Research notes

All here is public

1 Statistics

1.1 Introduction to Cosmology

With Hubble's discovery of the expanding universe, there has been great efforts to understand this expansion and the evolution of this expansion (and of the universe as a whole). A related idea, General Relativity has described the importance of the geometry of the universe. There are, in general, three possible geometries:

- A flat geometry which is equivalent to Euclidean space with zero curvature.
- An open geometry with constant negative curvature (Anti-de Sitter space).
- A Closed geometry with constant positive curvature (de Sitter space)

Given the relationship between curvature and the energy-momentum tensor given by the Einstein field equations, it seems reasonable to assume that any non-zero energy means the universe is not flat. Through quantum field theory (QFT) the Casimir effect shows that the vacuum has non-zero energy density, and through astronomical observation of distant galaxies it can be seen that our universe is flat. At first these two observations contradict each other. However, by introducing another term to the Einstein field equations, this discrepency can be resolved.

$$G_{\mu\nu} - \underbrace{\Lambda g_{\mu\nu}}_{\text{new term}} = T_{\mu\nu}$$

The constant Λ is called the *cosmological constant* which can absorb the contributions from the vacuum energy, allowing for a flat universe and non-zero vacuum energy density. A natural question to ask is 'what is the source of the cosmological constant?'

1.2 ΛCDM

1.3 The (Affine) Geodesic Equation

Suppose we have a (semi-) Riemannian manifold M with metric g and tangent bundle TM. An *affine connection* is a map

$$\Gamma(TM) \times \Gamma(TM) \to \Gamma(TM)$$
$$(X, Y) \to \nabla_X Y$$

That is, it parallel transports the vector field Y along the connection ∇ in the direction of vector field X. From this, we can write the affine geodesic equation for a path $\gamma(t)$

$$\nabla_{\dot\gamma}\dot\gamma(t) = 0$$

Thus, a geodesic is a path such that its tangent vector is parallel translated. Since we observe our world with coordinates, in physics it is often more instructive to work this out in a specific set of coordinates x^μ. Thus this can be written as

$$\ddot\gamma^\mu + \Gamma^\mu_{\rho\lambda}\dot\gamma^\rho\dot\gamma^\lambda$$

1.4 The FLRW Metric and Dark Energy

In general, metrics allow one to attribute distances between points in a space as $d = g_{\mu\nu}x^\mu x^\nu$. The 'flat' metric for relativity is the *Minkowski metric* given by $\mathrm{diag}(-1, 1, 1, 1)$, however, a metric to describe the expanding universe is given by the *FLRW metric* by $d^2 = t^2 - a^2(t)s^2$ with s^2 the standard Euclidean distance in $\mathbb{R}^3$. What is immediately appeerent from the FLRW metric is that spacial slices of the $d = 4$ spacetime remain curvature free under the expansion of the universe describe by the scale factor $a^2(t)$.

In cosmology, there are other distances which can be more useful than the distance given by the FLRW metric. In the FLRW metric the distance between two points grows in time. We can avoid this by defining the *comoving distance* in which distances remain fixed through time. If we look at a coordinate function x^μ at $t = t_0$, at a later time the coordinate function can be written as $x^\mu \to a(t)x^\mu$, thus by dividing by the scale factor $a(t)$ we can define the comoving coordinates as

$$\chi = \int_{t_0}^{t} \frac{1}{a(t')} dt' \tag{1}$$

with the standard Minkowski metric. This can be taken a step further by determining how far light has travelled since $t = 0$

$$\eta = \int_0^t \frac{1}{a(t')} dt' \tag{2}$$

Since we can't see anything beyond this distance, it is often called the *comoving horizon*. There is one last useful distance to define, the *angular distance* which is inferred by the angle subtended by two objects. This relates distances to the geometry discuss in the first section where the measured distance will be the radial distance D

$$D_A = \begin{cases} R & K = 0 \\ R\sin(D/R) & K > 0 \\ R\sinh(D/R) & K < 0 \end{cases} \tag{3}$$

When describing the structure of the universe, I will make a few assumptions (which hold up to small perturbations):

- Homogeneity. The cosmology describing the universe does not depend on location.
- Isotropy. The cosmology describing the universe does not depend on location.

