Astrostatistics: 04 Mar 2020

https://github.com/CambridgeAstroStat/PartIII-Astrostatistics-2020

- Today: finish MCMC / start Gaussian processes
  - sketch of MCMC theory: more details: see, e.g. Robert & Casella, "Monte Carlo Statistical Methods"
  - Mixed samplers / parameter blocking
  - Gaussian processes in astrophysics
- Lecture MCMC code examples online: lecture_codes/
- Example Class 3: Moved to:
  - Friday, 13 Mar, 12pm, MR5 (usual lecture time)

- Finish sketch of MCMC Theory
- Some remarks on mixed samplers and blocking

MCMC in Practice

1. Find the mode(s), using optimisation.

2. Begin multiple (4-8, parallel) chains at starting positions dispersed around the mode(s).

3. Scale Metropolis proposals to tune 25-50% acceptance rate (depending on dimensionality of jump)

4. Use proposal covariance matrix that reflects the shape of the posterior

5. After run, look at chains to check for obvious problems

6. Compute Gelman-Rubin ratio comparing within-chain-variance to between-chain variance to check that chains are well-mixed (should be very close to 1), and assess burn-in

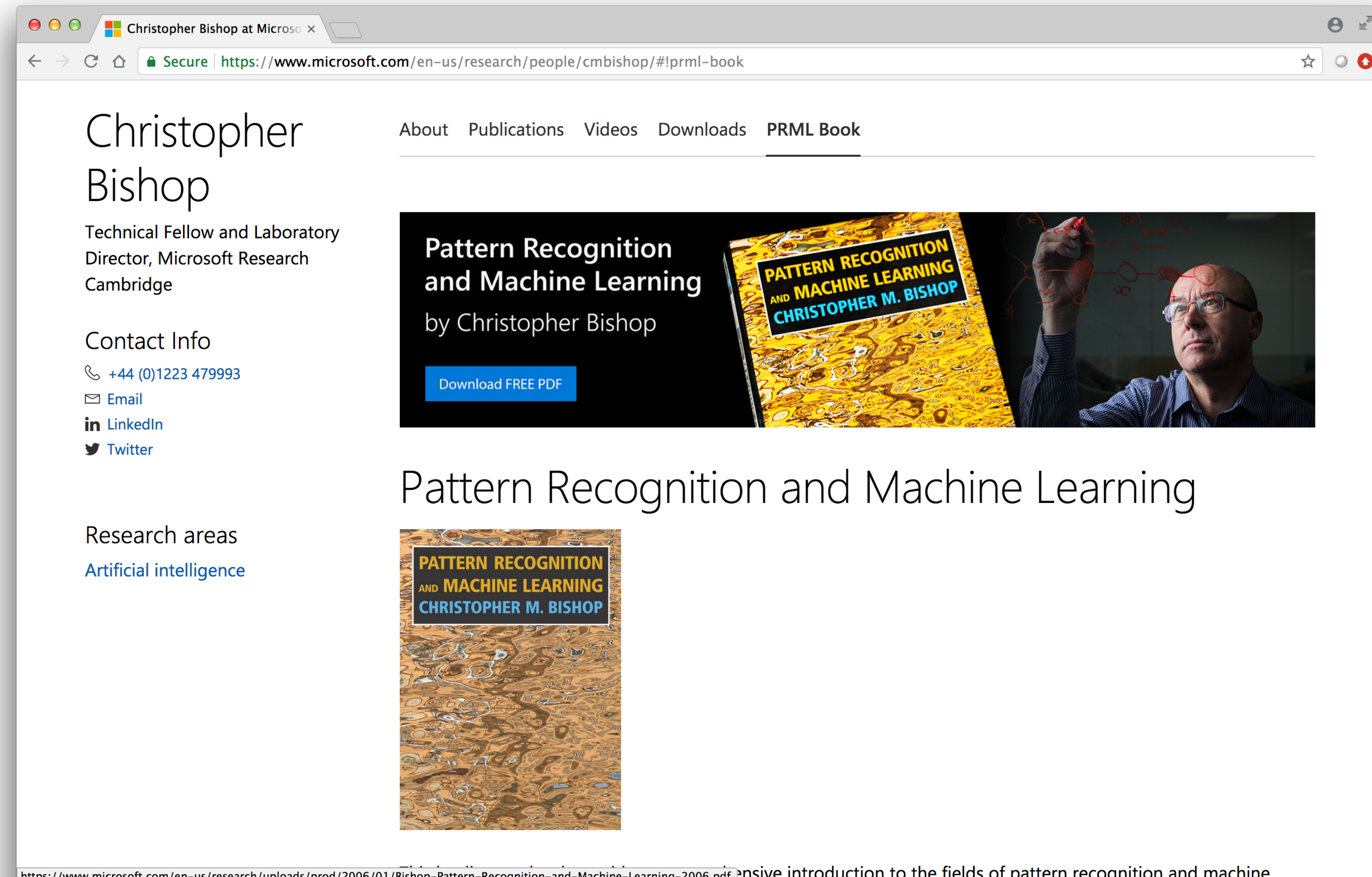
7. Compute autocorrelation timescale and effective sample size to make sure you have enough *independent* samples for inference.

