical system. The first item can be considered as a mathematical technicality: For any countable set $\{a_i\}$ of statements let us define the statement

$$\bigwedge_i a_i$$

to be the largest statement that implies all the a_i . Similarly, let us define the statement

$$\bigvee_i a_i$$

to be the smallest statement that is implied by all the a_i . We will then assume that these elements indeed exist for all countable sets $\{a_i\}$ of statements that we can make about our system.

Our second remaining item goes as follows: if $a, c \in \mathcal{L}$ with $a \leq c$ then there exists an element $b \in \mathcal{L}$ such that $b \leq a^{\perp}$ and $a \vee b = c$. In other words, we assume that the difference between the statements a and c can be encoded into a third statement b . This leads to our final result:

D) **The set of all statements $\mathcal{L}$ together with the relation $\leq$ and the orthocomplementation $\perp$ is *a logic.***

The designation *logic* is primarily just a mathematical term. But you can indeed think of it at a mathematical attempt to define the minimal properties satisfied by the statements that can be made by a "logical agent".

Exercise 76

In Figure 20 you see four examples of sets of (abstract) statements together with an implication relation between them. Only one of those four examples could actually represent a logic - which case is that? What are the elements **0** *and* **1** *in that case? What would be potential pairs of orthogonal statements in that example? (Note: there are several possible answers to the last question.) Why can the other cases not represent logics?*

Exercise 77

In the above we have tried to collect general properties that should be satisfied by the set of statements that any physical theory can make about a physical system at hand. Do you think there is something missing among the list of properties we came up with? Do you think any of the properties we came up with is unnecessary?

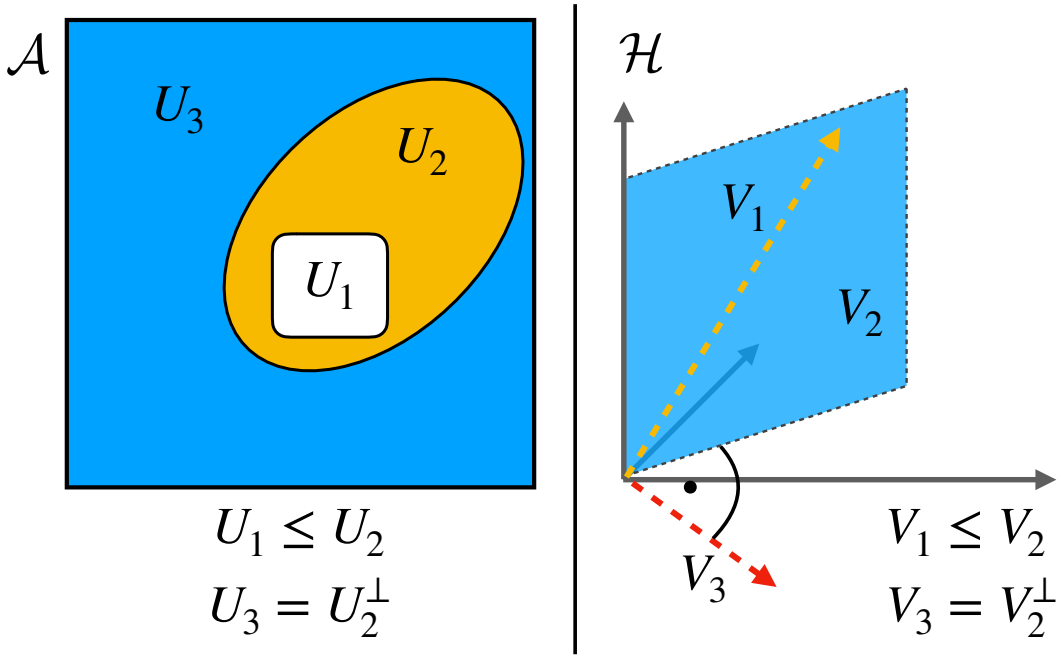

Figure 21. Left: sketch visualising the first of the two examples for logics mentioned at the end of Section 9.1 - the logic of subsets of a set $\mathcal{A}$. Right: visualisation of the second example - the logic of closed, linear subspaces of a Hilbert space $\mathcal{H}$.

Exercise 78
Let $\mathbb{R}^2 = \{(q,p)\}$ be the phase space of a (classical) 1-dimensional point particle. What are the possible statements you could make about that system?

There are two examples of logics that play a particularly important role in physics.

Example 1: Let $\mathcal{A}$ be a set and let $\mathcal{L}(\mathcal{A})$ be a set of subsets of $\mathcal{A}$, such that $\mathcal{A} \in \mathcal{L}(\mathcal{A})$ and such that for any $U \in \mathcal{L}(\mathcal{A})$ also $U^c \in \mathcal{L}(\mathcal{A})$, where U^c is the complement of U in $\mathcal{A}$. For technical reasons, let us assume that for any countable sequence of subsets $U_i \in \mathcal{L}(\mathcal{A})$ both the union

$$\bigcup_i U_i$$

and the intersection

$$\bigcap_i U_i$$

are also part of $\mathcal{L}(\mathcal{A})$. Then $\mathcal{L}(\mathcal{A})$ is a logic if we define the implication relation $\leq$ via

$$(U_1 \leq U_2) \Leftrightarrow (U_1 \subset U_2) \tag{9.11}$$

and if we define the orthocomplementation $\perp$ via

$$U^\perp \equiv U^c \ . \tag{9.12}$$

Example 2: Let $\mathcal{H}$ be a Hilbert space, and let $\mathcal{L}(\mathcal{H})$ be the set of all closed, linear subspaces of $\mathcal{H}$ (i.e. the set of all sub-vector spaces of $\mathcal{H}$, where the requirements of being "closed" is

again a mathematical technicality). Then $\mathcal{L}(\mathcal{H})$ is a logic if we define the implication relation $\leq$ between any two sub-spaces as

$$(V_1 \leq V_2) \Leftrightarrow (V_1 \subset V_2) \tag{9.13}$$

(i.e. via the inclusion relation, as in example 1) and if we define the orthocomplement $V^{\perp}$ of any sub-space V to be the largest sub-space that is orthogonal to V wrt. the inner product on $\mathcal{H}$.

We will see in the next subsection, that **these two examples are in some sense the only two examples of logics**. As a preparation for this, we will need the following definition: Let $a \in \mathcal{L}$ be a statement such that any other statement $b \in \mathcal{L}$ with $b \leq a$ is either given by $b = \mathbf{0}$ or $b = a$. The we call a an *elementary statement* (or an *atomic statement*, as they are also often called in the literature).

9.2 The theorems of Piron and of Solèr: characterising possible logics

It seems like the axioms by which we defined a logic above are quite broad and that they should be satisfied by a quite wide range of graphs connecting statements. But with two more assumptions we can pin-point the possible realisations of logics to a very narrow range of constructions.

Additional assumption 1: Let $a, b \in \mathcal{L}$ be two statements such that $a \leq b$ but $a \neq b$. Then we will assume that there is an elementary statement $c \in \mathcal{L}$ such that $c \leq b$ but $c \not\leq a$. In other words: "there is always an elementary statement that fits into the difference between two other statements". This property is often called that *atomicity assumption*.

Additional assumption 2: The lattice formed by the statements in $\mathcal{L}$ is *modular*. This means that for any three statements a, b, c with $c \leq a$ we have

$$a \wedge (b \vee c) = (a \wedge b) \vee c \ . \tag{9.14}$$

Exercise 79

Not graded: *Reading $\vee$ as "or" and $\wedge$ as "and", try to find everyday statements a, b, c that do not satisfy the additional assumption 2.*

Based on the above assumptions Constantin Piron [43] has proven the following theorem.

Piron's theorem: The statements a of any logic $\mathcal{L}$ that satisfies the above assumptions can be viewed as collections of statements $a_1, \ \ldots \ , \ a_N$ of smaller logics $\mathcal{L}_1, \ \ldots \ , \ \mathcal{L}_N$,

$$a = (a_1, \ \ldots \ , \ a_N) \ , \tag{9.15}$$

such that the implication relation between two statements $a, b \in \mathcal{L}$ is defined as

$$a \leq b \ \Rightarrow \ a_i \leq b_i \ \forall i \ . \tag{9.16}$$

Furthermore, all the smaller logics $\mathcal{L}_i$ are given by the logics of linear subspaces in generalised Hilbert spaces.

In the above construction you can read a statement $a = (a_1, \ \dots \ , \ a_N)$ as saying "a_1 or a_2 or ... or a_N". This is why this construction is also sometimes referred to as a Hilbert space logic with *selection rules*: if $\mathcal{H}_i$ are the Hilbert spaces corresponding to the logics $\mathcal{L}_i$ then we can consider the direct sum $\mathcal{H} = \bigoplus_i \mathcal{H}_i$ of all these Hilbert spaces, and the logic $\mathcal{L}$ restricts us to making statements that only lie in particular linear subspaces of the Hilbert space $\mathcal{H}$.

Of course we haven't yet specified what the *generalised Hilbert spaces* in the above theorem are. Fortunately, under a further additional assumption Maria Solèr has proven a theorem that restricts these to simply be Hilbert spaces over either the real numbers, the complex number or the quaternions. The additional assumption that she makes is rather technical, but in simple words it says that "for any orthogonal, elementary statements a and b there is an elementary statement c that is exactly in the middle of a and b". You can e.g. find an exact but still somewhat digestible version of this assumption in [44], who also tries to give a physical justification for it.

