Part III Astrostatistics: Example Sheet 3
Example Class: Friday, 15 Mar 2018, 1:00pm, MR12 TBC

1 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

$$M_s \sim N(M_0, \sigma^2_{\text{int}}) \tag{1}$$

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:

$$m_s = M_s + \mu(z_s; H_0, w, \Omega_M) \tag{2}$$

where the true distance modulus at the observed redshift z_s is

$$\mu(z_s; H_0, w, \Omega_M) = 25 + 5 \log_{10} \left[\frac{c}{H_0} \tilde{d}(z_s; w, \Omega_M) \ \text{Mpc}^{-1} \right] \tag{3}$$

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,

$$\tilde{d}(z; w, \Omega_M) = (1+z) \int_0^z \frac{dz'}{\sqrt{\Omega_M (1+z')^3 + (1-\Omega_M)(1+z')^{3(1+w)}}} \tag{4}$$

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:

$$L(M_0, \sigma^2_{\text{int}}, H_0, w, \Omega_M) = P(\{m_s\} | \{z_s\}, M_0, \sigma^2_{\text{int}}, H_0, w, \Omega_M) \tag{5}$$

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

2. Suppose the prior is of the form $P(M_0)P(\sigma^2_{\text{int}})P(\theta)P(w)P(\Omega_M)$. Write down the unnormalised posterior density of $(M_0, \sigma^2_{\text{int}}, \theta, w, \Omega_M)$ given data $\mathcal{D}$.

3. Assume flat improper priors for $w, \theta \sim U(-\infty, \infty)$, and flat positive improper priors

$$P(X) \propto \begin{cases} 1, & X \geq 0 \\ 0, & X < 0 \end{cases} \tag{6}$$

for Ω_M and σ^2_{int}. For the prior on M_0, assume a broad Gaussian: $M_0 \sim N(-19, 2^2)$. Derive the following conditional posteriors: (a) $P(M_0|\, \theta, \sigma^2, w, \Omega_M)$, (b) $P(\theta|\, M_0, \sigma^2, w, \Omega_M)$, (c) $P(\sigma^2|\, M_0, \theta, w, \Omega_M)$. Use these conditionals to construct an MCMC to sample the posterior $P(M_0, \sigma^2_{\text{int}}, H_0, w, \Omega_M|\, \mathcal{D})$ over the 5 parameters.

4. Make a change of variables to $\mathcal{M} = M_0 - 5\log h$. Let the prior on $\mathcal{M} \sim U(-\infty, \infty)$, and use the same ones above for $w, \Omega_M, \sigma^2_{\text{int}}$. Write down the joint posterior $P(\mathcal{M}, \sigma^2_{\text{int}}, w, \Omega_M|\, \mathcal{D})$ and derive the conditional posterior

$$P(\mathcal{M}, \sigma^2_{\text{int}}|w, \Omega_M;\, \mathcal{D}) = P(\mathcal{M}|\, \sigma^2_{\text{int}}, w, \Omega_M, \mathcal{D}) \times P(\sigma^2_{\text{int}}|\, w, \Omega_M, \mathcal{D}) \tag{7}$$

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.

5. Describe in practice how you implement and assess your chains. If appropriate, use your chains to compute the marginal posterior estimates of w and Ω_M, and plot their joint posterior. Compare the performance of your two algorithms.

2 Periodic Gaussian Processes

Many astronomical time-domain phenomena exhibit periodic signals (e.g. variable stars or exoplanet transits). Consider the zero-mean Gaussian process on the plane $\boldsymbol{x} \in \mathbb{R}^2$ with the squared exponential kernel: $f(\boldsymbol{x}) \sim \mathcal{GP}(\boldsymbol{0}, k(\boldsymbol{x}, \boldsymbol{x}'))$:

$$k(\boldsymbol{x}, \boldsymbol{x}') = A^2 \exp\left(-\frac{|\boldsymbol{x} - \boldsymbol{x}'|^2}{2l^2}\right). \tag{8}$$

Now consider the process $g(t) = f(\boldsymbol{u}(t))$ restricted to the circle:

$$\boldsymbol{u}(t) = \left(r \sin\frac{2\pi t}{T}, r\cos\frac{2\pi t}{T}\right). \tag{9}$$

1. Derive the covariance function $k(t, t')$ between $g(t)$ and $g(t')$ on the circle. Show that the Gaussian process on the circle is stationary. What is the period of functions drawn from this GP? Justify.

2. Suppose we have irregularly timed time-series observations of the brightness of a periodic variable star. The true brightness (in magnitudes) light curve $m(t) = g(t)$ of the star repeats every P days. The mean brightness has been subtracted, so $m(t)$ may be assumed to have a long-term average of zero. The measurement of the latent brightness $m(t_i)$ at observation time t_i is y_i with zero-mean heteroskedastic Gaussian error with known variance σ_i^2, for $i = 1, \dots, N$ data points. Assume a zero-mean Gaussian process prior and the covariance function you derived, with hyperparameters $\boldsymbol{H} = (A, \tilde{l} \equiv l/r, T)$, for the underlying light curve. Derive a marginal likelihood function $P(\boldsymbol{y}|\, \boldsymbol{t}, \boldsymbol{H})$.

3. Use the dataset provided ("variable_star.txt") containing brightness time series measurements of a new type of variable star over a time span of 1 to $\sim$ 1000 days. Estimate the period of the variable star and its 1σ uncertainty. The log likelihood function is highly multimodal, so it is important to identify the major mode. One way to begin is to compute the profile likelihood:

$$L_{\text{prof}}(T) = \max_{A, \tilde{l}} L(A, \tilde{l}, T). \tag{10}$$

To do this, loop through a fine, but wide, grid of trial T's. At each of trial T, maximise the marginal likelihood function over the other hyperparameters, and record the log likelihood. After completing the loop, plot the log profile likelihood values versus T to identify the major mode. (If you cannot get your optimiser to work, you can just plot the marginal likelihood with fixed $A = 1, \tilde{l} = 1$).

4. Having now determined estimates $\hat{\boldsymbol{H}}$ of the hyperparameters $\boldsymbol{H}$, we would like to infer the true, latent light curve at times spanning the gaps in the observations, as well as future times beyond the last observation. Let $\boldsymbol{t}^*$ be a fine grid of times spanning 1 to 2000 days. Fixing, $\boldsymbol{H} = \hat{\boldsymbol{H}}$, derive an expression for the joint posterior predictive probability of the future light curve $\boldsymbol{m}(\boldsymbol{t}^*)$, which has elements $m(t_j^*)$: $P(\boldsymbol{m}(\boldsymbol{t}^*) | \boldsymbol{t}^*, \boldsymbol{y}, \boldsymbol{t}, \hat{\boldsymbol{H}})$. Plot the expected value and standard deviation as a function of time. What is your prediction of the magnitude and its 1-σ uncertainty at $t = 1800$?

3 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:

$$df(t) = \tau^{-1}[c - f(t)]\, dt + \sigma dW_t \tag{11}$$

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

$$f(t) \sim \mathcal{GP}(m(t), k(t, t')\,) \tag{12}$$

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

$$\mathrm{Cov}[f(t), f(t')] = k(t, t') = A^2 \exp(-|t - t'|/\tau) \tag{13}$$

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. Now write down 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.

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 $\boldsymbol{\Sigma}\boldsymbol{x} = \boldsymbol{b}$, 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.