These two conditions for what is sometimes referred to as *the cosmological principle*. In general, they don't hold on small scales, however averaging over a sufficiently large distance these assumptions give an accurate description. The isotropy condition means that the universe should have 0 net momentum, and by assumming the universe is smooth the energy-momentum tensor can be written

$$T^{\mu}_{\nu} = \begin{pmatrix} -\mathcal{E} & 0 & 0 & 0 \\ 0 & \mathcal{P} & 0 & 0 \\ 0 & 0 & \mathcal{P} & 0 \\ 0 & 0 & 0 & \mathcal{P} \end{pmatrix} \tag{4}$$

The usual conservation law holds

$$\nabla_{\mu} T^{\mu}_{\nu} = 0 \Rightarrow \partial_t \mathcal{E} + \frac{\dot{a}}{a}(3\mathcal{E} + 3\mathcal{P}) = 0 \tag{5}$$

We can use the geodesic equation to examine how the energy of the massless particles evolves through time.

1.5 Einstein's Field Equation from the Bianchi Identity

1.6 Dynamics of the FRLW Metric

If we examine the 00 component of Einstein's Field Equation, the result is the *first Friedman equation* (note that ρ is shorthand for $\sum_i \rho_i$ and $R(G)$ to note that R depends on the geometry of spaceial slices)

$$H^2(a) + \frac{1}{a^2 R^2(G)} = \frac{8\pi G}{3}\rho \tag{6}$$

We can interpret the second term on the left (which is spacial curvature) as some density asociated with the curvature ρ_k. If we divide by ρ_{crit} at $z = 0$ / $a = 1$ we find the usual form.

$$\omega + \omega_k = 1 \tag{7}$$

The second Friedman equation comes from the trace of Einstein's equation.

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3P) \tag{8}$$

2 Computing

2.1 Markov Chain Monte Carlo (MCMC)

Interestingly, MCMC algorithms have heavy analogies with statistical mechanics which are useful to demonstrate the concept. To examine this, lets first define what a Markov Chain is.

Definition 1. *A sequence $X_1, \ldots, X_n$ of random elements is a Markov Chain if the conditional distribution X_{n+1} depends only on X_n. The set in which X_i take values is called the state space of the chain.*

2.2 The Metropolis-Hastings Algorithm

Suppose we want to sample from a distribution $p(x)$. $p(x)$ can be high dimensional and is generally difficult to calculate (evidence is hard to compute since it requires integration over the entire parameter space). The goal is to use Markov Chains to sample from $p(x)$ without needing to compute the evidence. This will be represented as a path through state space until the chain reaches a stable point (stationary state).

We start with a proposal distribution $g(x_n)$. Sample from the proposal distribution to find the next state x_{n+1} with probability $g(x_{n+1}|x_n)$. This transition from state n to state $n+1$ must follow the *detailed balance condition*

$$p(x_n)g(x_{n+1}|x_n)A(x_n \to x_{n+1}) = p(x_{n+1})g(x_n|x_{n+1})A(x_{n+1} \to x_n)$$

where A is an *acceptance probability* which I will define more precisely later. Using Bayes' Theorem on $p(x)$, the evidence cancels out on each side, and thus the detailed balance condition can be simplified to only rely on the likelihood and prior of $p(x)$, which I will denote π and $\mathcal{L}$.

$$\pi(x_n)\mathcal{L}(x_n)g(x_{n+1}|x_n)A(x_n \to x_{n+1}) = \pi(x_{n+1})\mathcal{L}(x_{n+1})g(x_n|x_{n+1})A(x_{n+1} \to x_n)$$

$$\Rightarrow \frac{A(x_n \to x_{n+1})}{A(x_{n+1} \to x_n)} = \frac{\pi(x_{n+1})\mathcal{L}(x_{n+1})g(x_n|x_{n+1})}{\pi(x_n)\mathcal{L}(x_n)g(x_{n+1}|x_n)} \equiv R_{n,n+1}$$

This allows us to define the acceptance probability as

$$A(x_n \to x_{n+1}) = \min(1, R_{n,n+1})$$

This probability is used to determine whether the chain moves to x_{n+1} or stays at x_n. The chain converges when it reaches a stationary state.

