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

So far includes solutions to Problems 1 & 2

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 $h = H_0/100$ km s^{-1} Mpc^{-1}.

Solution: Let $\boldsymbol{\phi} = (w, \Omega_M)^T$. For one supernova s the probability density of its magnitude conditional on the redshift and the parameters is:

$$P(m_s|\, z_s; M_0, \sigma^2; H_0, w, \Omega_M) = N(m_s|\, M_0 + \mu(z_s; H_0, w, \Omega_M), \sigma^2)$$

Note that we can rewrite:

$$\mu(z_s; H_0, w, \Omega_M) = 25 - 5\log h + f(z; w, \Omega_M)$$

where

$$f(z;\boldsymbol{\phi}) \equiv 25 + 5\log_{10}\left[\frac{c}{100 \textbf{ km s}^{-1} \textbf{ Mpc}^{-1}}\,\tilde{d}(z;\boldsymbol{\phi})\,\textbf{Mpc}^{-1}\right]$$

Therefore the likelihood function for the sample is:

$$L(M_0, \sigma^2, w, \Omega_M) = \prod_{s=1}^{N} N(m_s|\, M_0 - \theta + f(z_s;\boldsymbol{\phi}), \sigma^2)$$

QED.

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 $\mathcal{D}$.

   Solution: The unnormalised posterior is obtained via Bayes' Theorem:

$$P(M_0, \sigma^2, \theta, \boldsymbol{\phi}|\, \mathcal{D}) \propto \left[\prod_{s=1}^{N} N(m_s|\, M_0 - \theta + f(z_s;\boldsymbol{\phi}), \sigma^2)\right] \times P(M_0)P(\sigma^2)P(\theta)P(w)P(\Omega_M)$$

   QED.

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) $P(M_0|\, \theta, \sigma^2_{\text{int}}, w, \Omega_M; \mathcal{D})$,

   Solution: Keeping track of only factors of M_0 (as everything else is held constant), we notice this conditional is

$$P(M_0|\, \theta, \sigma^2_{\textbf{int}}, w, \Omega_M; \mathcal{D}) \propto \left[\prod_{s=1}^{N} N(m_s|\, M_0 - \theta + f(z_s;\boldsymbol{\phi}), \sigma^2)\right] \times N(M_0|\, \tilde{M}_0, \tau^2)$$

   where $\tilde{M}_0 = -19$ is the prior mean and τ^2 is the prior variance. We notice that this is a product of $N+1$ Gaussian densities in the running variable M_0, and by examining the factors inside the exponentials, can be rewritten:

$$P(M_0|\, \theta, \sigma^2_{\textbf{int}}, w, \Omega_M; \mathcal{D}) \propto \left[\prod_{s=1}^{N} N(M_0|\, m_s + \theta - f(z_s;\boldsymbol{\phi}), \sigma^2)\right] \times N(M_0|\, \tilde{M}_0, \tau^2).$$

   The product of Gaussian densities is proportional to a Gaussian density, with a resulting precision (inverse variance) equal to the sum of the individual precisions, and the resulting mean equal to the precision-weighted average of the individual means.

$$P(M_0|\, \theta, \sigma^2_{\textbf{int}}, w, \Omega_M; \mathcal{D}) \propto N(M_0|\, \hat{M}_0, \hat{\sigma}^2_0)$$

where

$$\hat{M}_0 \equiv \frac{\tau^{-2}\tilde{M}_0 + \sigma^{-2}\sum_{s=1}^{N}(m_s + \theta - f(z_s;\boldsymbol{\phi}))}{\tau^{-2} + N\sigma^{-2}}$$
$$\hat{\sigma}_0^{-2} = N\sigma^{-2} + \tau^{-2}$$

By inspection, we see that the above conditional is already normalised, so

$$P(M_0|\,\theta, \sigma^2_{\mathbf{int}}, w, \Omega_M; \mathcal{D}) = N(M_0|\,\hat{M}_0, \hat{\sigma}_0^2)$$

is the conditional. QED.

