(DOD) 11.4.3. FORECASTING

Here, we want to predict the uncertainty of cosmological parameters from future experiment. As an example, consider a CMB experiment. Start with the following:

- A set of C_ℓ's describing "true" universe
- δC_ℓ from a given future experiment
- A set of parameters $\{\lambda_\alpha\}$ for which we would like to forecast.

The observed C_ℓ's; $C_\ell^{obs.}$ will be close to the true ones. Indeed,

$$\chi^2 \equiv \sum_\ell \frac{[\, C_\ell(\{\lambda_\alpha\}) - C_\ell^{obs} \,]^2}{(\delta C_\ell)^2}$$

should have its minimum at $\lambda_\alpha = \bar{\lambda}_\alpha$. Of course, we don't know $\bar{\lambda}_\alpha$, but we can quantify how fast χ^2 changes as we move away from it, i.e. error (depending on pivot λ^0). So we assume that to good approximation,

$$\text{prob}(\vec{\lambda}) = e^{-\chi^2/2}$$

and that expanding the χ^2 in terms of λ terminates to good approximation at 2nd order:

$$\chi^2(\vec{\lambda}) = \chi^2(\bar{\lambda}) + \text{linear} + \frac{1}{2}\frac{\partial^2\chi^2}{\partial\lambda_i\partial\lambda_j}(\lambda_i-\bar{\lambda}_i)(\lambda_j-\bar{\lambda}_j)$$

And so

$$-\left\langle L^{(2)}_{ij} \right\rangle = +\left\langle \frac{1}{2}\frac{\partial^2 \chi^2}{\partial\lambda_i \partial\lambda_j} \right\rangle = \frac{1}{2}\frac{\partial^2 \chi^2}{\partial\lambda_i \partial\lambda_j} ,$$

i.e. the gaussian posterior (and likelihood, because (uniform prior) enables us to immediate relate the Fisher matrix to the covariance: they are inverses of each other in the gaussian case?

Hence

$$F_{ij} = \frac{1}{2}\frac{\partial^2 \chi^2}{\partial\lambda_i \partial\lambda_j} = \sum_\ell \frac{\partial}{\partial\lambda_j}\left[\frac{\partial C_\ell}{\partial\lambda_i}\{C_\ell(\{\lambda_i\}) - C_\ell^{obs}\} \right]\frac{1}{\delta C_\ell^2}$$

$$= \sum_\ell \frac{1}{(\delta C_\ell)^2}\left[\frac{\partial C_\ell}{\partial\lambda_i}\frac{\partial C_\ell}{\partial\lambda_j} + \frac{\partial^2 C_\ell}{\partial\lambda_i \partial\lambda_j}\{C_\ell - C_\ell^{obs}\} \right]$$

The second term is usually neglected, because sometimes $\{C_\ell - C_\ell^{obs}\}$ is positive, sometimes negative. And in then $\sum_\ell$, these contributions wash out.

So,

$$F_{ij} = \sum_\ell \frac{1}{(\delta C_\ell)^2}\frac{\partial C_\ell}{\partial\lambda_i}\frac{\partial C_\ell}{\partial\lambda_j}$$

⇒ if you know the specifications (i.e. w^{-1}, f_{sky})

⇒ δC_ℓ

in addition, get C_ℓ derivative of C_ℓ around assumed "true" values.

Marginalization is simple. For instance as Doodson says, if you initially have 5 parameters, compute F^{-1}. Then consider only the 2x2 submatrix (e.g.) for the two (e.g.) parameters, you are interested in. These 2x2 matrices define the error ellipses.

Markov Chain Monte Carlo

We pick most of the material for MCMC methods from the book "Monte Carlo Methods in statistical physics" from M.E.J Newman and G.T. Barkema. I will keep on labeling the sections by an indication of the source (Newman) and the section number of the book.