There are a few properties that can be observed for this algorithm:

- Having an asymmetrical proposal $g(x)$ can allow for faster convergence of the chain.
- The initial sampling may not accurately reflect samples for $p(x)$. This is regarded as the 'burn-in' and is generally discarded from the samples.
- MCMC Sampling loses sampling power for multi-modal distributions.

2.3 Parameter Difference Distribution

Define the posterior probability of parameters θ within model $\mathcal{M}$ given data d as (Bayes' theorem)

$$\mathcal{P}(\theta) := P(\theta|d, \mathcal{M}) = \frac{P(\theta|\mathcal{M})P(d|\theta, \mathcal{M})}{P(d|\mathcal{M})} \tag{9}$$

The following notation will simplify this expression

$$\Pi(\theta) = P(\theta|\mathcal{M}), \text{ Prior probability}$$
$$\mathcal{L}(\theta) = P(d|\theta, \mathcal{M}), \text{ the likelihood}$$
$$\mathcal{E} = P(d|\mathcal{M}), \text{ the evidence}$$

$$\mathcal{P}(\theta) := P(\theta|d, \mathcal{M}) = \frac{\Pi(\theta)\mathcal{L}}{\mathcal{E}}$$

These are (maybe?) understandable as the prior probability is the parameters given the model, the likelihood as the data given some parameters of the model, and the evidence as the data give a model. From now on, the model $\mathcal{M}$ will be implied.

Consider now two data sets d_1, d_2. They have a joint likelihood

$$\mathcal{L}(\theta) = P(d_1, d_2|\theta)$$

This **quantifies the likelihood of d_1 and d_2 coming from the same set of parameters of a given model.** Denote now $\mathcal{L}_1, \mathcal{P}_1$ as the marginalized likelihood and Posterior over d_2. (**Is there a procedure/algorithm to find the marginalized probability? Seems generally non-trivial**).

Now duplicate the parameter set to θ_1, θ_2 and perscibe a new joint likelihood $\mathcal{L}(\theta_1, \theta_2) = P(d_1, d_2|\theta_1, \theta_2)$. **This is generally a choice, but how can one make a good choice?** The choice however is not unique.

As such, the constraints are imposed:

1. $\mathcal{L}(\theta_1 = \theta, \theta_2 = \theta) = \mathcal{L}(\theta)$, or in plain english, the joint likelihood if $\theta_1 = \theta = \theta_2$ needs to coincide with the likelihood of θ
2. $P(d_1|\theta_1, \theta_2) = P(d_1|\theta_1)$ once marginalized over d_2. Marginalizing the likelihood over one of the data sets removes the dependancy on the corresponding parameter sets.

This ensures the datasets are conditionally independent,

$$P(d_1, d_2|\theta) = P(d_1|\theta)P(d_2|\theta)$$

Proof? Which means we can choose

$$\mathcal{L}(\theta_1, \theta_2) = \mathcal{L}_1(\theta_1)\mathcal{L}_2(\theta_2)$$

If we further assume the prior distribution can be factorized, the joint posterior is then

$$\mathcal{P}(\theta_1, \theta_2) = \mathcal{L}(\theta_1, \theta_2)\Pi(\theta_1)\Pi(\theta_2)$$

Defining $\Delta\theta = \theta_1 - \theta_2$, the parameter difference posterior is given by

$$\mathcal{P}(\Delta\theta) = \int_{V_\Pi} \mathcal{P}(\theta, \theta - \Delta\theta)d\theta$$

2.4 Quantifying Results

Given some probability P of a parameter shift, the following formula can give you the number of standard deviations if the probability shift comes from a gaussian distribution

$$n_\sigma = \sqrt{2}\text{Erf}^{-1}(P)$$

I have a notebook using two unit gaussian priors separated by a distance a. This example can be computed analytically.

$$\begin{aligned}
\mathcal{P}(\Delta\theta) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\theta^2/2}e^{-(\theta-\Delta\theta)^2/2}d\theta \\
&= \frac{1}{2\pi}\cdot\sqrt{\pi}e^{-(\Delta\theta)^2/4} \\
&= \frac{1}{\sqrt{4\pi}}e^{-(\Delta\theta)^2/4}
\end{aligned}$$