(b) $P(\theta|\,M_0, \sigma^2_{\text{int}}, w, \Omega_M; \mathcal{D})$,
Solution: Similarly, we examine the only factors in the posterior only depending on θ (with everything else held constant).

$$P(\theta|\,M_0, \sigma^2, \boldsymbol{\phi}, \mathcal{D}) \propto \left[\prod_{s=1}^{N} N(m_s|\,M_0 - \theta + f(z_s;\boldsymbol{\phi}), \sigma^2)\right] \times P(\theta)$$

By reorganising inside the exponentials, and with a flat prior, we see that this conditional is proportional to the product of N Gaussian densities. Therefore,

$$\begin{aligned} P(\theta|\,M_0, \sigma^2, \boldsymbol{\phi}, \mathcal{D}) &\propto \left[\prod_{s=1}^{N} N(\theta|\,M_0 - m_s + f(z_s;\boldsymbol{\phi}), \sigma^2)\right] \\ &= N(\theta|\,\hat{\theta}, \hat{\sigma}_\theta^2) \end{aligned}$$

where $\hat{\sigma}_\theta^2 = \sigma^2/N$ and

$$\hat{\theta} = \frac{1}{N}\sum_{s=1}^{N} M_0 - m_s + f(z_s;\boldsymbol{\phi})$$

QED.

(c) $P(\sigma^2_{\text{int}}|\,M_0, \theta, w, \Omega_M; \mathcal{D})$.
Solution: For $\sigma^2 > 0$, and defining $\tilde{m}_s = M_0 - \theta + f(z;\boldsymbol{\phi})$, the conditional is

$$\begin{aligned} P(\sigma^2|\,M_0, \theta, w, \Omega_M; \mathcal{D}) &\propto \prod_{s=1}^{N} N(m_s|\,\tilde{m}_s, \sigma^2) \\ &\propto (\sigma^2)^{-N/2} \exp\left(-\frac{1}{2}Ns_N^2/\sigma^2\right) \end{aligned}$$

where $s_N^2 \equiv \frac{1}{N}\sum_{s=1}^{N}(m_s - \tilde{m}_s)$. This can be expressed as a standard Inv-χ^2 density:

$$\textbf{Inv-}\chi^2(x|\,\nu, s^2) \propto x^{-(\nu/2+1)}\exp(-\nu s^2/2x)$$

The inverse χ^2 distribution is the same as the inverse gamma distribution:

$$\textbf{Inv-}\chi^2(x|\,\nu, s^2) = \textbf{Inv-Gamma}(x|\,\alpha = \nu/2, \beta = \nu s^2/2)$$

where the inverse gamma distribution:

$$\textbf{Inv-Gamma}(x|\,\alpha, \beta) \propto x^{-(\alpha+1)}\exp(-\beta/x)$$

Thus the conditional can be expressed as:

$$\begin{aligned} P(\sigma^2 |\, M_0, \theta, w, \Omega_M; \mathcal{D}) &= \textbf{Inv-}\chi^2(\sigma^2 |\, N-2, \frac{N}{N-2} s_N^2) \\ &= \textbf{Inv-Gamma}(\sigma^2 |\, (N-2)/2, N s_N^2/2) \end{aligned}$$

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.

Solution We describe a mixed Gibbs sampler that uses these conditionals. After initialising our chains properly, at the i-th step, we sample

a) $M_0^i \sim P(M_0 | \theta_{i-1}, \sigma^2_{i-1}, \boldsymbol{\phi}_{i-1}, \mathcal{D})$**, which is a Gaussian draw.**

b) $\theta_i \sim P(\theta | M_0^i, \sigma^2_{i-1}, \boldsymbol{\phi}_{i-1}, \mathcal{D})$**, which is a Gaussian draw.**

c) $\sigma^2_i \sim P(\sigma^2 |\, M_0^i, \sigma^2_i, \boldsymbol{\phi}_{i-1}, \mathcal{D})$**, which is an inverse** χ^2 **or inverse Gamma draw.**

d) The conditional $P(\boldsymbol{\phi} | M_0, \sigma^2, \theta, \mathcal{D})$ **is not tractable (analytic), so we cannot simply draw from it. However, we can replace this Gibbs step with a Metropolis proposal and accept/reject. Using a suitable proposal 2x2 matrix** Σ_P**, we can propose a new** $\boldsymbol{\phi}^* \sim N(\boldsymbol{\phi}_{i-1}, \boldsymbol{\Sigma}_P)$**. We calculate the Metropolis ratio**

$$r = \frac{P(M_0^i, \sigma_i^2, \theta_i, \boldsymbol{\phi}^* | \mathcal{D})}{P(M_0^i, \sigma_i^2, \theta_i, \boldsymbol{\phi}_{i-1} | \mathcal{D})}$$

and accept $\boldsymbol{\phi}_i = \boldsymbol{\phi}^*$ **with probably** r **(i.e if** $u \sim U(0,1)$ **and** $u < r$**), otherwise** $\boldsymbol{\phi}_i = \boldsymbol{\phi}_{i=1}$**. QED.**

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.