From the perspective of the above theorems it seems like any logic can be decomposed into Hilbert space logics. So where does this leave the more classical logics of subsets we discussed e.g. in Figure 21? Note that the construction of Piron's theorems contains statements of the form $(\mathbf{1}_1, \mathbf{1}_2, \mathbf{0}_3, \ \dots \ , \ \mathbf{1}_N)$ etc. , i.e. the statements that contain either the $\mathbf{0}$ or $\mathbf{1}$ of the sublogics $\mathcal{L}_i$. These statements actually form a classical logic, that can be interpreted as a logic of subsets.

9.3 PDFs as measures on a logic

Let us conclude this section by discussing measured on logics, and how they are related to PDFs. Let $\mathcal{L}$ be any logic and let P be a map

$$\begin{aligned} P : \mathcal{L} &\longrightarrow \mathbb{R} \\ a &\mapsto P(a) \ . \end{aligned}$$

We call p a measure on $\mathcal{L}$ if it satisfies the following properties:

- $0 \leq P(a) \leq 1$ for all statements a ,
- $P(\mathbf{0}) = 0$ and $P(\mathbf{1}) = 1$ and
- if $a_1, \ a_2, \ \dots$ is a sequence of mutually orthogonal statements, then

$$P\left(\bigvee_i a_i\right) = \sum_i P(a_i) \ . \tag{9.17}$$

It turns out that the measures on the logics of subsets U of a set $\mathcal{A}$ are exactly the PDFs on $\mathcal{A}$. So if $\mathcal{A}$ is e.g. the phase space of a 1-dimensional particle, then any measure P that satisfies the above properties is represented by a PDF $p(q,p)$ such that

$$P(U) = \int_{(q,p)\in U} \mathrm{d}q\mathrm{d}p \ p(q,p) \tag{9.18}$$

for any $U \subset \mathcal{A}$ in the logic.

Furthermore, it can be shown that the measures on the logics corresponding to linear subspaces in a Hilbert space $\mathcal{H}$ are exactly the state operators $\hat{\rho}$ we discussed in Section 7 .

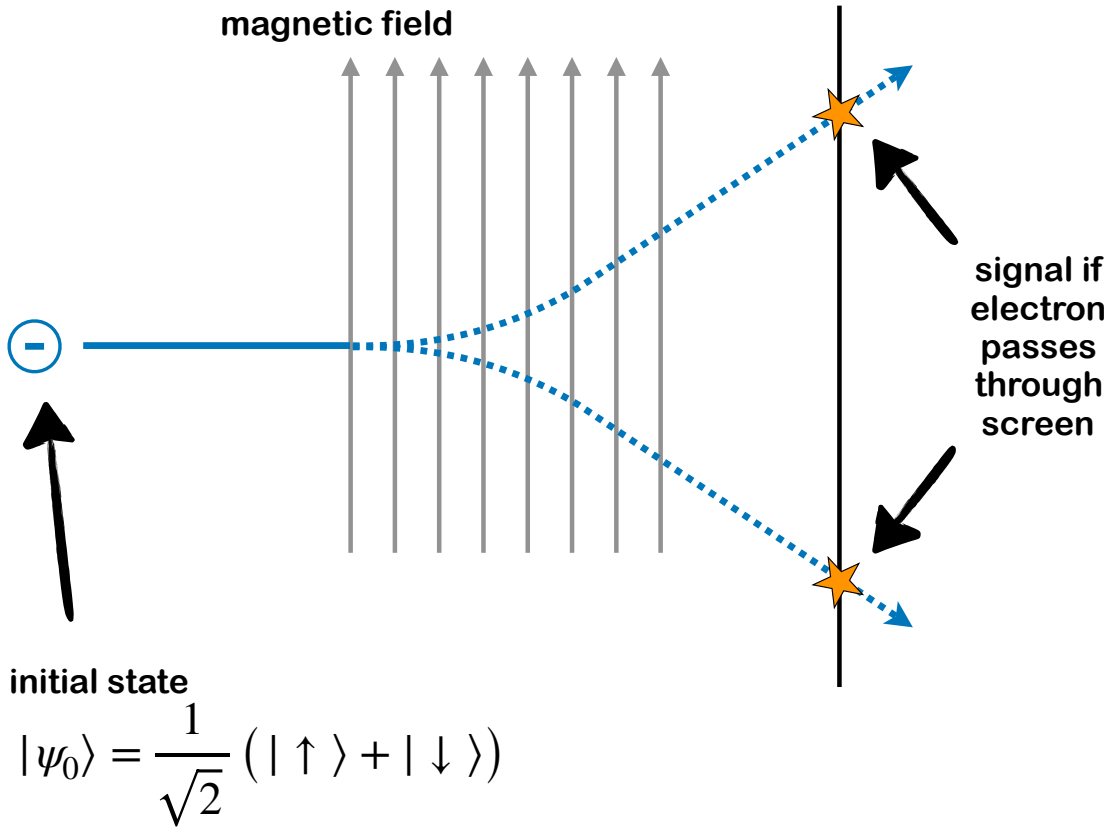

Figure 22. Sketch of the Stern-Gerlach experiment: An electron is prepared in a superposition of "spin up" ($|\uparrow\rangle$) and "spin down" ($|\downarrow\rangle$) and is then shot though a magnetic field. The $|\uparrow\rangle$-part of the wave function will then be deflected upwards while the $|\downarrow\rangle$-part will be deflected downwards. For the purpose of our discussion we have assumed that a screen is placed behind the magnetic field, and that that screen emits a light signal once the electron passes through it. Note also that we have neglected the Lorentz force in our discussion which would accelerate the electron in a direction perpendicular to the magnetic field.

For any subspace $V \subset \mathcal{H}$ such an operator defines a measure as

$$P(V) = \sum_{i=1}^{\dim(V)} \langle\psi_i|\,\hat{\rho}\,|\psi_i\rangle \ , \tag{9.19}$$

where $\{|\psi_i\rangle\}$ is any orthonormal (sub-)basis of V . In particular, the above definition of $P(V)$ is independent of the choice of the basis $\{|\psi_i\rangle\}$.

The above results were the motivation for why we interpreted state operators $\hat{\rho}$ as the quantum analogs of PDFs in Section 7 .

10 Discussion: the measurement problem and the many worlds attempt to explain probabilities

According to the inflationary paradigm, the primordial fluctuations of the cosmic density field were quantum fluctuations. In particular, the ground state in which we let our harmonic oscil-

lators $\tilde{Q}_{\boldsymbol{k}}$ start in Section 8.3 are quantum super-positions of many possible configurations of $\tilde{Q}_{\boldsymbol{k}}$ (and hence of the density contrast field $\tilde{\delta}_{\boldsymbol{k}}$). Contrary to that, the density field we observe today is only one of those possible configurations - cf. the DES observations in Figure 1 . How does this transition from a quantum super-position to a concrete realisation (e.g. drawn from a classical random field) happen? Before we examine this question in detail in Section 11 , let us have a very simplified look at the measurement process in quantum mechanics.

A good situation to study this process is the Stern-Gerlach experiment (see the sketch in Figure 22). Assume that we have prepared an electron in a quantum super position of the states $|\uparrow\rangle$ - i.e. "spin up" - and $|\downarrow\rangle$ - i.e. "spin down" - wrt. the z-direction of a lab frame. This e.g. means that the spin of the electron is described by the pure state

$$|\psi_0\rangle = \frac{1}{\sqrt{2}} \left(|\uparrow\rangle + |\downarrow\rangle\right) \ , \tag{10.1}$$

where for the purpose of this section we return to the more standard convention of describing pure states by Hilbert space vectors $|\psi\rangle$ as opposed to state matrices $|\psi\rangle \langle\psi|$.

We now shoot our electron along the lab-frame x-axis through a magnetic field that is aligned with the z-axis. The $|\uparrow\rangle$-part of the electron will then be deflected upwards and the $|\downarrow\rangle$-part will be deflected downward[7]. By separating the two parts in that way we are able to perform "a spin measurement" - e.g. by catching the electron on a screen behind the magnetic field. If the electron hits the screen at a z-value > 0 the the measurement result is "spin up", and if it hits the screen at a z-value < 0 the measurement result is "spin down". Let us assume that we have designed the screen in such a way that its interaction with the electron does not destroy said electron nor its spin state. E.g. the electron could pass through the screen, leaving behind only a luminescent signal at the point of impact. Then our spin measurement changes the state of the electron! There will only be one concrete outcome of the measurement - either spin up or spin down. Say the outcome is spin up. Then this change is

$$|\psi_0\rangle \to |\uparrow\rangle \ . \tag{10.2}$$

And the standard ("Kopenhagen") interpretation of quantum mechanics would say that either of the two possible outcomes happens with a probability of $1/2$.