(Newman) 2. Principles of equilibrium thermal Monte Carlo simulation

(Newman) 2.1. The estimator

The usual goal is to calculate the expectation value of some observable quantity Q. Ideally, we average over all possible states μ of the system which we take to be in thermal equilibrium. So

$$\langle Q \rangle = \frac{\sum_\mu Q_\mu e^{-\beta E_\mu}}{\sum_\mu e^{-\beta E_\mu}}$$

Unfortunately, it will be impossible for any realistic system to sum over all states in practice. Just think of the 2^n states of a spin lattice! The trick of monte carlo techniques is to choose a subset from a probability distribution P_μ such that the subset is much smaller and hence tractable.

Suppose we chose M such states $\{\mu_1 \dots \mu_M\}$.
Our best estimate is then

$$Q_M = \frac{\sum_i^M Q_{\mu_i} p_{\mu_i}^{-1} e^{-\beta E_{\mu_i}}}{\sum_i^M p_{\mu_i}^{-1} e^{-\beta E_{\mu_i}}}$$

For $M \to \infty$, we clearly get $\lim_{M\to\infty} Q_M = \langle Q \rangle$.
How should we choose p_μ? Suppose we chose $p_\mu = 1$ for all states. Then

$$Q_M = \frac{\sum_i^M Q_{\mu_i} e^{-\beta E_{\mu_i}}}{\sum_i^M e^{-\beta E_{\mu_i}}}$$

Unfortunately, we would spend a lot of time computing states with very low $e^{-\beta E_{\mu_i}}$ while surely missing the few very probable states.
If we knew a better p_μ, we could sample exactly the important states. This is the idea of MCMC. Picking the important states is called "importance sampling".

(Newman) 2.2. Importance sampling

A common choice for picking important states is to choose

$$p_\mu = Z^{-1} e^{-\beta E_\mu}$$

then

$$Q_M = \frac{\sum_i^M p_{\mu_i}^{-1} Q_{\mu_i} e^{-\beta E_{\mu_i}}}{\sum_i^M p_{\mu_i}^{-1} e^{-\beta E_{\mu_i}}}$$

$$= \frac{\sum_i^M Z^M e^{+\beta E_{\mu_i}} e^{-\beta E_{\mu_i}} Q_{\mu_i}}{\sum_i^M Z e^{\beta E_{\mu_i}} e^{-\beta E_{\mu_i}}} = \frac{\sum_i^M Z Q_{\mu_i}}{\sum_i^M Z}$$

$$= \frac{1}{M} \sum_{i}^{M} Q_{\mu_i}$$

Obviously, this choice of P_μ is much better, because the sample weighs each state equally.
But how do we pick M states $\{\mu_1 .. \mu_M\}$ such that $P_\mu = Z^{-1} e^{-\beta E_\mu}$? The standard solution is to use a "Markov process".

(Neuau) 2.2.1. Markov processes

We cannot pick state at random and accept or reject according to $e^{-\beta E}$, because that would cost as much as performing the sum over all states.
Instead, we use a Markov process.
A Markov process ~~has~~ generates a new state ν given a current state μ at random with probability $P(\mu \to \nu)$. It must satisfy:

1. $P(\mu \to \nu)$ should not vary over time
2. They should only depend on μ and ν and not on any former state.
3. $\sum_\nu P(\mu \to \nu) = 1$

Note that $P(\mu \to \mu)$ need not be 0!
Using Markov processes repeatedly generates a Markov chain.

As we want to reach the equilibrium distribution, we place two more constraints on our Markov process: "ergodicity" and "detailed balance".

(Newman) 2.2.2. Ergodicity

Ergodicity means that any state of the system should be reachable from any other state if we run long enough. This is necessary, because any state appears with probability $e^{-\beta E_\mu}$ in the true system. If we start our chain from a point that never reaches μ, we clearly don't sample to the full system.

Note that this does not mean that $P(\mu \to \nu)$ cannot be 0 for states. But ν must be reached by some intermediate steps.