Solution: Let $\mathcal{M} = M_0 - \theta$**. Then the likelihood is**

$$L(\mathcal{M}, \sigma^2, \boldsymbol{\phi}) = \left[\prod_{i=1}^{N} N(m_s |\, \mathcal{M} + f(z_s; \boldsymbol{\phi}), \sigma^2) \right]$$

And the posterior is

$$P(\mathcal{M}, \sigma^2, \boldsymbol{\phi} |\, \mathcal{D}) \propto L(\mathcal{M}, \sigma^2, \boldsymbol{\phi}) \times P(\mathcal{M}, \sigma^2, \boldsymbol{\phi})$$

Defining $\hat{\mathcal{M}}_s \equiv m_s - f(z_s; \boldsymbol{\phi})$**, the conditional is**

$$\begin{aligned} P(\mathcal{M}, \sigma^2 |\, \boldsymbol{\phi}, \mathcal{D}) &\propto \left[\prod_{i=1}^{N} N(m_s |\, \mathcal{M} + f(z_s; \boldsymbol{\phi}), \sigma^2) \right] \\ &\propto (\sigma^2)^{-N/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{N} (\hat{\mathcal{M}}_s - \mathcal{M})^2 \right) \\ &\propto (\sigma^2)^{-N/2} \exp\left(-\frac{(N-1) S^2_{\mathcal{M}}}{2\sigma^2} \right) \exp(-\frac{N}{2\sigma^2} (\mathcal{M} - \bar{\mathcal{M}})^2) \end{aligned}$$

where $S^2_{\mathcal{M}} \equiv (N-1)^{-1}\sum_{i=1}^N (\hat{\mathcal{M}}_i - \bar{\mathcal{M}})^2$ and $\bar{\mathcal{M}} = N^{-1}\sum_{i=1}^N \hat{\mathcal{M}}_i$. The last line follows analogously to Example Sheet 1, Problem 4. Similarly, this can be decomposed into the product of a conditional and marginal.

$$P(\mathcal{M}, \sigma^2 | \, \boldsymbol{\phi}, \mathcal{D}) = P(\mathcal{M}|\sigma^2, \, \boldsymbol{\phi}, \mathcal{D}) \times P(\sigma^2 | \, \boldsymbol{\phi}, \mathcal{D})$$

Similarly to Example Sheet 1, Problem 4, these factors can be found:

$$P(\mathcal{M}|\sigma^2, \, \boldsymbol{\phi}, \mathcal{D}) = N(\mathcal{M}| \, \bar{\mathcal{M}}, \sigma^2/N)$$

$$P(\sigma^2 | \, \boldsymbol{\phi}, \mathcal{D}) \propto (\sigma^2)^{-N/2+1/2} \exp\left(-\frac{(N-1)S^2_{\mathcal{M}}}{2\sigma^2}\right)$$

$$\propto \textbf{Inv-}\chi^2(\sigma^2 | \, N+1, (N-1)/(N+1)S^2_{\mathcal{M}})$$

A mixed Gibbs sampler can be composed as follows. After initialising the chains, at each step i,

a) Jointly draw $(\mathcal{M}_i, \sigma_i^2) \sim P(\mathcal{M}, \sigma^2 | \boldsymbol{\phi}, \mathcal{D})$ directly by sampling the sequence:

a.1) $\sigma^2 \sim P(\sigma^2 | \boldsymbol{\phi}_{i-1}, \mathcal{D})$, which is an inverse χ^2 or inverse Gamma distribution.

a.2) $\mathcal{M}_i | \, \sigma_i^2 \sim P(\mathcal{M}| \, \sigma_i^2, \boldsymbol{\phi}_{i-1}, \mathcal{D})$, which is an Gaussian draw.