But there is a short coming of our above description of the Stern-Gerlach setup. Quantum mechanics is believed to be a theory of universal validity, applying to every aspect of the Universe. Hence, it should also apply to the screen with which we measure the deflection of the electron. The joint system of screen and electron before the measurement would then be described by the tensor product state

$$|\Psi_0\rangle \equiv |\psi_0\rangle \, |\text{screen}_0\rangle = \frac{1}{\sqrt{2}} \left(|\uparrow\rangle + |\downarrow\rangle\right) |\text{screen}_0\rangle \ , \tag{10.3}$$

where $|\text{screen}_0\rangle$ describes the state of the screen has not yet been hit by an electron. Once the electron passes the screen, the quantum state will actually evolve as

$$|\Psi_0\rangle \to |\Psi_1\rangle \equiv \frac{1}{\sqrt{2}} \left(|\uparrow\rangle \, |\text{screen}_\uparrow\rangle + |\downarrow\rangle \, |\text{screen}_\downarrow\rangle\right) \ , \tag{10.4}$$

[7]There will in fact also be a Lorentz force which would accelerate the electron in a direction perpendicular to the magnetic field. We could e.g. suppress this force by switching on a carefully chosen electric field perpendicular to the magnetic field. But for the purpose of our discussion we will simply ignore the Lorentz force.

where $|\text{screen}_\uparrow\rangle$ is the state in which the screen shows a luminescent signal at $z > 0$ and $|\text{screen}_\downarrow\rangle$ is the state where it shows a luminescent signal at $z < 0$. So in this more complete picture, there is actually no concrete measurement outcome! The quantum state simply evolves into a quantum super-position of "the electron has spin up and the screen has a signal at $z > 0$" and of "the electron has spin down and the screen has a signal at $z < 0$" . But real Stern-Gerlach experiment never yield such an outcome. Instead, they always result in either the ourcome $|\downarrow\rangle |\text{screen}_\uparrow\rangle$ or the outcome $|\downarrow\rangle |\text{screen}_\downarrow\rangle$. So what has gone wrong here? Maybe the answer is simply: once the experimenter actually looks at the screen they perceive with their own eyes whether the luminescent signal appeared at $z > 0$ or at $z < 0$, and this process of "looking" at the outcome is the actual measurement process that e.g. changes the initial state as

$$|\Psi_1\rangle \to |\uparrow\rangle |\text{screen}_\uparrow\rangle \ . \tag{10.5}$$

This "measurement by looking" should again yield either of the results $|\downarrow\rangle |\text{screen}_\uparrow\rangle$ or $|\downarrow\rangle |\text{screen}_\downarrow\rangle$ with a probability of $1/2$.

But, of course the true scientific believer would say that also the experimenter themselves is a quantum system. Then we would actually start with the state

$$|\Psi_0\rangle \equiv |\psi_0\rangle |\text{screen}_0\rangle |\text{🧍}_0\rangle = \frac{1}{\sqrt{2}} (|\uparrow\rangle + |\downarrow\rangle) |\text{screen}_0\rangle |\text{🧍}_0\rangle \ , \tag{10.6}$$

where $|\text{🧍}_0\rangle$ is the state in which the experimenter has not yet looked at the screen. The above product state would then evolve as

$$|\psi_0\rangle |\text{screen}_0\rangle |\text{🧍}_0\rangle \to |\Psi_1\rangle \equiv \frac{1}{\sqrt{2}} \left(|\uparrow\rangle |\text{screen}_\uparrow\rangle |\text{🧍}_\uparrow\rangle + |\downarrow\rangle |\text{screen}_\downarrow\rangle |\text{🧍}_\downarrow\rangle \right) \ , \tag{10.7}$$

i.e. now we seem have a super-position of the experimenter thinking that they have measured spin up and of the experimenter thinking that they have measured spin down! What does it take for this super-position to collapse into one well defined mindset of the observer? E.g. a conversation with a second person about the measurement outcome? This would clearly only shift the problem to the next level, and we would always remain in weird super-position states which none of us in real-life have ever observed. So quantum mechanics seems to be in contradiction with reality.

How could we try to solve this problem? Well first of all, note that the state $|\Psi_1\rangle$ is actually NOT a superposition of the states $|\text{🧍}_\uparrow\rangle$ and $|\text{🧍}_\downarrow\rangle$. It is only a superposition of the full product state

$$|\uparrow\rangle |\text{screen}_\uparrow\rangle |\text{🧍}_\uparrow\rangle$$

and the full product state

$$|\downarrow\rangle |\text{screen}_\downarrow\rangle |\text{🧍}_\downarrow\rangle \ .$$

To truly obtain a superposition of different states-of-mind of the observer the states of screen and the electron would have to evolve as

$$|\uparrow\rangle |\text{screen}_\uparrow\rangle \to |?\rangle |\text{screen}_?\rangle$$
$$|\downarrow\rangle |\text{screen}_\downarrow\rangle \to |?\rangle |\text{screen}_?\rangle$$

such that the overall state evolves into

$$|\Psi_1\rangle \rightarrow |\Psi_?\rangle \equiv \frac{1}{\sqrt{2}} |?\rangle \, |\text{screen}_?\rangle \left(|\text{🧍}_\uparrow\rangle + |\text{🧍}_\downarrow\rangle \right) \ . \tag{10.8}$$

Leaving aside the question of whether such an evolution is possible in a unitary way, it is anyway probably quite unlikely, that the above transition will take place. This becomes even more true if we consider a 4th system which is always present in any experiment: the environment. There will e.g. be countless photons that are bouncing of the observer as they're writing down the result of there measurement. If they e.g. writing the words "spin up" into their notebook the photons in the room will bounce off their hands in a different way than if they had noted down the words "spin down". This will result in to different states of the photons - or more generally: the environment - and these different states will quickly become almost perfectly orthogonal in the environment Hilbert space. The total state after the measurement will then actually be

$$|\Psi_1\rangle = \frac{1}{\sqrt{2}} \left(|\uparrow\rangle \, |\text{screen}_\uparrow\rangle \, |\text{🧍}_\uparrow\rangle \, |\text{Env}_\uparrow\rangle + |\downarrow\rangle \, |\text{screen}_\downarrow\rangle \, |\text{🧍}_\downarrow\rangle \, |\text{Env}_\downarrow\rangle \right) \ , \tag{10.9}$$

and there is for all practical purposes no way in which the two states $|\text{Env}_\uparrow\rangle$ and $|\text{Env}_\downarrow\rangle$ of the complicated, messy environment ever evolve into an identical state $|\text{Env}_?\rangle$ again such that we could obtain a true superposition of the observer states $|\text{🧍}_\uparrow\rangle$ and $|\text{🧍}_\downarrow\rangle$.

The above situation lends itself to a radical interpretation that was first proposed by [24] and which has come to be known as the *many worlds* interpretation of quantum mechanics. This interpretation builds on the two observations

A) the linearity of the Schrödinger equation implies that the two states

$$|\uparrow\rangle \, |\text{screen}_\uparrow\rangle \, |\text{🧍}_\uparrow\rangle \, |\text{Env}_\uparrow\rangle$$

and

$$|\downarrow\rangle \, |\text{screen}_\downarrow\rangle \, |\text{🧍}_\downarrow\rangle \, |\text{Env}_\downarrow\rangle$$

evolve independently of each other.

B) The entanglement of the observer with the other systems (in our situation: electron, screen, environment) prevents superpositions between different observer states.

Based on there observations the many worlds interpretation concludes that the two terms on the right-hand side of Equation 10.9 are two independent branches of the quantum state that represent two different, independent universes. In each of those universes, the observer then has a well defines state(-of-mind). And if you were that observer, then the probability that you find yourself on either of these branches after the measurement is 1/2 . To make this - and in particular observation B) from above - more clear, let us return to our notation of Section 7 and describe the total quantum state by the state operator

$$\hat{\rho}_{1,\text{total}} = |\Psi_1\rangle \, \langle\Psi_1| \ . \tag{10.10}$$

The environment system is so vast and messy that we will never know its exact state. We can take into account this ignorance by taking the trace of $\hat{\rho}_{1,\text{total}}$ over the basis states of the

environment Hilbert space. A standard calculation then shows that the state operator in the remaining part of the Hilbert space (electron-screen-observer) is given by

$$\hat{\rho}_{1,\text{no}-\text{Env}} = \frac{1}{2} |\uparrow\rangle |\text{screen}_\uparrow\rangle |\text{🧍}_\uparrow\rangle \langle \text{🧍}_\uparrow| \langle \text{screen}_\uparrow| \langle\uparrow| + \frac{1}{2} |\downarrow\rangle |\text{screen}_\downarrow\rangle |\text{🧍}_\downarrow\rangle \langle \text{🧍}_\downarrow| \langle \text{screen}_\downarrow| \langle\downarrow| \ . \tag{10.11}$$

This should remind you of the situation we observed in the middle panel of Figure 19 . There we had considered a (classical) mixture of two Gaussian wave packages and we had seen that those two wave packages did not interfere with each other and instead behaved like two independent quantum states.

So the many worlds interpretation seems to achieve two things. It explains why each observer actually experiences concrete measurement outcomes. And it seems to explain the origin of quantum randomness: there are indeed two different versions of the universe that emerge from the measurement process and there is a randomness to whether or not you as an observer would find yourself on either of the emerging branches.

Exercise 80

Done during lecture - not graded: *Form groups of 3-4 people and discuss the above line of thought for a concrete example. First, enumerate the members of your group. And at the end of your discussion revisit the webpage* `https://qrng.anu.edu.au/dice-throw/` *to throw their "quantum dice" such that the number of possible outcomes corresponds to the number of people in your group. The person who's number gets drawn will be the one to report the results of your group's discussion back to the plenary.*
Now what should you discuss? First, formulate the experiment you are about to run as a quantum measurement in analogy to how we considered the Stern-Gerlach experiment above. Then, try to spot a logical shortcoming in the way in which the many worlds interpretation derives the emergence of randomness/probability in such an experiment. There is at least one, but maybe you can come up with more?