8. If all checks out, remove burn-in, thin, and combine chains to compute posterior inferences

Human Learning of Gaussian Processes

- Classic Text: Rasmussen & Williams (2006)
  - “Gaussian Processes for Machine Learning”,Ch 1-2,4-5
  - Free Online: http://www.gaussianprocess.org/gpml/
- Ivezic, Sec 8.10 GP Regression, (Ch 8 is Regression)
- Bishop: Pattern Recognition & Machine Learning, Ch 6
  - Also free online: https://www.microsoft.com/en-us/research/people/cmbishop/#!prml-book
- Gelman, Bayesian Data Analysis 3rd Ed., Chapter 21
- “Practical Introduction to GPs for Astronomy” - D. Foreman-Mackey
  - http://hea-www.harvard.edu/AstroStat/aas231_2018/DForeman-Mackey_20180110_aas231.pdf

Human Learning of Gaussian Processes



Review: Properties of Multivariate Gaussians

Full probability density:

$\boldsymbol{\Sigma}$ is positive definite

$$N(\boldsymbol{f}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \equiv [\det(2\pi\boldsymbol{\Sigma})]^{-1/2} \exp[-\tfrac{1}{2}\,(\boldsymbol{f}-\boldsymbol{\mu})^T\,\boldsymbol{\Sigma}^{-1}\,(\boldsymbol{f}-\boldsymbol{\mu})]$$

Joint distribution of components:

$$\boldsymbol{f} = \begin{pmatrix} \boldsymbol{U} \\ \boldsymbol{V} \end{pmatrix} \sim N\left(\begin{bmatrix} \boldsymbol{U}_0 \\ \boldsymbol{V}_0 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_U & \boldsymbol{\Sigma}_{UV} \\ \boldsymbol{\Sigma}_{VU} & \boldsymbol{\Sigma}_V \end{bmatrix}\right)$$

If you observe/know/condition on V:

Conditional dist'n: $\boldsymbol{U}|\,\boldsymbol{V} \sim N\,(\mathbb{E}[\boldsymbol{U}|\boldsymbol{V}],\, \mathrm{Var}[\boldsymbol{U}|\boldsymbol{V}])$

Conditional Mean: $\mathbb{E}[\boldsymbol{U}|\boldsymbol{V}] = \boldsymbol{U}_0 + \boldsymbol{\Sigma}_{UV}\,\boldsymbol{\Sigma}_V^{-1}(\boldsymbol{V}-\boldsymbol{V}_0)$

Conditional Variance: $\mathrm{Var}[\boldsymbol{U}|\boldsymbol{V}] = \boldsymbol{\Sigma}_U - \boldsymbol{\Sigma}_{UV}\,\boldsymbol{\Sigma}_V^{-1}\,\boldsymbol{\Sigma}_{VU}$

If V = observed data, U = unobserved parameters, then P(**U** | **V**) is a posterior pdf!

What is a Gaussian Process?

- A GP is a collection of random variables $\{f_t\}$, (typically with some ordering in time, space or wavelength), such that any finite subset of r.v.s have a jointly multivariate Gaussian distribution.

- Any vector $\mathbf{f} = \{f_t : t = 1 \ldots N\}$ of a finite subset is multivariate Gaussian, therefore it is completely described by a mean $\mathbf{E}[\mathbf{f}]$ and covariance matrix $\mathbf{Var}[\mathbf{f}] = \mathbf{Cov}[\mathbf{f}, \mathbf{f}^T]$.

