PARTICLES AND COSMOLOGY

17th Baksan School
on Astroparticle Physics

Modern Statistical Methods and Tools

Lecture 3

Grigory I. Rubtsov
Institute for Nuclear Research of the
Russian Academy of Sciences

Terskol, Kabardino-Balkarian Republic, Russia
April 4-11, 2025

Bayesian approach summary

$$P(M|obs) \sim P(obs|M)\,P(M)$$

- P(obs|M) – likelihood, P(M) – prior, P(M|obs) – posterior probability
- Calculate marginal distributions:

$$p(m_l) = \frac{\iint\limits_{m_1..m_K \setminus m_l} P(M|obs)\,dM}{\iint\limits_{m_1..m_K} P(M|obs)\,dM} \qquad p(m_l, m_q) = \frac{\iint\limits_{m_1..m_K \setminus m_l m_q} P(M|obs)\,dM}{\iint\limits_{m_1..m_K} P(M|obs)\,dM}$$

Bayesian approach summary

$$P(M|obs) \sim P(obs|M)\,P(M)$$

- P(obs|M) – likelihood, P(M) – prior, P(M|obs) – posterior probability
- Calculate marginal distributions:

$$p(m_l) = \frac{\iint\limits_{m_1..m_K \setminus m_l} P(M|obs)\,dM}{\iint\limits_{m_1..m_K} P(M|obs)\,dM} \qquad p(m_l, m_q) = \frac{\iint\limits_{m_1..m_K \setminus m_l m_q} P(M|obs)\,dM}{\iint\limits_{m_1..m_K} P(M|obs)\,dM}$$

- How to calculate marginal distribution if M has many dimensions?
  - E.g. Fermi 4FGL catalog has 7195 sources, full sky model has more than 10k parameters

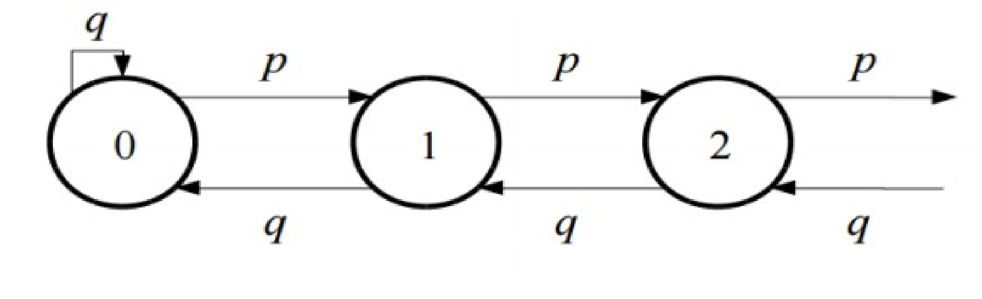
Markov chains

- Stochastic process – sequence of random variables
- Markov chain – special type of stochastic process
  - the probability of occurrence of one or another outcome at step n+1 depends only on the state of the system at step n
- $q_1,..,q_{n-1},q_n,..$
  - Probability distribution function $p(q_n)$ is completely determined by q_{n-1}
- Note: the requirement of no memory is often relaxed, but these chains are not Markovian

Andrey Andreyevich Markov, 1856-1922

Markov chains: discrete case with finite number of states

- Provide probability to move from one state to another
- Example: one dimensional random walk (p=q=0.5)