DO NOT CONTINUE TO READ IF YOU DON'T WANT THE LECTURER'S OPINION TO INFLUENCE YOU IN EXERCISE 80.

Imagine that we had initially prepared the electron in the state

$$|\psi_0\rangle = \sqrt{\frac{2}{3}}\,|\uparrow\rangle + \sqrt{\frac{1}{3}}\,|\downarrow\rangle \ . \tag{10.12}$$

Then the final state of the combined system electron-screen-observer-environment would be

$$|\Psi_1\rangle = \sqrt{\frac{2}{3}}\,|\uparrow\rangle\,|\text{screen}_\uparrow\rangle\,|\text{🧍}_\uparrow\rangle\,|\text{Env}_\uparrow\rangle + \sqrt{\frac{1}{3}}\,|\downarrow\rangle\,|\text{screen}_\downarrow\rangle\,|\text{🧍}_\downarrow\rangle\,|\text{Env}_\downarrow\rangle \ . \tag{10.13}$$

This state still contains exactly two independent branches, and the prefactors $\sqrt{2/3}$ and $\sqrt{1/3}$ are in a sense just a "decoration" of those branches. So according to the many worlds argument, the probability of finding yourself on either of the two should be $1/2$. But quantum mechanics indeed predicts that you find yourself in the first branch with probabilities given by the absolute square of the decorative prefactors, i.e. by $2/3$ and $1/3$. This is a well known criticism (see e.g. [13, 23, 39, 48, 62]) which seems to indicate that the probabilities in quantum theory can't be "explained away" from within the theory itself. In other words: quantum theory seems to fundamentally be a theory about probabilities.

Exercise 81

Done during lecture - not graded: *Return to your groups and discuss the following questions.*

- *Do you share the above criticism? Or are you convinced by the many worlds explanation for the probabilities in quantum mechanics? If so, what are your counter arguments to the above objections?*
- *Do you think that quantum probabilities are Frequentist or Bayesian? What are the reasons for your answer? Can quantum mechanics be the fundamental theory of nature if your answer is correct?*

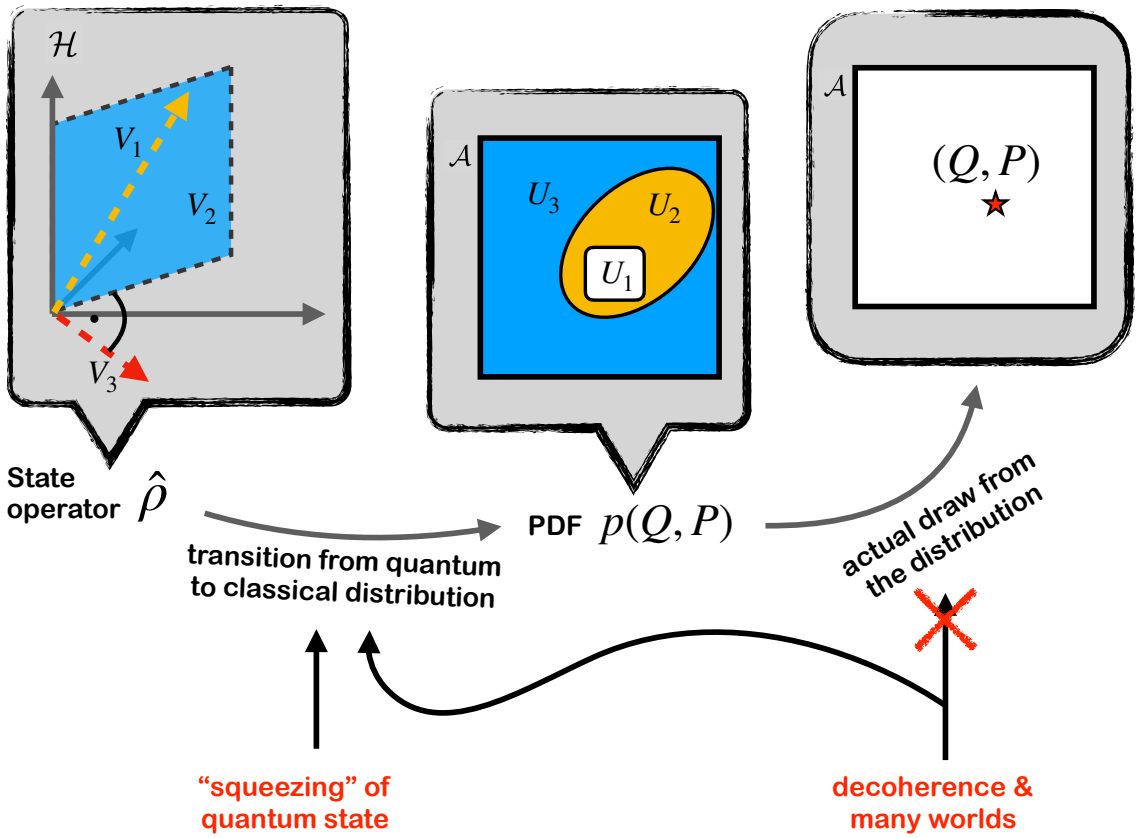

Figure 23. The quantum-to-classical transition can be split into two steps. First, we need the full quantum state $\hat{\rho}$ to evolve into something that closely resembles a classical PDF in phase-space. And in a second step, we have to end up in an actual draw from that PDF. Today, we have a very good understanding of how the first step takes place. But the second step - which is the actual measurement problem - remains unexplained. As we have tried to argue in Section 10, even the many worlds interpretation of quantum mechanics ultimately has to postulate that the prefactors in front of independent branches of the quantum state represent transition probabilities.

11 How does the Universe become classical?

In Section 8.3 we said that the distribution of density fluctuations in the very early Universe are described by a quantum field (as opposed to a classical random field). To make this statement precise we defined a field Q from the density perturbation field $\delta\rho$ and showed that this new field behaves like a collection of harmonic oscillators - one for each Fourier mode $\boldsymbol{k}$[8]. We then replaced each of these oscillators by quantum harmonic oscillators, which means that we replaces the field $Q(\boldsymbol{x})$ by an operator valued field $\hat{Q}(\boldsymbol{x})$ - cf. our notation in Section 7.6 . The eigenvectors $|Q\rangle$ of this field are a basis of the Hilbert space of our quantum field - cf. our discussion around Equation 7.58 . We then made the assumption, that the Universe starts in the ground state of the quantum Harmonic oscillators at some early moment in time.

[8] Actually the more precise statement would be: two oscillators for half of the Fourier modes. Recall that because $Q(\boldsymbol{x})$ is a real field, only half of the Fourier modes $\boldsymbol{k}$ are independent. But the field $\tilde{Q}(\boldsymbol{k})$ at those modes has both a real and imaginary part, and both of those parts behave as harmonic oscillators.

Say that we are only interested in Fourier modes with $|\boldsymbol{k}| > k_{\rm min}$, where the wave number $k_{\rm min}$ could e.g. correspond to scales much larger than our present horizon, such that we can never observe those Fourier modes of the density field. Then there is an early enough time where all modes $\boldsymbol{k}$ of interest are deep inside the Hubble horizon during inflation. Then we can assume that all the oscillators we are interested in are in their ground state. Hence, we can summarize our results for the ground state wave function of the individual modes into the wave functional

$$\Psi[Q] \;\propto\; \exp\left(-\frac{1}{4}\int\limits_{k>k_{\rm min}} \frac{{\rm d}^3k}{(2\pi)^3}\;\frac{|\tilde{Q}_{\boldsymbol{k}}|^2}{\sigma_k^2}\right) \;, \tag{11.1}$$

where the variance σ_k^2 is given by Equation 8.78 . You can think of the above functional as the ground state $|\Psi\rangle$ expressed in the basis $|Q\rangle$. Of course, in Section 7 we had identified quantum probability distributions not with Hilbert space vectors $|\Psi\rangle$ but with state matrices $\hat{\rho}$ which satisfy Equations 7.61 to 7.63 . In that language, saying that we are in the ground state of all oscillators means that we are in the state

$$\hat{\rho} = |\Psi\rangle\,\langle\Psi| \;, \tag{11.2}$$

where $|\Psi\rangle$ is the functional given above. The state $\hat{\rho}$ can be seen as a matrix, whose elements in the basis $|Q\rangle$ are given by

$$\rho(Q_1, Q_2) = \Psi[Q_1]\Psi[Q_2]^* \;. \tag{11.3}$$

At the same time, we had seen that the field Q which we "quantized" is proportional to the density contrast field $\delta \equiv \delta\rho/\bar{\rho}$ (cf. Exercise 68). We can than consider the operator valued field $\hat{\delta}(\boldsymbol{x})$ and its eigenbasis $|\delta\rangle$ instead of the basis $|Q\rangle$.

Exercise 82

NOT graded: *Show that the power spectrum of δ is given by*

$$\sigma^2_{\delta,k} = \frac{(1+\bar{w})k}{\sqrt{w_\delta}\;\bar{\rho}a^4} \;. \tag{11.4}$$

Hint: this follows from Equation 8.78 together with your result from Exercise 68 .