The parameter difference posterior is a gaussian with standard deviation $\sqrt{2}$. The separation is fixed by a, hence the shift is $\mathcal{P}(a)$. Hence the shift probability is

$$\Delta = \int_{-a}^{a} e^{-(\Delta\theta)^2/4}d\Delta\theta$$

Lets use the example $a = 2$. Then $n_\sigma = 2/\sqrt{2} = \sqrt{2}$. Using this we can work backwards to find Δ from a z-table to find $\Delta = 0.9207 - 0.0793 = 0.8414$.

2.5 Normalizing Flows

The method of normalizing flows (MAF) implemented here uses Masked Autoencoders (MADE) to construct the flow. Suppose we have an input to the flow x_i. The output of the map is $y_i = \mu(x_{1:i-1}) + \sigma(x_{1:i-1})x_i$. The μ and σ are found using neural networks which recieve masked inputs $x_{1:i-1} = (x_1, \ldots, x_{i-1}, 0, \ldots, 0)$. Since the input only depends on the first $i-1$ inputs, the normalizing flow is *autoregressive* and the Jacobian is triangular.

The implementation in tensorflow uses *bijectors* which implements a local diffeomorphism between a manifold M and a target manifold N (which are our parameter spaces), i.e. $\phi : M \to N$ such that ϕ is differentiable and injective. In tensorflow it has three operations, Forward, Inverse, and log_deg_jacobian, which are exactly the three we want. By constructing a bijector for each masked input, the full normalizing map can be constructed.

References

https://www.colorado.edu/amath/sites/default/files/attached-files/2_28_2018.pdf

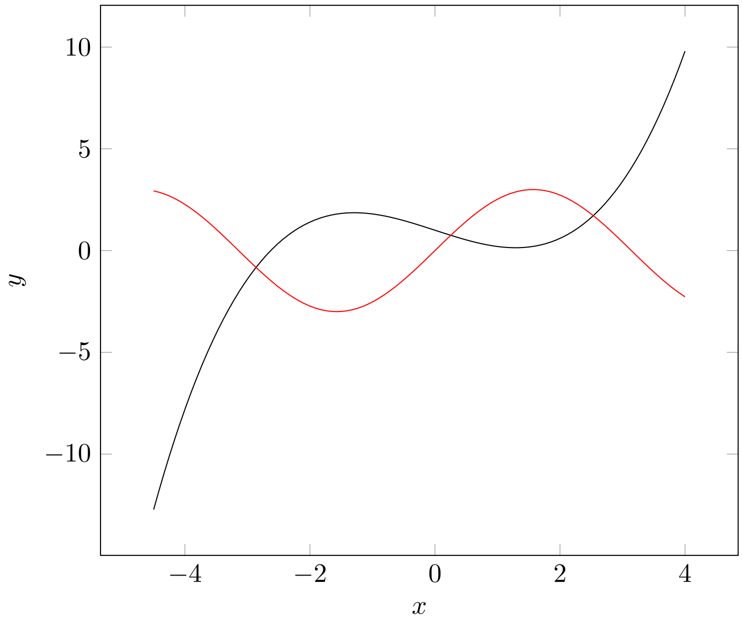
https://si.biostat.washington.edu/sites/default/files/modules/Geyer-Introduction%20to%20markov%20chain%20Monte%20Carlo_0.pdf

https://towardsdatascience.com/monte-carlo-markov-chain-mcmc-explained-94e3a6c8de11

https://www.sheffield.ac.uk/polopoly_fs/1.60510!/file/MCMC.pdf

---

Test graph



3 Fields

3.1 Multifield Dark Energy

So I don't keep having to look at this

Definition 2. *Given a semi-riemannian manifold M with metric g, the christoffel symbols are given by*

$$\nabla_a \partial_b = \Gamma^c_{ab} \partial_c$$

$$\partial_a g_{bc} + \partial_b g_{ca} - \partial_c g_{ab} = 2 g_{dc} \Gamma^d_{ab}$$