- Probability of each state may be defined as a vector

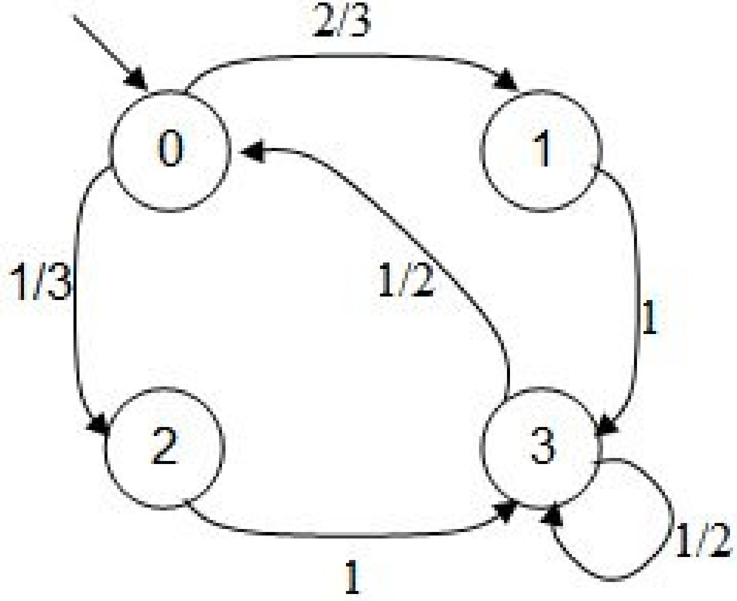
$$x_0 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

- Transition matrix P: P**x** gives the probability of each state at the next step
- P^2**x** – probability after 2 steps
- P^n**x** – aften n steps

$$P = \begin{pmatrix} q & p & 0 & 0 \\ q & 0 & p & 0 \\ 0 & q & 0 & p \\ 0 & 0 & q & p \end{pmatrix}$$

Markov chains: discrete example

State diagram



Matrix

$$\begin{pmatrix} 0 & 2/3 & 1/3 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1/2 & 0 & 0 & 1/2 \end{pmatrix}$$

Probability distribution

Start.	(1,	0,	0,	0)
1.	(0,	2/3,	1/3,	0)
2.	(0,	0,	0,	1)
3.	(1/2,	0,	0,	1/2)
4.	(1/4,	1/3,	1/6,	1/4)

Phil Hooker, the University of Edinburgh UK

Markov chains: ergodicity

- For Markov chains ergodic hypothesis has a form of a theorem
- For any irreducible and aperiodic Markov chain

$$\exists \lim_{n \to \infty} P^n = A$$

- Irreducible: one may reach any state starting from any other

$$P = \begin{pmatrix} 0.3 & 0.7 & 0 & 0 \\ 0.2 & 0.8 & 0 & 0 \\ 0 & 0 & 0.6 & 0.4 \\ 0 & 0 & 0.9 & 0.1 \end{pmatrix}$$

reducible

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

periodic

Markov chains: ergodicity

$$\exists \lim_{n \to \infty} P^n = A$$

$$\boldsymbol{\alpha} = \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix}$$

- All columns of the matrix A are the same (vector **α**)
- For any starting vector x:

$$\exists \lim_{n \to \infty} P^n \boldsymbol{x} = \boldsymbol{\alpha}$$

$$P\boldsymbol{\alpha} = \boldsymbol{\alpha}$$

- **α** – equilibrium distribution of the Markov chain

Markov Chain Monte Carlo

- Metropolis-Hastings algorithm, 1953
- The idea of the method:
  - Construct a Markov chain for which the equilibrium distribution is the posterior probability

$$\rho(M) \sim P(M|obs)^{1/T}$$

- T – temperature, 1 by default

Metropolis-Hastings algorithm

- The transition to the new state is two-stage:
  - random step to another point M → M’
  - accept new state or stay at M
- Transition probability is defined as a product

$$p(M \rightarrow M') = step(M \rightarrow M')\, accept(M \rightarrow M')$$

- step is symmetric

$$step(M \rightarrow M') = step(M' \rightarrow M)$$

$$accept(M \rightarrow M') = \begin{Bmatrix} 1, if\ \rho(M') > \rho(M) \\ \rho(M')/\rho(M), else \end{Bmatrix}$$

Metropolis-Hastings algorithm

- The detailed equilibrium condition

$$\rho_{eq}(M)\,p(M \to M') = \rho_{eq}(M')\,p(M' \to M)$$

$$\rho_{eq}(M)\,step(M \to M')\,accept(M \to M') = \rho_{eq}(M')\,step(M' \to M)\,accept(M' \to M)$$