b. The conditional $P(\phi|\mathcal{M}, \sigma^2, \mathcal{D})$ is not tractable (analytic), so we cannot simply draw from it. However, we can replace this Gibbs step with a Metropolis proposal and accept/reject. Update $\boldsymbol{\phi}$ by Metropolis. Using a suitable proposal 2x2 matrix Σ_P, we can propose a new $\boldsymbol{\phi}^* \sim N(\boldsymbol{\phi}_{i-1}, \boldsymbol{\Sigma}_P)$. We calculate the Metropolis ratio

$$r = \frac{P(\mathcal{M}_i, \sigma_i^2, \boldsymbol{\phi}^* | \, \mathcal{D})}{P(\mathcal{M}_i, \sigma_i^2, \boldsymbol{\phi}_{i-1} | \, \mathcal{D})}$$

and accept $\boldsymbol{\phi}_i = \boldsymbol{\phi}^*$ with probably r (i.e if $u \sim U(0,1)$ and $u < r$), otherwise $\boldsymbol{\phi}_i = \boldsymbol{\phi}_{i=1}$. QED.

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.

   Solution: We would typically start by attempting to find the mode of the joint posterior by optimisation. Then we can start several chains with initial parameter values dispersed around the mode according to the inverse of the observed Fisher matrix at the mode. We can assess burn-in and convergence using the Gelman-Rubin ratio to compare within-chain variance to between-chain variance. We can also compute the autocorrelation timescale of each parameter in each single chain to assess the number of independent samples obtained, and make sure we have a sufficient number. We can (optionally) thin the chain by the largest autocorrelation timescale. We remove burn in and thin the chains and combine the chains to compute posterior inferences of the parameters.

   We expect the first mixed Gibbs sampler to be slow. The likelihood function only constrains the linear combination $M_0 - \theta$ so they are separately non-identifiable within the likelihood. These parameters will be highly correlated in the posterior, causing the sampling to be slow. By reparametrisation $\mathcal{M} = M_0 - \theta$ we remove that degeneracy. Furthermore, we draw $(\mathcal{M}, \sigma^2)$ jointly as a block Gibbs step rather than via separate Gibbs steps, so that also improves the sampling of the posterior.

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: $f(\boldsymbol{x}) \sim \mathcal{GP}(\boldsymbol{0}, k(\boldsymbol{x}, \boldsymbol{x}'))$:

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

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

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

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.

   Solution: The covariance of the function on the circle is:

   $$k(t, t') = \mathbf{Cov}[g(t), g(t')] = \mathbf{Cov}[f(\boldsymbol{u}(t)), f(\boldsymbol{u}(t'))] = A^2 \exp\left(-\frac{|\boldsymbol{u}(t) - \boldsymbol{u}(t')|^2}{2l^2}\right)$$

   Now we make use of trigonometric identities, with $\theta = 2\pi t/T$ and $\theta' = 2\pi t'/T$,

   $$\begin{aligned}
   |\boldsymbol{u} - \boldsymbol{u}'|^2 &= r^2(\sin\theta - \sin\theta')^2 + r^2(\cos\theta - \cos\theta')^2 \\
   &= r^2\left[2\sin\left(\frac{\theta-\theta'}{2}\right)\cos\left(\frac{\theta+\theta'}{2}\right)\right]^2 + r^2\left[-2\sin\left(\frac{\theta+\theta'}{2}\right)\sin\left(\frac{\theta-\theta'}{2}\right)\right]^2 \\
   &= r^2\sin^2\left(\frac{\theta-\theta'}{2}\right)\left[4\left(\cos^2\left(\frac{\theta+\theta'}{2}\right) + \sin^2\left(\frac{\theta+\theta'}{2}\right)\right)\right] \\
   &= 4r^2\sin^2\left(\frac{\theta-\theta'}{2}\right)
   \end{aligned}$$

   Therefore the covariance on the circle is:

   $$k(t, t') = A^2 \exp\left(-\frac{2r^2}{l^2}\sin^2\left(\pi(t-t')/T\right)\right)$$

   This kernel is stationary because it is only a function of t and t' through their difference $t - t'$. Thus the covariance is invariant to time shifts $t, t' \to t + c, t' + c$. Since $\sin^2\theta$ repeats every π radians, the maximal covariance is when $t = t' + kT$ where k is any integer. Thus, the pattern will repeat with period T.