Consider a metric of the form $\mathrm{diag}(-1, a^2(t))$, so its determinant is $-a^6$. The action is

$$S = \int d^4x \sqrt{-g} \left[\frac{1}{2} M_p^2 R - \frac{1}{2} \gamma_{ab} \partial_\mu \phi^a \partial^\mu \phi^b - V(\phi) + \mathcal{L}_m \right]$$

I only want to describe a homogeneous background, so the field is only a function of time. The action becomes

$$S = \int d^4x \sqrt{-g} \left[\frac{1}{2} M_p^2 R - \frac{1}{2} \gamma_{ab} \dot{\phi}^a \dot{\phi}^b - V(\phi) + \mathcal{L}_m \right]$$

Varying the field gives

$$\delta S = \int d^4x \sqrt{-g} \left[-\frac{1}{2} \partial_a (\gamma_{bc}) \dot{\phi}^b \dot{\phi}^c - \frac{1}{2} \gamma_{ab} \frac{d}{dt} (\delta \phi^a) \dot{\phi}^b - \partial_a V \delta \phi^a \right]$$

$$\delta S = -\frac{1}{2} (\delta \phi^a)(\sqrt{-g} \gamma_{ab} \dot{\phi}^b + \int d^4x \sqrt{-g} \left[-\frac{1}{2} \gamma(\nabla_a \partial_b, \partial_c) \dot{\phi}^b, \dot{\phi}^c - \frac{1}{2} \gamma(\partial_b, \nabla_a \partial_c) \dot{\phi}^b \dot{\phi}^c + 3 \frac{\dot{a}}{a} \gamma_{ab} \delta \phi^a \dot{\phi}^b + \gamma_{ab} \delta \phi^a \ddot{\phi}^b - V_a \delta \phi^a \right]$$

$$\delta S = -\int d^4x \sqrt{-g} \left[\frac{1}{2} \Gamma^d_{ab} \gamma_{dc} \dot{\phi}^b \dot{\phi}^c + \frac{1}{2} \Gamma^d_{ac} \gamma_{bd} \dot{\phi}^b \dot{\phi}^c + 3H \gamma_{ab} \dot{\phi}^b + \gamma_{ab} \ddot{\phi}^b + V_a \right]$$

Multiply everything by γ^{aa}. Now lets do some reshuffling of the indices.

$$\gamma^{aa} \Gamma^d_{ab} \gamma_{dc} \dot{\phi}^b \dot{\phi}^c$$

$$a \leftrightarrow d$$
$$\gamma^{ad}\Gamma^a_{db}\gamma_{ac}\dot\phi^b\dot\phi^c$$
$$b \leftrightarrow d$$
$$\gamma^{ad}\Gamma^a_{bd}\gamma_{ac}\dot\phi^b\dot\phi^c$$
$$\gamma^{ab}\Gamma^a_{bd}\gamma_{ac}\dot\phi^d\dot\phi^c$$
$$c \leftrightarrow d$$
$$\gamma^{ab}\Gamma^a_{bc}\gamma_{ad}\dot\phi^d\dot\phi^c = \Gamma^a_{bc}\dot\phi^b\dot\phi^c$$

and

$$\gamma^{aa}\Gamma^d_{ac}\gamma_{bd}\dot\phi^b\dot\phi^c$$
$$a \leftrightarrow d$$
$$\gamma^{ad}\Gamma^a_{dc}\gamma_{ba}\dot\phi^b\dot\phi^c$$
$$b \leftrightarrow d$$
$$\gamma^{ab}\Gamma^a_{bc}\gamma_{da}\dot\phi^d\dot\phi^c = \Gamma^a_{bc}\dot\phi^b\dot\phi^c$$

Thus the equation of motion is found,

$$\ddot\phi^a + \Gamma^a_{bc}\dot\phi^b\dot\phi^c + 3H\dot\phi^a + V^a = 0$$
$$D_t\dot\phi^a + 3H\dot\phi^a + V^a = 0$$

The riemann curvature is

$$R^\alpha_{\beta\gamma\delta} = \partial_\beta\Gamma^\alpha_{\gamma\delta} - \partial_\gamma\Gamma^\alpha_{\beta\delta} + \Gamma^\sigma_{\gamma\delta}\Gamma^\alpha_{\beta\sigma} - \Gamma^\sigma_{\beta\delta}\Gamma^\alpha_{\gamma\sigma}$$