- Using the equations from the previous slide

$$step(M \to M') = step(M' \to M)$$

$$accept(M \to M') = \begin{cases} 1, if\ \rho(M') > \rho(M) \\ \rho(M')/\rho(M), else \end{cases}$$

- Arrive to:

$$\frac{\rho_{eq}(M)}{\rho_{eq}(M')} = \frac{\rho(M)}{\rho(M')}$$

Example: Markov chains from Planck experiment

0.2000000E+01	0.3903386E+04	0.2216171E-01	0.1173768E+00	0.1040830E+01	0.1574385E+00	0.9742281E+00	0.3212016E+01	0.1061088E+03
0.3922832E+02	0.1188850E+03	0.9237343E+01	0.1926690E+02	0.7390182E+01	0.7961410E+00	0.5182232E+00	0.3118709E+00	0.1000757E+01
0.9952326E+00	0.6375330E+00	0.4118827E+01	0.8931399E-01	0.6976213E+00	0.3023787E+00	0.8757837E+00	0.1634064E+02	0.0000000E+00
0.6808838E+02	0.0000000E+00	0.2482908E+01	0.1401836E+00	0.9544873E-01	0.2477506E+00	0.1812222E+01	0.6450616E-03	0.1380927E+02
0.1090054E+04	0.1452632E+03	0.1040979E+01	0.1059322E+04	0.1480165E+03	0.1396357E+00	0.1612278E+00	0.3334282E+04	0.8249804E+00
0.7211274E-01	0.3104359E-03	0.1382280E+04						
0.1000000E+01	0.3903445E+04	0.2217904E-01	0.1188319E+00	0.1040756E+01	0.1460309E+00	0.9710000E+00	0.3193086E+01	0.1753512E+03
0.5447629E+02	0.1136970E+03	0.6445500E+01	0.2598853E+02	0.8612640E+01	0.8185609E+00	0.5569459E+00	0.4452574E+00	0.1000235E+01
0.9954719E+00	0.4192213E+00	0.1044009E+01	0.5696056E+00	0.6896567E+00	0.3103433E+00	0.8716791E+00	0.1555023E+02	0.0000000E+00
0.6756104E+02	0.0000000E+00	0.2436350E+01	0.1416560E+00	0.9570426E-01	0.2477580E+00	0.1819277E+01	0.6450616E-03	0.1381598E+02
0.1090160E+04	0.1448694E+03	0.1040910E+01	0.1059475E+04	0.1476068E+03	0.1400761E+00	0.1611470E+00	0.3369476E+04	0.8185305E+00
0.7164725E-01	0.3099108E-03	0.1388806E+04						
0.3000000E+01	0.3904980E+04	0.2218395E-01	0.1185423E+00	0.1041055E+01	0.1350443E+00	0.9711457E+00	0.3170773E+01	0.2926837E+03
0.5244221E+02	0.1104685E+03	0.5307461E+01	0.2981380E+02	0.6029230E+01	0.7927495E+00	0.5571131E+00	0.5019521E+00	0.1000000E+01
0.9962095E+00	0.7369567E-01	0.1321914E+01	0.1464696E+01	0.6921521E+00	0.3078479E+00	0.8612775E+00	0.1471100E+02	0.0000000E+00
0.6776611E+02	0.0000000E+00	0.2382589E+01	0.1413713E+00	0.9580185E-01	0.2477601E+00	0.1818658E+01	0.6450616E-03	0.1380543E+02
0.1090129E+04	0.1449410E+03	0.1041203E+01	0.1059475E+04	0.1476777E+03	0.1400052E+00	0.1611942E+00	0.3362671E+04	0.8200120E+00
0.7180792E-01	0.3102621E-03	0.1385930E+04						
0.3000000E+01	0.3901667E+04	0.2201057E-01	0.1207360E+00	0.1041001E+01	0.7594523E-01	0.9629134E+00	0.3058624E+01	0.1121257E+03
0.4801457E+02	0.9877199E+02	0.4680260E+01	0.2947102E+02	0.6287884E+01	0.8137509E+00	0.6816312E+00	0.3961482E+00	0.1000331E+01
0.9953284E+00	0.2752744E+00	0.5962238E+01	-0.1688007E+00	0.6789775E+00	0.3210225E+00	0.8198106E+00	0.9901428E+01	0.0000000E+00
0.6683350E+02	0.0000000E+00	0.2129823E+01	0.1433916E+00	0.9583365E-01	0.2476856E+00	0.1829693E+01	0.6450616E-03	0.1383389E+02
0.1090547E+04	0.1445049E+03	0.1041159E+01	0.1059246E+04	0.1472881E+03	0.1402746E+00	0.1613635E+00	0.3410962E+04	0.8108123E+00