- Elements of the **covariance matrix** are determined by a function of the coordinates, e.g. $\mathrm{Cov}[f_t, f_{t'}] = k(t, t')$, called the *covariance function* or *kernel*

- A Gaussian Process with mean function m(t) is denoted:

$$f(t) \sim \mathcal{GP}(m(t), k(t, t'))$$

What are GPs used for in astrophysics?

- Some physical models are actually very nearly Gaussian Processes (e.g. Cosmic Microwave Background is a GP on the sphere) or approximately so (damped random walks for quasar light curves)

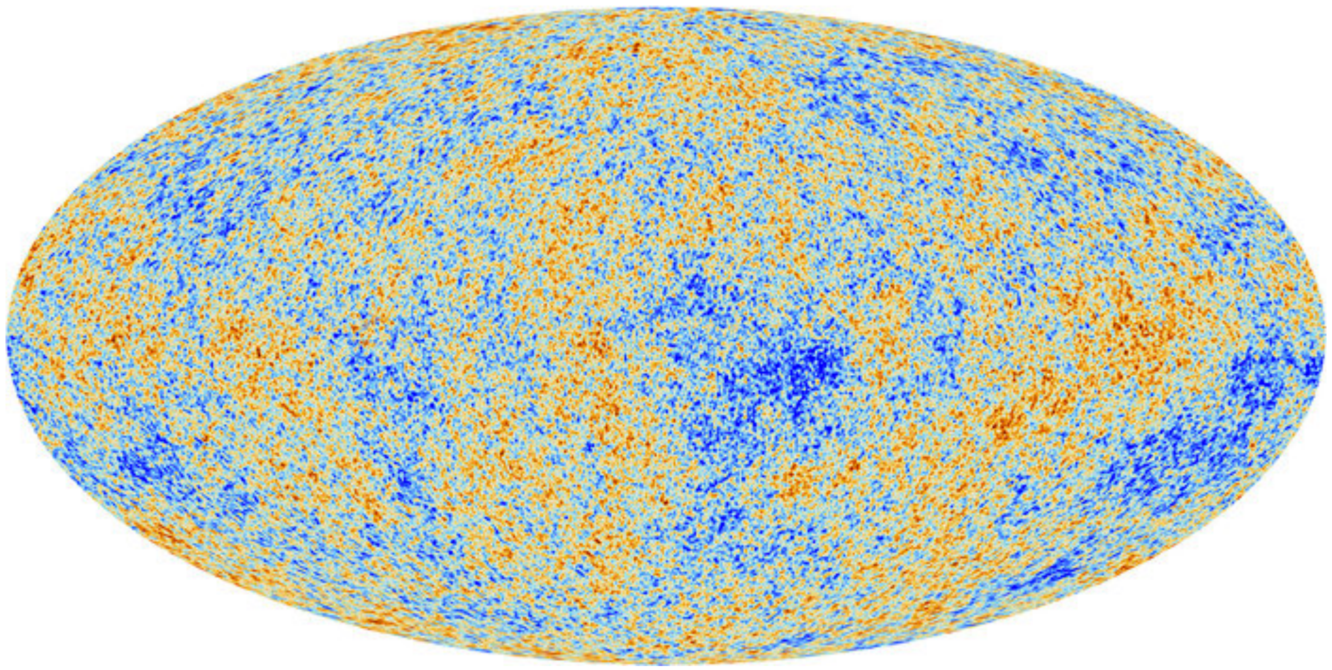
- “Nonparametric” models: flexible functions to use when an accurate astrophysical function is not available or is imperfect

- Nonparametric ~ number of parameters grows with the dataset

- Interpolation/Emulation: To generate a smooth curve going through some observation or simulation points

- Correlated noise/error model: When you marginalise out the latent error function, you are effectively accounting for correlated noise over time/space/wavelength.
  e.g. *D. Jones et al.* “Improving Exoplanet Detection Power: Multivariate Gaussian Process Models for Stellar Activity.” arXiv:1711.01318

Example: A Gaussian Process for the Spatial Variations of Intensity/Temperature



Cosmic Microwave Background (Planck)
~ Gaussian Random Field
(mean = 2.7 K,
std dev ~ 10^{-5})

Power Spectrum
(~Fourier Transform of Covariance Function)