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 $i = 1, \ldots, N$ 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 $P(\boldsymbol{y}|\, \boldsymbol{t}, \boldsymbol{H})$.

Solution: The GP kernel is:

$$k_{\boldsymbol{\theta}}(t,t') = A^2 \exp\left(-\frac{2}{\tilde{l}^2}\sin^2\left(\pi(t-t')/T\right)\right)$$

where the hyperparameters are: $\boldsymbol{\theta} = (A, \tilde{l}, T)$. Let $\boldsymbol{t} = (t_1, \dots t_N)^T$ and $\boldsymbol{m} = (m(t_1), \dots, m(t_N))^T$. The the GP prior for the latent (true) magnitudes is:

$$\boldsymbol{m}|\,\boldsymbol{\theta} \sim N(\boldsymbol{0}, \boldsymbol{K}_{\boldsymbol{\theta}}(\boldsymbol{t},\boldsymbol{t}))$$

where the elements $[K_{\boldsymbol{\theta}}(\boldsymbol{t},\boldsymbol{t})]_{ij} = k_{\boldsymbol{\theta}}(t_i, t_j)$ are obtained by evaluating the kernel between every pair of observation times. The noisy measurement process for each datum $i = 1, \dots, N$ is $y_i|m_i \sim N(m_i, \sigma_i^2)$, which yields a likelihood function

$$P(\boldsymbol{y}|\,\boldsymbol{m}) = \prod_{i=1}^{N} N(y_i|\,m_i, \sigma_i^2) = N(\boldsymbol{y}|\,\boldsymbol{m}, \boldsymbol{W})$$

where $\boldsymbol{W}$ is diagonal with elements $W_{ij} = \sigma_i^2\,\delta_{ij}$. The joint distribution of the observed data and latent GP values is

$$P(\boldsymbol{y}, \boldsymbol{m}|\,\boldsymbol{\theta}) = P(\boldsymbol{y}|\,\boldsymbol{m})P(\boldsymbol{m}|\boldsymbol{\theta}).$$

The marginal likelihood is obtained by integrating out the unseen latent values $\boldsymbol{m}$.

$$\begin{aligned} L(\boldsymbol{\theta}) = P(\boldsymbol{y}|\,\boldsymbol{\theta}) = \int P(\boldsymbol{y}|\,\boldsymbol{m})P(\boldsymbol{m}|\,\boldsymbol{\theta})\,d\boldsymbol{m} &= \int N(\boldsymbol{y}|\,\boldsymbol{m}, \boldsymbol{W})N(\boldsymbol{m}|\,\boldsymbol{0}, \boldsymbol{K}_{\boldsymbol{\theta}}(\boldsymbol{t}, \boldsymbol{t}'))\,d\boldsymbol{m} \\ &= N(\boldsymbol{y}|\boldsymbol{0}, \boldsymbol{K}_{\boldsymbol{\theta}}(\boldsymbol{t},\boldsymbol{t}) + \boldsymbol{W}) \end{aligned}$$

QED.

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:

$$L_{\text{prof}}(T) = \max_{A,\tilde{l}} L(A, \tilde{l}, T). \tag{8}$$

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 $A = 1, \tilde{l} = 1$).

Solution: We scan the profile log likelihood function over a grid of T values. We find that the best period is $\hat{T} = 199.50$. Insert plots.