In terms of the basis $|\delta\rangle$ the state matrix becomes

$$\rho[\delta_1, \delta_2] = \langle\delta_1|\Psi\rangle\,\langle\Psi|\delta_2\rangle \;. \tag{11.5}$$

On the diagonal this gives

$$\mathcal{P}[\delta] \;\equiv\; \rho[\delta,\delta] \;=\; |\,\langle\delta|\Psi\rangle\,|^2 \;\propto\; \exp\left(-\frac{1}{2}\int\limits_{k>k_{\rm min}} \frac{{\rm d}^3k}{(2\pi)^3}\;\frac{|\tilde{\delta}_{\boldsymbol{k}}|^2}{\sigma^2_{\delta,k}}\right) \;, \tag{11.6}$$

where $\sigma^2_{\delta,k}$ is your result from Exercise 82 . Note the extreme analogy between this and the classical PDFunctional for a Gaussian random field in Equation 2.47 ! Given this perfect match between the classical, Gaussian PDFunctional and our quantum expression, you might ask - what even is the difference between the two situations?

<– Everything until here is relevant for the exam. Everything after this point is NOT relevant for the exam.

There are in fact two differences. Firstly, to fully characterise the behaviour of the quantum state $\hat{\rho}$ it is not sufficient to know the diagonal elements $\rho[\delta, \delta]$. To e.g. calculate the expectation value of any observable $\hat{A}$ which is not a function of the field operators $\hat{\delta}$ we need to know all off-diagonal elements $\rho[\delta_1, \delta_2]$. But the probably more surprising gap in the apparent similarity between the quantum and the classical PDF is the fact that the classical PDFunctional of Equation 2.47 is also not a complete characterisation of the statistical properties of the classical field $\delta(\boldsymbol{x})$.

Exercise 83

NOT graded: *A PDFunctional like the one in Equation 2.47 can not be a sufficient characterisation of the statistical properties of the (classical) density contrast field $\delta(\boldsymbol{x})$ - even if the field is indeed well approximated as a Gaussian random field. What is missing?*

Recall that $\delta(\boldsymbol{x})$ satisfies a second order ordinary differential equation in the limit where the Euler equation and the continuity equation are well approximated by their linearised versions, which we had combined into Equation 1.28 . A solution to such an equation is only fully determined once we specify both δ and its time derivative at some moment in time. Or, speaking within the Hamiltonian framework, we need to specify both δ and its canonically conjugate momentum in order to describe a possible phase space trajectory of our field! We saw in Section 8 that the conjugate momentum field corresponding to δ is proportional to the velocity potential ϑ . So we can alternatively characterise a phase space trajectory by specifying both δ and ϑ at a given moment in time. To fully characterise the relevant statistical properties of the cosmic density field at that time we would then need to specify a PDFunctional $\mathcal{P}[\delta, \vartheta]$, i.e. of *both* δ and ϑ .

So why did we brush over this in Section 2.2 and only consider $\mathcal{P}[\delta]$? The answer is that we ignored the decaying solution of Equation 1.28 . This equation will have two independent solutions, which we can write as

$$\delta_+(\boldsymbol{x}, t) = \delta_+(\boldsymbol{x}, t_i)\frac{D_+(t)}{D_+(t_i)} \ , \ \delta_-(\boldsymbol{x}, t) = \delta_-(\boldsymbol{x}, t_i)\frac{D_-(t)}{D_-(t_i)} \ . \tag{11.7}$$

Here D_+ and D_- are two different linear growth factors which both solve Equation 1.28 . A general solution $\delta(\boldsymbol{x}, t)$ will be the sum of these two solutions. But one of those - which we take to be δ_- - will rapidly decay, so that ultimately $\delta \approx \delta_+$. It is only in that limit that the ansatz of Equation 1.29 becomes general, such that

$$\delta(\boldsymbol{x}, t) = \delta(\boldsymbol{x}, t_i)\frac{D_+(t)}{D_+(t_i)} \ . \tag{11.8}$$

Once the decaying mode D_- is negligible there is actually a one-to-one correspondence between δ and ϑ .

Exercise 84

NOT graded: *Show that Equation 1.26 (i.e. the linearised version of the continuity equation) can be re-written as*

$$\frac{\mathrm{d}\delta}{\mathrm{d}t} + \frac{1}{a^2}\Delta_{\boldsymbol{x}}\vartheta = 0 \ , \tag{11.9}$$

where the derivative wrt. t is take at fixed co-moving coordinates $\boldsymbol{x}$. Furthermore, by assuming the ansatz of Equation 11.8, show that Equation 11.9 can be written as

$$\Delta_{\boldsymbol{x}}\vartheta = -\delta \, \frac{a^2\dot{D}_+}{D_+} \, , \tag{11.10}$$

which in Fourier space becomes

$$\tilde{\vartheta}_{\boldsymbol{k}} = \tilde{\delta}_{\boldsymbol{k}} \, \frac{a^2\dot{D}_+}{k^2 D_+} \, . \tag{11.11}$$

Your result from Exercise 84 means that in the limit where the growing mode D_+ dominates the phase-space PDFunctional $\mathcal{P}[\tilde{\delta}, \tilde{\vartheta}]$ (in Fourier space) becomes

$$\mathcal{P}[\tilde{\delta}, \tilde{\vartheta}] \;\propto\; \mathcal{P}[\tilde{\delta}] \;\; \prod_{\boldsymbol{k}} \delta_D \left(\tilde{\vartheta}_{\boldsymbol{k}} - \tilde{\delta}_{\boldsymbol{k}} \, \frac{a^2\dot{D}_+}{k^2 D_+} \right) \, , \tag{11.12}$$

i.e. it collapses along the $\tilde{\vartheta}$-direction, as indicated by the Dirac delta distributions.

In order to transition from the early-time quantum regime of the cosmic density field to the classical field we observe today, the quantum state $\hat{\rho}$ of the initial density fluctuations somehow has to start to resemble the above classical phase-space distribution. This transition, which is the topic of this section, is only one of the two steps required for the so called quantum-to-classical transition. We sketch this in Figure 23: First, we need to move from the quantum state to a classical probability distribution. And then we somehow need to obtain a draw from that classical distribution. The second step is the so-called measurement problem of quantum mechanics, and we had argued in Section 10 that it is as of yet unsolved. But at the first step in the quantum-to-classical transition is well understood. And as we will see later in this section, the reduction of the full phase-space distribution into the so called *phase-space sheet* plays a crucial role in this step.

11.1 A quantum phase-space PDF

For the rest of this discussion we will return to the fields Q and P as opposed to δ and ϑ. This is more convenient because Q and P are indeed a pair of canonically conjugate variables. But keep in mind that $Q \propto \delta$ and $P \propto \vartheta$ (though the proportionality factor between P and ϑ is indeed time dependent). To keep our discussion simple, we will focus much of the following on just one fixed Fourier mode $\boldsymbol{k}$ of the fields Q and P . Also, we will think of $\tilde{Q}_{\boldsymbol{k}}$ and $\tilde{P}_{\boldsymbol{k}}$ as individual, real degrees of freedom. In principle they will be both complex numbers, and represent two pairs of degrees of freedom. But this is just a technicality and does not qualitatively change anything in the following derivations.

To transition from the quantum distribution of the initial density fluctuations to the classical distribution we expect at later times, there should be a phase-space PDF emerging $p(\tilde{Q}_{\boldsymbol{k}}, \tilde{P}_{\boldsymbol{k}})$ emerging from the quantum state $\hat{\rho}$. This seems to be impossible, because it is one of the corner stones of quantum mechanics, that one cannot simultaneously measure conjugate variables like $\tilde{Q}_{\boldsymbol{k}}$ and $\tilde{P}_{\boldsymbol{k}}$, such that the question about their joint distribution at any given time seems to be meaningless.

What we can however define is something like a "quantum moment generating function" in phase-space. Denoting by $\hat{Q}_{\boldsymbol{k}}$ and $\hat{P}_{\boldsymbol{k}}$ the Hermitian operators whose spectrum is given by

the different values of $\tilde{Q}_{\boldsymbol{k}}$ and $\tilde{P}_{\boldsymbol{k}}$ (i.e. $\hat{Q}_{\boldsymbol{k}}$ and $\hat{P}_{\boldsymbol{k}}$ are the position and momentum operators) we can construct the quantum MGF as

$$\begin{aligned}\phi(\lambda_Q, \lambda_P) &\equiv \left\langle \exp\left(\lambda_Q \hat{Q}_{\boldsymbol{k}} + \lambda_P \hat{P}_{\boldsymbol{k}}\right)\right\rangle \\ &= \text{Trace}\left(\hat{\rho}\ \exp\left(\lambda_Q \hat{Q}_{\boldsymbol{k}} + \lambda_P \hat{P}_{\boldsymbol{k}}\right)\right)\ . \end{aligned} \tag{11.13}$$

Exercise 85

NOT graded: *Show that the operator* $\exp\left(\lambda_Q \hat{Q}_{\boldsymbol{k}} + \lambda_P \hat{P}_{\boldsymbol{k}}\right)$ *is a Hermitian operator , i.e. it represents an observable (whose expectation value in the state* $\hat{\rho}$ *is given by* $\phi(\lambda_Q, \lambda_P)$*).*

Note that the above definition of $\phi(\lambda_Q, \lambda_P)$ follows closely our definition of the MGF for classical random variables in Equation 2.11 (though that equation was for a 1-dimensional random variable). It the case where we are in a pure state, $\hat{\rho} = |\psi\rangle\langle\psi|$ the above definition reads

$$\phi(\lambda_Q, \lambda_P) \ = \ \langle\psi| \exp\left(\lambda_Q \hat{Q}_{\boldsymbol{k}} + \lambda_P \hat{P}_{\boldsymbol{k}}\right) |\psi\rangle\ . \tag{11.14}$$

For classical random variables, the MGF and the PDF are related via an inverse Laplace transform - cf. Equation 4.11 . This motivates the following definition of a "quantum phase-space PDF":

$$w(\tilde{Q}_{\boldsymbol{k}}, \tilde{P}_{\boldsymbol{k}}) \equiv \int \frac{\mathrm{d}\lambda_Q \mathrm{d}\lambda_P}{(2\pi)^2}\ \phi(i\lambda_Q, i\lambda_P)\ e^{i\lambda_Q \tilde{Q}_{\boldsymbol{k}} + i\lambda_P \tilde{P}_{\boldsymbol{k}}}\ . \tag{11.15}$$

This function is called the *Wigner function*. It is not a real PDF. In fact, it does not even have to be positive everywhere in phase space. But it is clear that in the quantum-to-classical transition it some how has to "become" the classical phase-space PDF. So let us study the evolution of the Wigner function for two toy examples that mimic important aspects of the evolution of the early-time quantum state of the cosmic density field.