(Newman) 2.2.3 Detailed balance

Detailed balance ensures that we generate the Boltzmann distribution in equilibrium and not any other distribution. In equilibrium, the rate of transitions into and out of some state must be equal:

$$\sum_\nu P_\mu \, \Pi(\mu \to \nu) = \sum_\nu P_\nu \, \Pi(\nu \to \mu)$$

(μ: out of; ν: into)

using $\sum_\nu \Pi(\mu \to \nu) = 1$, we get

$$P_\mu = \sum_\nu P_\nu \, \Pi(\nu \to \mu)$$

The transition probabilities $P(\mu \to \nu)$ can be thought of forming a "Markov matrix" or "stochastic matrix" for the process. Let us denote the probability that our chain is in state μ at time step t by $w_\mu(t)$ then

$$\vec{w}(t+1) = P\,\vec{w}(t) \quad ; \quad w_\nu(t+1) = \sum_\mu P(\mu \to \nu) w_\mu(t)$$

$$P_{\nu\mu} = P(\mu \to \nu)$$

If we reach equilibrium, then $\vec{w}(\infty)$ as $t \to \infty$ then

$$\vec{w}(\infty) = P\,\vec{w}(\infty)$$

However, it is also possible that we reach a dynamical equilibrium in which $\vec{w}$ rotates around a number of different values. This is called a "limit cycle":

$$\vec{w}(\infty) = P^n \vec{w}(\infty)$$

We can eliminate limit cycles by detailed "balance":

$$P_\mu P(\mu \to \nu) = P_\nu P(\nu \to \mu)$$

which if summed over ν is our original requirement. The l.h.s. of this equation is the overall rate of transition from μ to ν. The r.h.s is the reverse process. Hence detailed balance says that on average, the system goes as frequently from μ to ν as it goes from ν to μ.

In a limit cycle, the probability for of occupation for some of the states changes in cyclic fashion. So detailed balance must be violated for some states. Detailed balance forbids limit cycles.

As $t \to \infty$, $\vec{w}(t)$ tends exponentially towards the eigenvector of $\underline{P}$ with largest eigenvalue. From $\vec{w}(\infty) = \underline{P}\,\vec{w}(\infty)$, we see that 1 is the largest eigenvalue. Looking back at

$$** \quad P_\mu = \sum_\nu P_\nu P(\nu \to \mu) \; ; \qquad \vec{p} = \underline{P}\,\vec{p}$$

we see that if $\vec{p} = \underline{P}\,\vec{p}$ holds for our process, then $\vec{p}$ which holds the p_μ's is precisely the correctly normalized eigenvector to eigenvalue **1**. Hence $\vec{w}(t) \to \vec{p}$ for $t \to \infty$, because in equilibrium $\vec{w}(\infty) = \underline{P}\,\vec{w}(\infty)$.

We have seen that we can arrange for the probability distribution of our chain to tend to <u>any</u> distribution p_μ we like by specifying $\underline{P}$

$$p_\mu P(\mu \to \nu) = p_\nu P(\nu \to \mu)$$

in a suitable fashion. For a system in thermal equilibrium, we would like $p_\mu = e^{-\beta E_\mu}$ hence:

$$\frac{P(\mu \to \nu)}{P(\nu \to \mu)} = \frac{p_\nu}{p_\mu} = e^{-\beta(E_\nu - E_\mu)}$$

2.3. Acceptance Ratios

Our requirement of $\sum_\nu P(\mu\to\nu)=1$ gives us some flexibility of adjusting $P(\mu\to\mu)$, the probability of staying at μ. So the trick is to split P in two parts: selection and acceptance:

$$P(\mu\to\nu) = g(\mu\to\nu)\, A(\mu\to\nu)$$

g: selection probability; A: acceptance ratio

A says that we should accept a new state with probability $A\in[0\ldots1]$. Choosing $A=0$ means never move, which is not clever of course ☺. The freedom to choose A gives us freedom to choose g, because

$$\frac{P(\mu\to\nu)}{P(\nu\to\mu)} = \frac{P_\nu}{P_\mu} = e^{-\beta(E_\nu - E_\mu)}$$

only fixes the ratio

$$\frac{P(\mu\to\nu)}{P(\nu\to\mu)} = \frac{g(\mu\to\nu)\,A(\mu\to\nu)}{g(\nu\to\mu)\,A(\nu\to\mu)}$$

So our strategy is to choose a selection function g and then adjust A to satisfy the above equation. Obviously, A should be as large as possible to sample as many different states as possible.