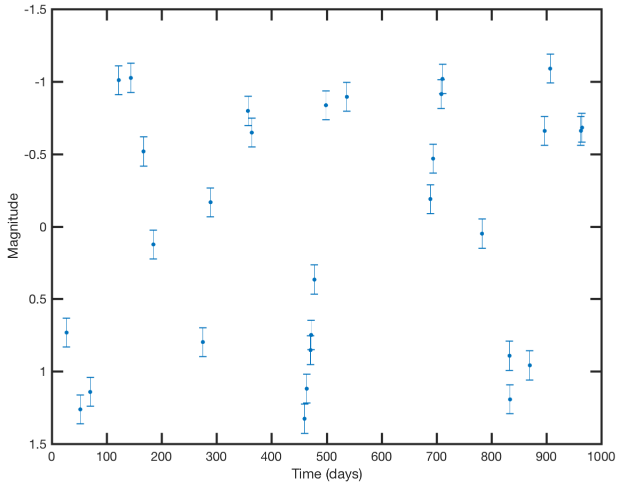

Figure 1: **Plot of data.**

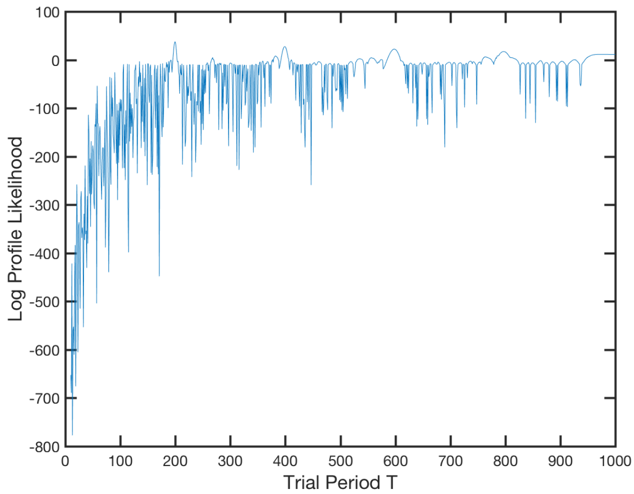

Figure 2: **Plot of log profile likelihood on a wide grid. Note that the likelihood is highly multimodal, which could make MCMC sampling difficult. We see that there are major modes roughly at multiples of 200 days. Aliasing means that a signal with period T is also periodic on integer multiples of T. We seek the smallest period.**

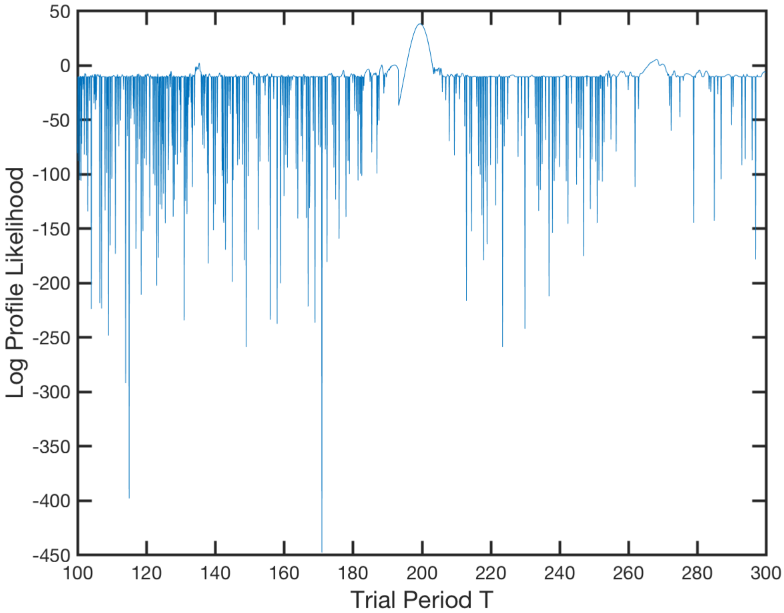

Figure 3: **Plot of log profile likelihood on a narrower grid.**

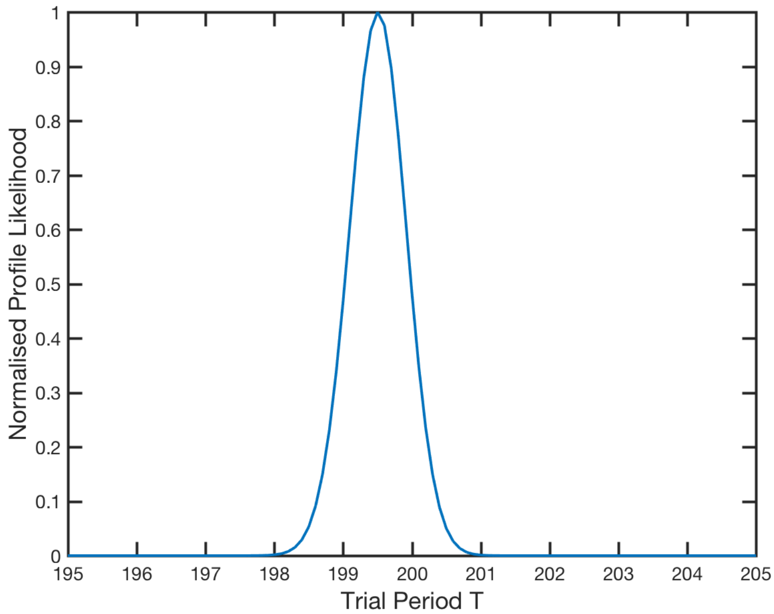

Figure 4: **Plot of profile likelihood on a narrower grid, focusing on the main mode.**

Once we find the maximum profile likelihood L, we run the optimiser again near the peak to obtain the minimum negative log likelihood and its Hessian over the 3 hyperparamers.