11.2 Toy model 1: inverted harmonic oscillator

See lecture recording. Will add this to the script soon. (But it is NOT relevant for the exam.)

11.3 Toy model 2: oscillator with increasing mass and decreasing frequency

See lecture recording. Will add this to the script soon. (But it is NOT relevant for the exam.)

11.4 When the Wigner function behaves like a classical PDF

See my comments in the lecture recording. Will add this to the script soon. (But it is NOT relevant for the exam.)

Acknowledgments

©2022. We are indebted to the invaluable work of the teams of the public python packages `NumPy` [31], `SciPy` [58], `mpmath` [37], `Matplotlib` [34], `ChainConsumer`[9] and `exsheets`[10].

References

[1] D. Aerts and S. Pulmannová. Representation of state property systems. *Journal of Mathematical Physics*, 47(7):072105, July 2006.

[2] Ning Bao, Sean M. Carroll, and Ashmeet Singh. The Hilbert space of quantum gravity is locally finite-dimensional. *International Journal of Modern Physics D*, 26(12):1743013, January 2017.

[3] A. Barthelemy, S. Codis, C. Uhlemann, F. Bernardeau, and R. Gavazzi. A nulling strategy for modelling lensing convergence in cones with large deviation theory. *MNRAS*, 492(3):3420–3439, March 2020.

[4] Alexandre Barthelemy, Francis Bernardeau, Sandrine Codis, and Cora Uhlemann. Numerical complexity of the joint nulled weak-lensing probability distribution function. *arXiv e-prints*, page arXiv:2106.11632, June 2021.

[5] Alexandre Barthelemy, Sandrine Codis, and Francis Bernardeau. Post-Born corrections to the one-point statistics of (CMB) lensing convergence obtained via large deviation theory. *MNRAS*, 494(3):3368–3382, May 2020.

[6] Alexandre Barthelemy, Sandrine Codis, and Francis Bernardeau. Probability distribution function of the aperture mass field with large deviation theory. *MNRAS*, 503(4):5204–5222, June 2021.

[7] F. Bernardeau. The effects of smoothing on the statistical properties of large-scale cosmic fields. *A&A*, 291:697–712, November 1994.

[8] F. Bernardeau, S. Codis, and C. Pichon. The joint statistics of mildly non-linear cosmological densities and slopes in count in cells. *MNRAS*, 449:L105–L109, April 2015.

[9] F. Bernardeau, S. Colombi, E. Gaztañaga, and R. Scoccimarro. Large-scale structure of the Universe and cosmological perturbation theory. *Physics Reports*, 367:1–248, September 2002.

[10] F. Bernardeau, C. Pichon, and S. Codis. Statistics of cosmic density profiles from perturbation theory. *Phys. Rev. D*, 90(10):103519, November 2014.

[11] F. Bernardeau and P. Valageas. Construction of the one-point PDF of the local aperture mass in weak lensing maps. *A&A*, 364:1–16, December 2000.

[12] Garrett Birkhoff and John von Neumann. The logic of quantum mechanics. *Annals of Mathematics*, 37(4):823–843, 1936.

[13] Florian Boge. *On Probabilities in the Many Worlds Interpretation of Quantum Mechanics*. PhD thesis, Universität zu Köln, April 2016.

[14] Aoife Boyle, Cora Uhlemann, Oliver Friedrich, Alexandre Barthelemy, Sandrine Codis, Francis Bernardeau, Carlo Giocoli, and Marco Baldi. Nuw CDM cosmology from the weak lensing convergence PDF. *MNRAS*, May 2021.

[15] Pierre Burger, Oliver Friedrich, Joachim Harnois-Déraps, and Peter Schneider. A revised density split statistic model for general filters. *arXiv e-prints*, page arXiv:2106.13214, June 2021.

[9] https://samreay.github.io/ChainConsumer/
[10] https://bitbucket.org/cgnieder/exsheets/

[16] Lucy Calder and Ofer Lahav. Dark energy: back to Newton? *Astronomy and Geophysics*, 49(1):1.13–1.18, February 2008.

[17] Chunjun Cao, Aidan Chatwin-Davies, and Ashmeet Singh. How low can vacuum energy go when your fields are finite-dimensional? *International Journal of Modern Physics D*, 28(14):1944006, January 2019.

[18] S. Codis, C. Pichon, F. Bernardeau, C. Uhlemann, and S. Prunet. Encircling the dark: constraining dark energy via cosmic density in spheres. *MNRAS*, 460:1549–1554, August 2016.

[19] Sandrine Codis, Francis Bernardeau, and Christophe Pichon. The large-scale correlations of multicell densities and profiles: implications for cosmic variance estimates. *MNRAS*, 460(2):1598–1613, Aug 2016.

[20] DES Collaboration. Dark Energy Survey Year 3 Results: Cosmological Constraints from Galaxy Clustering and Weak Lensing. *arXiv e-prints*, page arXiv:2105.13549, May 2021.

[21] Planck Collaboration. Planck 2018 results. IX. Constraints on primordial non-Gaussianity. *A&A*, 641:A9, September 2020.

[22] Planck Collaboration. Planck 2018 results. VI. Cosmological parameters. *A&A*, 641:A6, September 2020.

[23] David Deutsch. The logic of experimental tests, particularly of Everettian quantum theory. *Studies in the History and Philosophy of Modern Physics*, 55:24–33, August 2016.

[24] Hugh Everett. "Relative State" Formulation of Quantum Mechanics. *Reviews of Modern Physics*, 29(3):454–462, July 1957.

[25] O. Friedrich, F. Andrade-Oliveira, H. Camacho, O. Alves, R. Rosenfeld, J. Sanchez, X. Fang, T. F. Eifler, E. Krause, C. Chang, Y. Omori, A. Amon, E. Baxter, J. Elvin-Poole, D. Huterer, A. Porredon, J. Prat, V. Terra, A. Troja, A. Alarcon, K. Bechtol, G. M. Bernstein, R. Buchs, A. Campos, A. Carnero Rosell, M. Carrasco Kind, R. Cawthon, A. Choi, J. Cordero, M. Crocce, C. Davis, J. DeRose, H. T. Diehl, S. Dodelson, C. Doux, A. Drlica-Wagner, F. Elsner, S. Everett, P. Fosalba, M. Gatti, G. Giannini, D. Gruen, R. A. Gruendl, I. Harrison, W. G. Hartley, B. Jain, M. Jarvis, N. MacCrann, J. McCullough, J. Muir, J. Myles, S. Pandey, M. Raveri, A. Roodman, M. Rodriguez-Monroy, E. S. Rykoff, S. Samuroff, C. Sánchez, L. F. Secco, I. Sevilla-Noarbe, E. Sheldon, M. A. Troxel, N. Weaverdyck, B. Yanny, M. Aguena, S. Avila, D. Bacon, E. Bertin, S. Bhargava, D. Brooks, D. L. Burke, J. Carretero, M. Costanzi, L. N. da Costa, M. E. S. Pereira, J. De Vicente, S. Desai, A. E. Evrard, I. Ferrero, J. Frieman, J. García-Bellido, E. Gaztanaga, D. W. Gerdes, T. Giannantonio, J. Gschwend, G. Gutierrez, S. R. Hinton, D. L. Hollowood, K. Honscheid, D. J. James, K. Kuehn, O. Lahav, M. Lima, M. A. G. Maia, F. Menanteau, R. Miquel, R. Morgan, A. Palmese, F. Paz-Chinchón, A. A. Plazas, E. Sanchez, V. Scarpine, S. Serrano, M. Soares-Santos, M. Smith, E. Suchyta, G. Tarle, D. Thomas, C. To, T. N. Varga, J. Weller, R. D. Wilkinson, R. D. Wilkinson, and DES Collaboration. Dark Energy Survey year 3 results: covariance modelling and its impact on parameter estimation and quality of fit. *MNRAS*, 508(3):3125–3165, December 2021.

[26] O. Friedrich, D. Gruen, J. DeRose, D. Kirk, E. Krause, T. McClintock, E. S. Rykoff, S. Seitz, R. H. Wechsler, and the DES Collaboration. Density split statistics: Joint model of counts and lensing in cells. *Phys. Rev. D*, 98(2):023508, July 2018.