Planck Legacy Archive, COM_CosmoParams_base_planck_lowl_R1.10.tar.gz

Note: first column – weight, how long the chain stayed in the state

MCMC: marginal distributions

- Finally, there is a chain $\{m_1^a, .. m_k^a\}$, k – number of parameter, a – number of step
- We need 2D marginal distribution for m_l, m_q
- Q: How to get them?

MCMC: marginal distributions

- Finally, there is a chain $\{m_1^a,..m_k^a\}$, k – number of parameter, a – number of step
- We need 2D marginal distribution for m_l,m_q
- Q: How to get them?
- A: no need to integrate, just take these parameters from the chain and ignore others $\{m_l^a m_q^a\}$
- Calculate the distribution of $\{m_l^a m_q^a\}$

Demonstration by Chi Feng

https://chi-feng.github.io/mcmc-demo/app.html?algorithm=RandomWalkMH&target=banana

MCMC: final notes

- Multiple chains are generated simultaneously
- When the chain is started, its states may be far from equilibrium
  - remove first “warm up” part of the chain
- Several methods exist to test the chain for convergence. One of them is the Gelman-Rubin statistic

(Statistical Science, 7, 4 (1992) 457)

MCMC: final notes

- If the subspace of good models is not simply connected, the chain will not move from one to another part
  - Train several networks with different temperatures in parallel and allow switching states between them

    d'Avigneau et al., Anytime parallel tempering, Statistics and Computing (2021) 31:74
- For some models it's easy to make step along some of the parameters and hard along others. E.g. for CMB is easy to change overall normalization.
  - Slow/fast optimization: make one slow step and then multiple fast steps
- Packages are available for MCMC: PyMC, Stan (PyStan)

Takeout from lecture 3

- Markov chain is the stochastic process without memory
- Ergodic assertion may be proved for Markov chains
- Metropolis-Hastings algorithm: construct Markov chain with the equilibrium distribution proportional to posterior probability
- There are convergence criteria for the Markov chain (Gelman-Rubin statistic is one of them)
- Given the Markov chain, marginalized distributions may be obtained without integration (by ignoring unnecessary variables)

Task for self-check

- Use the list of Fermi LAT photons and exposure from data directory at Yandex disk.
- Construct a model of gamma-ray radiation with two sources:
  - Isotropic flux
  - Constant flux in a circle with a radius of 1^{o} around Crab
- Calculate likelihood and posterior probability distribution
- Implement MCMC and plot posterior distributions of the parameters

Task for self-check

- Make the model more complicated:
  - Include the flux near Galactic plane
  - In addition to Crab, add another source with both the flux and the coordinates as free parameters
- Implement MCMC and plot posterior marginal distributions of the parameters
- (*) test the chain for convergence
- Add 2 free sources to the model, add 100 free sources

Hands-on session

- Download the code
- https://disk.yandex.ru/d/bPrpOq2Z-oJlOw
- Run jupyter notebook
- Go through exercises in the notebook


Thank you!

Backup slides