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} -\log P(\boldsymbol{y}|\boldsymbol{\theta})$$

$$\boldsymbol{H} = \boldsymbol{\nabla\nabla}[-\log P(\boldsymbol{y}|\,\boldsymbol{\theta})]\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \hat{\boldsymbol{I}}$$

which is the observed Fisher matrix:

$$\hat{I}_{ij} = -\frac{\partial^2 P(\boldsymbol{y}|\,\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}.$$

An estimate of the uncertainty on the period is given by the $T-T$ element of the inverse of the observed Fisher matrix (although this is really a lower bound). Let $\boldsymbol{\Sigma} = \boldsymbol{I}^{-1}$. Then $\sigma_T^2 \geq \Sigma_{TT}$. We obtain $\hat{A} = 0.96 \pm 1.09$, $\hat{\hat{l}} = 1.11 \pm 0.68$, and $\hat{T} = 199.51 \pm 0.41$. See code.

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 $t = 1800$?

Solution: Suppose at other M times $\boldsymbol{t}^* = (t_1^*, \ldots t_N^*)^T$, we want to infer the latent values $\boldsymbol{m}^* = (m(t_1^*), \ldots, m(t_M^*))^T$. We can construct the joint density of the observed data $\boldsymbol{y}$ and these latent values $\boldsymbol{m}^*$. The joint distribution of $\boldsymbol{y}$ and m is

$$\begin{pmatrix} \boldsymbol{y} \\ \boldsymbol{m} \end{pmatrix} \sim N\left(\begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{K}(\boldsymbol{t},\boldsymbol{t}) + \boldsymbol{W} & \boldsymbol{K}(\boldsymbol{t},\boldsymbol{t}^*) \\ \boldsymbol{K}(\boldsymbol{t}^*,\boldsymbol{t}) & \boldsymbol{K}(\boldsymbol{t}^*,\boldsymbol{t}^*) \end{pmatrix}\right).$$

The covariance matrix components can be derived by considering $\boldsymbol{y} = \boldsymbol{m} + \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon} \sim N(0, \boldsymbol{W})$, and computing Cov$[\boldsymbol{y}, \boldsymbol{y}]$, Cov$[\boldsymbol{y}, \boldsymbol{m}^*]$, Cov$[\boldsymbol{m}, \boldsymbol{m}^*]$, Cov$[\boldsymbol{m}^*, \boldsymbol{m}^*]$, etc. For example:

$$\mathbf{Cov}[\boldsymbol{y},\boldsymbol{y}] = \mathbf{Cov}[\boldsymbol{m},\boldsymbol{m}] + \mathbf{Cov}[\boldsymbol{\epsilon},\boldsymbol{\epsilon}] = \boldsymbol{K_\theta}(\boldsymbol{t},\boldsymbol{t}) + \boldsymbol{W}$$

because the GP and measurement process are independent. Then using conditional property of multivariate Gaussians random vectors, we can compute the posterior of $\boldsymbol{f}_g$ given $\boldsymbol{y}_1$.

$$\boldsymbol{m}^*|\,\boldsymbol{y} \sim N\left(\mathbb{E}[\boldsymbol{m}^*|\,\boldsymbol{y}], \mathbf{Var}[\boldsymbol{m}^*|\,\boldsymbol{y}]\right)$$

$$\mathbb{E}[\boldsymbol{m}^*|\,\boldsymbol{y}] = \boldsymbol{0} + \boldsymbol{K_\theta}(\boldsymbol{t}^*,\boldsymbol{t})[\,\boldsymbol{K_\theta}(\boldsymbol{t},\boldsymbol{t}) + \boldsymbol{W}]^{-1}(\boldsymbol{m}^* - \boldsymbol{0})$$

$$\mathbf{Var}[\boldsymbol{m}^*|\,\boldsymbol{y}] = \boldsymbol{K_\theta}(\boldsymbol{t}^*,\boldsymbol{t}^*) - \boldsymbol{K_\theta}(\boldsymbol{t}^*,\boldsymbol{t})[\,\boldsymbol{K_\theta}(\boldsymbol{t},\boldsymbol{t}) + \boldsymbol{W}]^{-1}\boldsymbol{K_\theta}(\boldsymbol{t},\boldsymbol{t}^*)$$