[27] Oliver Friedrich, Anik Halder, Aoife Boyle, Cora Uhlemann, Dylan Britt, Sandrine Codis, Daniel Gruen, and ChangHoon Hahn. The PDF perspective on the tracer-matter connection: Lagrangian bias and non-Poissonian shot noise. *arXiv e-prints*, page arXiv:2107.02300, July 2021.

[28] Oliver Friedrich, Ashmeet Singh, and Olivier Doré. Toolkit for Scalar Fields in Universes with finite-dimensional Hilbert Space. *arXiv e-prints*, page arXiv:2201.08405, January 2022.

[29] Oliver Friedrich, Cora Uhlemann, Francisco Villaescusa-Navarro, Tobias Baldauf, Marc Manera, and Takahiro Nishimichi. Primordial non-Gaussianity without tails - how to measure f_{NL} with the bulk of the density PDF. *MNRAS*, August 2020.

[30] D. Gruen, O. Friedrich, E. Krause, J. DeRose, R. Cawthon, C. Davis, J. Elvin-Poole, E. S. Rykoff, R. H. Wechsler, A. Alarcon, G. M. Bernstein, J. Blazek, C. Chang, J. Clampitt, M. Crocce, J. De Vicente, M. Gatti, M. S. S. Gill, W. G. Hartley, S. Hilbert, B. Hoyle, B. Jain, M. Jarvis, O. Lahav, N. MacCrann, T. McClintock, J. Prat, R. P. Rollins, A. J. Ross, E. Rozo, S. Samuroff, C. Sánchez, E. Sheldon, M. A. Troxel, J. Zuntz, T. M. C. Abbott, F. B. Abdalla, S. Allam, J. Annis, K. Bechtol, A. Benoit-Lévy, E. Bertin, S. L. Bridle, D. Brooks, E. Buckley-Geer, A. Carnero Rosell, M. Carrasco Kind, J. Carretero, C. E. Cunha, C. B. D'Andrea, L. N. da Costa, S. Desai, H. T. Diehl, J. P. Dietrich, P. Doel, A. Drlica-Wagner, E. Fernandez, B. Flaugher, P. Fosalba, J. Frieman, J. García-Bellido, E. Gaztanaga, T. Giannantonio, R. A. Gruendl, J. Gschwend, G. Gutierrez, K. Honscheid, D. J. James, T. Jeltema, K. Kuehn, N. Kuropatkin, M. Lima, M. March, J. L. Marshall, P. Martini, P. Melchior, F. Menanteau, R. Miquel, J. J. Mohr, A. A. Plazas, A. Roodman, E. Sanchez, V. Scarpine, M. Schubnell, I. Sevilla-Noarbe, M. Smith, R. C. Smith, M. Soares-Santos, F. Sobreira, M. E. C. Swanson, G. Tarle, D. Thomas, V. Vikram, A. R. Walker, J. Weller, Y. Zhang, and DES Collaboration. Density split statistics: Cosmological constraints from counts and lensing in cells in DES Y1 and SDSS data. *Phys. Rev. D*, 98(2):023507, July 2018.

[31] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, September 2020.

[32] J. Hartlap, P. Simon, and P. Schneider. Why your model parameter confidences might be too optimistic. Unbiased estimation of the inverse covariance matrix. *A&A*, 464:399–404, March 2007.

[33] J. Y. Haw, S. M. Assad, A. M. Lance, N. H. Y. Ng, V. Sharma, P. K. Lam, and T. Symul. Maximization of extractable randomness in a quantum random-number generator. *Phys. Rev. Applied*, 3:054004, May 2015.

[34] J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95, 2007.

[35] Alan Hájek. Interpretations of Probability. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Fall 2019 edition, 2019.

[36] Mikhail M. Ivanov, Alexander A. Kaurov, and Sergey Sibiryakov. Non-perturbative probability distribution function for cosmological counts in cells. *JCAP*, 2019(3):009, Mar 2019.

[37] Fredrik Johansson et al. *mpmath: a Python library for arbitrary-precision floating-point arithmetic (version 0.18)*, December 2013. `http://mpmath.org/`.

[38] G. M. Kaufman. Some bayesian moment formulae. *Report No. 6710, Center for Operations Research and Econometrics, Catholic University of Louvain, Heverlee, Belgium*, 1967.

[39] Adrian Kent. Against Many-Worlds Interpretations. *International Journal of Modern Physics A*, 5(9):1745–1762, January 1990.

[40] Juan Maldacena. Non-gaussian features of primordial fluctuations in single field inflationary models. *Journal of High Energy Physics*, 2003(5):013, May 2003.

[41] V. Mukhanov. *Physical Foundations of Cosmology*. Cambridge University Press, November 2005.

[42] T. Padmanabhan. *Gravitation: Foundations and Frontiers*. Cambridge University Press, January 2010.

[43] Constantin Piron. Axiomatique quantique. *Helvetica physica acta*, 37(4-5):439, 1964.

[44] Itamar Pitowsky. Quantum mechanics as a theory of probability. *arXiv e-prints*, pages quant–ph/0510095, October 2005.

[45] Frank P. Ramsey. Truth and probability. In R. B. Braithwaite, editor, *The Foundations of Mathematics and other Logical Essays*, chapter 7, pages 156–198. McMaster University Archive for the History of Economic Thought, 1926.

[46] Adam G. Riess, Alexei V. Filippenko, Peter Challis, Alejandro Clocchiatti, Alan Diercks, Peter M. Garnavich, Ron L. Gilliland, Craig J. Hogan, Saurabh Jha, Robert P. Kirshner, B. Leibundgut, M. M. Phillips, David Reiss, Brian P. Schmidt, Robert A. Schommer, R. Chris Smith, J. Spyromilio, Christopher Stubbs, Nicholas B. Suntzeff, and John Tonry. Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant. *AJ*, 116(3):1009–1038, September 1998.

[47] Max Tegmark. The Interpretation of Quantum Mechanics: Many Worlds or Many Words? *Fortschritte der Physik*, 46(6-8):855–862, January 1998.

[48] Max Tegmark. Many Worlds in Context. *arXiv e-prints*, page arXiv:0905.2182, May 2009.

[49] C. Uhlemann, S. Codis, C. Pichon, F. Bernardeau, and P. Reimberg. Back in the saddle: large-deviation statistics of the cosmic log-density field. *MNRAS*, 460(2):1529–1541, Aug 2016.

[50] C. Uhlemann, M. Feix, S. Codis, C. Pichon, F. Bernardeau, B. L'Huillier, J. Kim, S. E. Hong, C. Laigle, C. Park, J. Shin, and D. Pogosyan. A question of separation: disentangling tracer bias and gravitational non-linearity with counts-in-cells statistics. *MNRAS*, 473:5098–5112, February 2018.

[51] C. Uhlemann, E. Pajer, C. Pichon, T. Nishimichi, S. Codis, and F. Bernardeau. Hunting high and low: disentangling primordial and late-time non-Gaussianity with cosmic densities in spheres. *MNRAS*, 474:2853–2870, March 2018.

[52] C. Uhlemann, C. Pichon, S. Codis, B. L'Huillier, J. Kim, F. Bernardeau, C. Park, and S. Prunet. Cylinders out of a top hat: counts-in-cells for projected densities. *MNRAS*, 477:2772–2785, June 2018.

[53] P. Valageas. Dynamics of gravitational clustering. 3. The quasi-linear regime for some non-gaussian initial conditions. *Astron. Astrophys.*, 382:431, 2002.

[54] P. Valageas. Dynamics of gravitational clustering. II. Steepest-descent method for the quasi-linear regime. *A&A*, 382:412–430, February 2002.

[55] P. Valageas. Dynamics of gravitational clustering. V. Subleading corrections in the quasi-linear regime. *A&A*, 382:477–487, February 2002.

[56] V.S. Varadarajan. *Geometry of Quantum Theory: Second Edition*. Springer New York, 2007.

[57] Susan Vineberg. Dutch Book Arguments. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Summer 2022 edition, 2022.

[58] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020.

[59] John von Neumann, Oskar Morgenstern, and Ariel Rubinstein. *Theory of Games and Economic Behavior (60th Anniversary Commemorative Edition)*. Princeton University Press, 1944.

[60] Carl Friedrich von Weizsäcker. *Zeit und Wissen*. Carl Hanser Verlag, 1992.

[61] Carl Friedrich von Weizsäcker, Thomas Görnitz, and Holger Lyre. *The Structure of Physics*, volume 155. 2006.

[62] Wojciech Hubert Zurek. Probabilities from entanglement, Born's rule $p_k = | \psi_k |^2$ from envariance. *Phys. Rev. A*, 71(5):052105, May 2005.

