Part III Astrostatistics: Example Sheet 3
Example Class: Friday, 13 Mar 2020, 12:00pm, MR5

1 Doubly Lensed Quasar Time Delay Estimation

Quasar light curves (brightness time series) $f(t)$ (in magnitudes) are often modelled using the Orstein-Uhlenbeck (O-U) process, which is described mathematically by a stochastic differential equation of the form:

$$df(t) = \tau^{-1}[c - f(t)]\, dt + \sigma dW_t \tag{1}$$

where the second term is a Brownian motion (continuous-time limit of a random walk), with the variability scaled by σ, and the first term is a drag that tends to return the brightness back to the mean level c. The O-U process is a Gaussian process

$$f(t) \sim \mathcal{GP}(m(t), k(t, t')\,) \tag{2}$$

with mean function $m(t) = c$ and covariance function or kernel:

$$\text{Cov}[f(t), f(t')] = k(t, t') = A^2 \exp(-|t - t'|/\tau) \tag{3}$$

with characteristic amplitude $A^2 = \tau\sigma^2/2$. The characteristic timescale for the quasar brightness to revert to the mean c is τ. Hence, astronomers often called this a "damped random walk". In a doubly-lensed quasar system, two images of the same quasar are observed. However, their brightness time series will have a time delay and magnification relative to each other due to the gravitational lensing effects. Find in the accompanying dataset, measurements of the brightness time series of two images of a lensed quasar, $y_1(t)$ and $y_2(t)$. Assume the measurement errors are Gaussian with the given standard deviations. Where possible, write down and derive the relevant equations before you implement them in code.

1. Plot the data. For each image (y_1 or y_2) time series separately, fit an O-U process by optimising the marginal likelihood to estimate c, A, and τ for each image. Are these estimates consistent between the two time series? Estimate the overall relative magnification factor (difference in magnitudes) between the two images. (The relative multiplicative magnification μ due to the gravitational lens is related to the magnitude shift by $\Delta m = -2.5 \log_{10} \mu$).

2. Fixing the values of c, A, and τ you found for each image separately, overplot random light curves drawn from the GP prior on each separate time series dataset. Use the Gaussian Process machinery to estimate the underlying light curve of each image separately. Plot the expectation and standard deviation of the posterior prediction as a function of time.

3. Derive a likelihood function for the two time series considered jointly, as two copies of the same realisation of the GP but shifted in time by the time delay Δt, and the magnification factor Δm (both relative to y_1), and measured with noise at the observed times. Thus $y_1(t)$ is a noisy measurement of $f(t)$ and $y_2(t)$ is a noisy measurement of $f(t - \Delta t) + \Delta m$. Using suitable non-informative priors, write down a posterior density $P(\Delta t, \Delta m, c, A, \tau \,|\, \boldsymbol{y}_1, \boldsymbol{y}_2)$.

4. Estimate Δt and Δm. Beware that the log likelihood is highly multimodal, so it is important to find the major mode. You may fix the O-U process parameters to reasonable values found previously, or estimate them jointly with $(\Delta t, \Delta m)$.

5. Overplot the two time series datasets, with y_2 shifted to the y_1 frame by subtracting the estimated Δt and Δm. Now using the O-U parameters you found, plot the posterior estimate of the underlying light curve using the two combined datasets.

Numerical Clue: A proper covariance matrix admits a Cholesky decomposition: $\boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{L}^T$, where $\boldsymbol{L}$ is the lower triangular Cholesky factor. The log of the determinant $|\boldsymbol{\Sigma}| = \det \boldsymbol{\Sigma}$ can stably be computed from Equation A.18 of Rasmussen & Williams. If you have computed the Cholesky factor $\boldsymbol{L}$, solutions $\boldsymbol{x}$ to linear equations of the form $\boldsymbol{\Sigma x} = \boldsymbol{b}$, i.e. $\boldsymbol{x} = \boldsymbol{\Sigma}^{-1}\boldsymbol{b}$, can be stably computed using forward/backward substitution, rather than by directly inverting $\boldsymbol{\Sigma}$, as described in Rasmussen & Williams, §A.4.