So the Ricci curvature tensor is

$$R_{\beta\delta} = R^\alpha_{\beta\alpha\delta} = \partial_\beta\Gamma^\alpha_{\alpha\delta} - \partial_\alpha\Gamma^\alpha_{\beta\delta} + \Gamma^\sigma_{\alpha\delta}\Gamma^\alpha_{\beta\sigma} - \Gamma^\sigma_{\beta\delta}\Gamma^\alpha_{\alpha\sigma}$$

And the Ricci scalar curvature is

$$R = R^\beta_\beta = g^{\beta\delta}R_{\beta\delta} = g^{\beta\delta}\partial_\beta\Gamma^\alpha_{\alpha\delta} - g^{\beta\delta}\partial_\alpha\Gamma^\alpha_{\beta\delta} + g^{\beta\delta}\Gamma^\sigma_{\alpha\delta}\Gamma^\alpha_{\beta\sigma} - g^{\beta\delta}\Gamma^\sigma_{\beta\delta}\Gamma^\alpha_{\alpha\sigma}$$

Since the fields only depend on time, we only need to consider the temporal component of the Ricci curvature tensor, so

$$R_{tt} = \partial_t\Gamma^\alpha_{\alpha t} - \partial_\alpha\Gamma^\alpha_{tt} + \Gamma^\sigma_{\alpha t}\Gamma^\alpha_{t\sigma} - \Gamma^\sigma_{tt}\Gamma^\alpha_{\alpha\sigma}$$

From the definition of the christoffel symbols, and noting $g = g(t)$ and g is diagonal, we have $\partial_t g_{aa} = 2g_{ba}\Gamma^b_{ta}$, thus the second term is necessarily 0 since $\partial_t g_{tt} = 0$. The first term is non-zero for $\alpha = 1, 2, 3$, in which it equals

$$\frac{1}{2}\partial_t(1/a^2\partial_t a^2) = \partial_t(\dot a/a) = -3\ddot a/a + 3\dot a^2/a^2$$

The third term is non-zero for $\alpha = \sigma$, In which case we get

$$\frac{3}{4a^4}(\partial_t(a^2))^2 = -3\dot a^2/a^2$$

The last term is vanishes since the temporal component of the metric is constant. Thus we find

$$R_{tt} = 3\frac{\ddot a}{a}$$

Now, if we consider spatial components of the Ricci curvature, we find that the first term now vanishes, the second term is equal to $-\ddot a/a - \dot a^2/a^2$, the thrid term is 0, and the last term is $-\dot a^2/a^2$, so the spacial components are

$$R_{xx} = -\ddot a a - 2\dot a^2$$

Thus the Ricci curvature scalar is

$$6\ddot a/a + 6\dot a^2/a^2$$

Hence the einstein equation gives

$$-3H^2 =$$

I need to write out the indices better if I want to get the signs correct.

$$\partial_a g_{bc} + \partial_b g_{ca} - \partial_c g_{ab} = 2g_{dc}\Gamma^d_{ab}$$

Considering the way this acts on the coordinate vector field ∂_α,

$$R\partial_\alpha = g^{\beta\delta}\partial_\beta\Gamma^\alpha_{\alpha\delta}\partial_\alpha - g^{\beta\delta}\partial_\alpha\Gamma^\alpha_{\beta\delta}\partial_\alpha + g^{\beta\delta}\Gamma^\sigma_{\alpha\delta}\Gamma^\alpha_{\beta\sigma}\partial_\alpha - g^{\beta\delta}\Gamma^\sigma_{\beta\delta}\Gamma^\alpha_{\alpha\sigma}\partial_\alpha$$
$$R\partial_\alpha = g^{\beta\delta}\partial_\beta\nabla_\alpha\partial_\delta - g^{\beta\delta}\partial_\alpha\nabla_\beta\partial_\delta + g^{\beta\delta}\Gamma^\sigma_{\alpha\delta}\nabla_\beta\partial_\sigma - g^{\beta\delta}\Gamma^\sigma_{\beta\delta}\nabla_\alpha\partial_\sigma$$