A relativistic equations

I'M SORRY - THIS IS STILL A MESS... WILL CLEAN UP SOON.

$$\nabla_\mu T^{\mu\nu} = \partial_\mu T^{\mu\nu} + \Gamma^\nu_{\kappa\mu} T^{\kappa\mu} + \Gamma^\mu_{\kappa\mu} T^{\kappa\nu} \tag{A.1}$$

For now in conformal time, in the conformal-Newtonian gauge (and at linear order in ϕ):

$$\Gamma^0_{00} = \mathcal{H} + \phi' \tag{A.2}$$
$$\Gamma^0_{0i} = \partial_i \phi \tag{A.3}$$
$$\Gamma^0_{ij} = \delta_{ij}\left(\mathcal{H} - 4\phi\mathcal{H} - \phi'\right) \tag{A.4}$$
$$\Gamma^i_{00} = \partial_i \phi \tag{A.5}$$
$$\Gamma^j_{0i} = \delta_{ij}\left(\mathcal{H} - \phi'\right) \tag{A.6}$$
$$\Gamma^k_{ij} = \delta_{ij}\partial_k\phi - \delta_{ik}\partial_j\phi - \delta_{jk}\partial_i\phi \ . \tag{A.7}$$

The four-velocity is

$$u^\mu = \frac{\mathrm{d}x^\mu}{\mathrm{d}\tau} \tag{A.8}$$
$$= \frac{\mathrm{d}\eta}{\mathrm{d}\tau}\frac{\mathrm{d}x^\mu}{\mathrm{d}\eta} \tag{A.9}$$
$$= \frac{1}{a\sqrt{(1+2\phi) + (1-2\phi)\left|\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}\eta}\right|^2}}\frac{\mathrm{d}x^\mu}{\mathrm{d}\eta} \tag{A.10}$$
$$\Rightarrow u^0 \approx \frac{1}{a}(1-\phi) \tag{A.11}$$
$$\Rightarrow u^i \approx \frac{1}{a}\frac{\mathrm{d}x^i}{\mathrm{d}\eta} = \frac{1}{a}v^i \ . \tag{A.12}$$

The energy momentum tensor is given by

$$T^{00} \approx \frac{1}{a^2}(\rho + p)(1-2\phi) - \frac{1}{a^2}(1-2\phi)p \tag{A.13}$$
$$\approx \frac{1}{a^2}(\bar\rho + \delta\rho)(1-2\phi) \tag{A.14}$$
$$\approx \frac{1}{a^2}(\bar\rho - 2\bar\rho\phi + \delta\rho) \tag{A.15}$$
$$T^{0i} \approx \frac{1}{a}(\rho + p)(1-\phi)u^i \tag{A.16}$$
$$\approx \frac{1}{a}(\bar\rho + \bar p)u^i \tag{A.17}$$
$$T^{ij} = (\rho + p)u^i u^j - g^{ij}p \tag{A.18}$$
$$\approx \frac{1}{a^2}(1+2\phi)(\bar p + \delta p) \tag{A.19}$$
$$\approx \frac{1}{a^2}(\bar p + \delta p + 2\phi\bar p) \tag{A.20}$$

The continuity equation reads

$$\nabla_\mu T^{\mu 0} = \partial_\mu T^{\mu 0} + \Gamma^0_{\kappa\mu} T^{\kappa\mu} + \Gamma^\mu_{\kappa\mu} T^{\kappa 0} \tag{A.21}$$

$$\tag{A.22}$$

The Einstein equations at linear order in the perturbations become (cf. [41])

$$\Delta\phi - 3\mathcal{H}(\phi' + \mathcal{H}\phi) = 4\pi a^2 \delta\rho \tag{A.23}$$

$$(a\phi)'_{,i} = 4\pi a^2 (\bar{\rho} + \bar{p}) \delta u_i \tag{A.24}$$

$$\phi'' + 3\mathcal{H}\phi' + (2\mathcal{H}' + \mathcal{H}^2)\phi = 4\pi a^2 \delta p \ . \tag{A.25}$$

Note that $\delta u_i = -a^2 \delta u^i = -a v^i = -\vartheta_{,i}$. So if both ϕ and ϑ are perturbations with an expectation value of 0 , then we can conclude that

$$(a\phi)' = -4\pi a^2 (\bar{\rho} + \bar{p}) \vartheta \ . \tag{A.26}$$

Changing from conformal time η to physical time t the above equations become

$$\frac{1}{a^2}\Delta\phi - 3H\frac{1}{a}\partial_t(a\phi) = 4\pi\delta\rho \tag{A.27}$$

$$\frac{1}{a}\partial_t(a\phi) = \dot{H}\vartheta \tag{A.28}$$

$$\ddot{\phi} + 4H\dot{\phi} + \left(2\dot{H} + 3H^2\right)\phi = 4\pi\delta p \ . \tag{A.29}$$

We can write the last equation as

$$4\pi\delta p = \partial_t\left(\frac{1}{a}\partial_t(a\phi)\right) + 3H\dot{\phi} + \left(\dot{H} + 3H^2\right)\phi \tag{A.30}$$

$$= \partial_t\left(\frac{1}{a}\partial_t(a\phi)\right) + 3H\frac{1}{a}\partial_t(a\phi) + \dot{H}\phi \tag{A.31}$$

$$= \partial_t\left(\dot{H}\vartheta\right) + 3H\dot{H}\vartheta + \dot{H}\phi \ . \tag{A.32}$$

Using this we arrive at

$$\frac{1}{a^2}\Delta\phi = 4\pi\delta\rho + 3H\dot{H}\vartheta \tag{A.33}$$

$$\frac{1}{a}\partial_t(a\phi) = \dot{H}\vartheta \tag{A.34}$$

$$\partial_t\left(\dot{H}\vartheta\right) + 3H\dot{H}\vartheta + \dot{H}\phi = 4\pi\delta p \tag{A.35}$$

$$\Rightarrow \dot{\vartheta} + \left(\frac{\ddot{H}}{\dot{H}} + 3H\right)\vartheta + \phi = \frac{4\pi}{\dot{H}}\delta p \ . \tag{A.36}$$

FROM HERE I ASSUME EOM WITH $\bar{w}$:

$$\Rightarrow \dot{\vartheta} - 3\bar{w}H\vartheta + \phi = \frac{4\pi w_\delta}{\dot{H}}\delta\rho \ . \tag{A.37}$$

A.1 $k/a \ll H$

$$0 \approx 4\pi\delta\rho + 3H\dot{H}\vartheta \tag{A.38}$$
$$= \frac{3}{2}H^2\delta - \frac{9}{2}(1+\bar{w})H^3\vartheta \tag{A.39}$$
$$\Rightarrow \delta \approx 3(1+\bar{w})H\vartheta \tag{A.40}$$
$$\Rightarrow \vartheta \approx \frac{\delta}{3(1+\bar{w})H} \tag{A.41}$$
$$\Rightarrow \frac{1}{a}\partial_t(a\phi) \approx \frac{\dot{H}}{3(1+\bar{w})H}\ \delta \tag{A.42}$$
$$= -\frac{H}{2}\ \delta \tag{A.43}$$
$$\tag{A.44}$$

$$0 = \dot{\vartheta} - 3\bar{w}H\vartheta + \phi - \frac{4\pi w_\delta}{\dot{H}}\delta\rho \tag{A.45}$$
$$= \dot{\vartheta} - 3\bar{w}H\vartheta + \phi - \frac{3}{2}\frac{w_\delta H^2}{\dot{H}}\delta \tag{A.46}$$
$$= \dot{\vartheta} - 3\bar{w}H\vartheta + \phi + \frac{w_\delta}{1+\bar{w}}\delta \tag{A.47}$$
$$\Rightarrow \frac{H}{2}\delta = \frac{1}{a}\partial_t\left(a\dot{\vartheta} - 3a\bar{w}H\vartheta + \frac{aw_\delta}{1+\bar{w}}\delta\right) \tag{A.48}$$
$$= \frac{1}{a}\partial_t\left(\frac{a\dot{\delta}}{3(1+\bar{w})H} - \frac{a\delta}{3(1+\bar{w})H^2}\dot{H} - \frac{a\bar{w}}{1+\bar{w}}\delta + \frac{aw_\delta}{1+\bar{w}}\delta\right) \tag{A.49}$$
$$= \frac{1}{a}\partial_t\left(\frac{a\dot{\delta}}{3(1+\bar{w})H} + \frac{a}{2}\delta + \frac{a(w_\delta-\bar{w})}{1+\bar{w}}\delta\right) \tag{A.50}$$
$$\Rightarrow 0 = \frac{\dot{\delta}}{3(1+\bar{w})} + \frac{(w_\delta-\bar{w})}{1+\bar{w}}H\delta + \frac{\ddot{\delta}}{3(1+\bar{w})H} - \frac{\dot{\delta}}{3(1+\bar{w})H^2}\dot{H} \tag{A.51}$$
$$+ \frac{1}{2}\dot{\delta} + \frac{(w_\delta-\bar{w})}{1+\bar{w}}\dot{\delta} \tag{A.52}$$
$$= \frac{\dot{\delta}}{3(1+\bar{w})} + \frac{(w_\delta-\bar{w})}{1+\bar{w}}H\delta + \frac{\ddot{\delta}}{3(1+\bar{w})H} + \dot{\delta} + \frac{(w_\delta-\bar{w})}{1+\bar{w}}\dot{\delta} \tag{A.53}$$
$$\Rightarrow 0 = \ddot{\delta} + H\dot{\delta} + 3(w_\delta-\bar{w})H^2\delta + 3(1+\bar{w})H\dot{\delta} + 3(w_\delta-\bar{w})H\dot{\delta} \tag{A.54}$$
$$= \ddot{\delta} + (4+3w_\delta)H\dot{\delta} + 3(w_\delta-\bar{w})H^2\delta \tag{A.55}$$
$$\tag{A.56}$$