2 Supernova Cosmology

Suppose Type Ia supernovae (SN) are standard candles: the true absolute magnitude M_s (proportional to the logarithm of the luminosity) of each individual supernova s is an independent draw from a narrow Gaussian population distribution

$$M_s \sim N(M_0, \sigma^2_{\text{int}}) \tag{4}$$

with unknown mean M_0 and unknown intrinsic "dispersion" or variance σ^2_{int}. The dimming effect of distance relates the true absolute magnitude M_s to the true apparent magnitude m_s for each SN s:

$$m_s = M_s + \mu(z_s; H_0, w, \Omega_M) \tag{5}$$

where the true distance modulus at the observed redshift z_s is

$$\mu(z_s; H_0, w, \Omega_M) = 25 + 5\log_{10}\left[\frac{c}{H_0}\tilde{d}(z_s; w, \Omega_M) \ \text{Mpc}^{-1}\right] \tag{6}$$

where Mpc is a mega-parsec (a unit of distance), c is the speed of light, H_0 is the Hubble constant, and (w, Ω_M) are other cosmological parameters, and, in a flat Universe,

$$\tilde{d}(z; w, \Omega_M) = (1+z)\int_0^z \frac{dz'}{\sqrt{\Omega_M(1+z')^3 + (1-\Omega_M)(1+z')^{3(1+w)}}} \tag{7}$$

is a dimensionless deterministic function. Assume we observed the apparent magnitude (data) m_s without measurement error. The redshift z_s for each SN s is known perfectly. In the provided table, find the data $\mathcal{D} = \{m_s, z_s\}$ for independent measurements of N supernovae.

1. Derive likelihood function for the sample of N supernovae:

$$L(M_0, \sigma^2_{\text{int}}, H_0, w, \Omega_M) = P(\{m_s\}|\,\{z_s\}, M_0, \sigma^2_{\text{int}}, H_0, w, \Omega_M) \tag{8}$$

Rewrite this in terms of $\mathcal{M} = M_0 - 5\log h$, where $h = H_0/100$ km s^{-1} Mpc^{-1}.

2. Assume flat improper priors for $w, \mathcal{M} \sim U(-\infty, \infty)$, and flat positive improper priors

$$P(X) \propto \begin{cases} 1, & X \geq 0 \\ 0, & X < 0 \end{cases} \tag{9}$$

for Ω_M and σ^2_{int}. Write down the unnormalised joint posterior $P(\mathcal{M}, \sigma^2_{\text{int}}, w, \Omega_M | \mathcal{D})$. Derive the conditionals $P(\mathcal{M} | \sigma^2_{\text{int}}, w, \Omega_M; \mathcal{D})$ and $P(\sigma^2_{\text{int}} | \mathcal{M}, w, \Omega_M; \mathcal{D})$. Use these conditionals to construct an MCMC algorithm to sample $P(\mathcal{M}, \sigma^2_{\text{int}}, w, \Omega_M | \mathcal{D})$ over the 4 parameters.

3. Run an optimiser to find the maximum a posteriori (MAP) estimate. Implement your MCMC algorithm to analyse the data. (Example code for $\tilde{d}(z; w, \Omega_M)$ is provided: "lumdist" computes $\tilde{d}$ for a single redshift z. "lumdist_vec" computes the $\tilde{d}$'s for a sorted vector of redshifts in an efficient manner. Set $\Omega_L = 1 - \Omega_M$ in a flat Universe). Assess your chains. If appropriate, use your chains to compute the marginal posterior distribution of w and Ω_M, and plot their joint posterior.

4. Estimate the posterior covariance matrix of $(\mathcal{M}, \sigma^2_{\text{int}}, w, \Omega_M)$ using your chains. Use a scaled version of this as the proposal covariance matrix in a 4D Metropolis algorithm. Implement this algorithm. Compare the performance of your two algorithms.