At $t^* = 1800$, we find $m(t^*) = 0.47 \pm 0.16$.

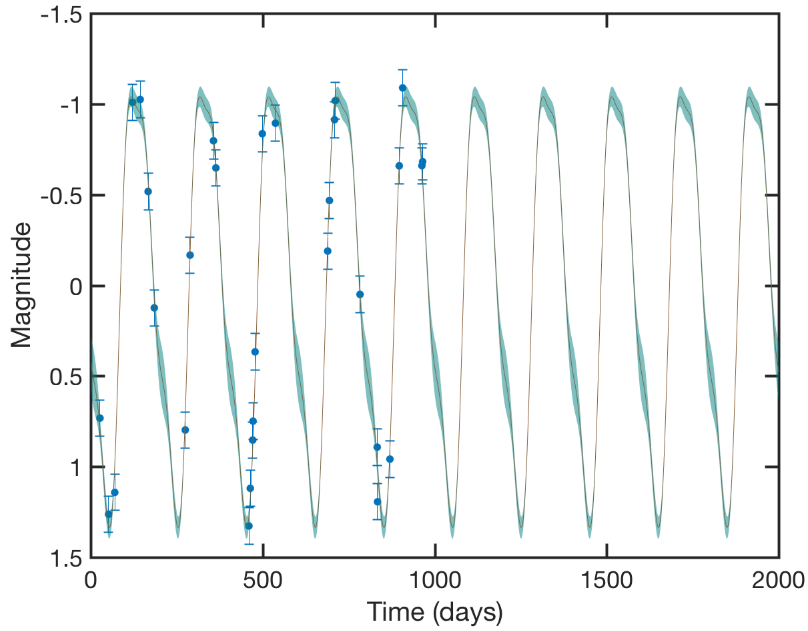

Figure 5: **GP posterior predictive mean and standard deviation of the underlying latent and future brightness curve.**

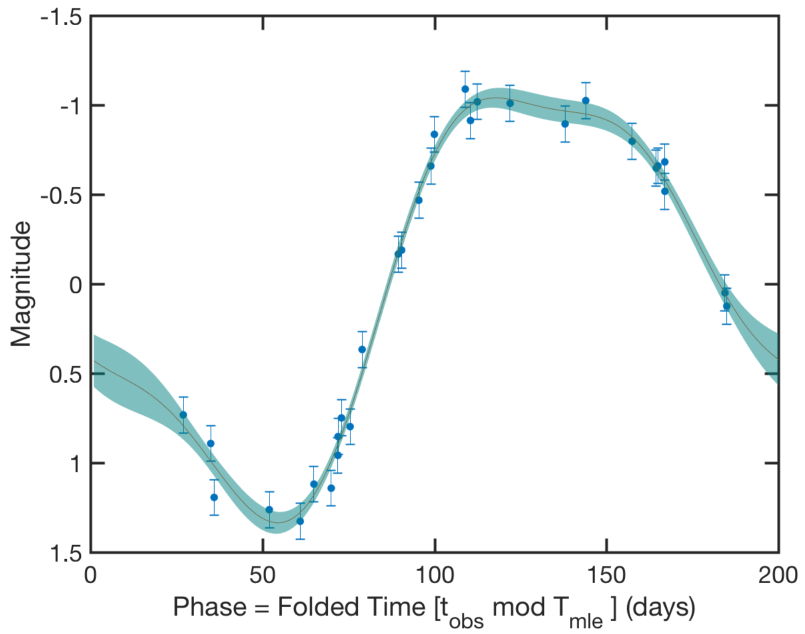

Figure 6: **Phase curve of GP posterior predictive mean and standard deviation vs. data phased by the estimated period. Phase = time mod period.**

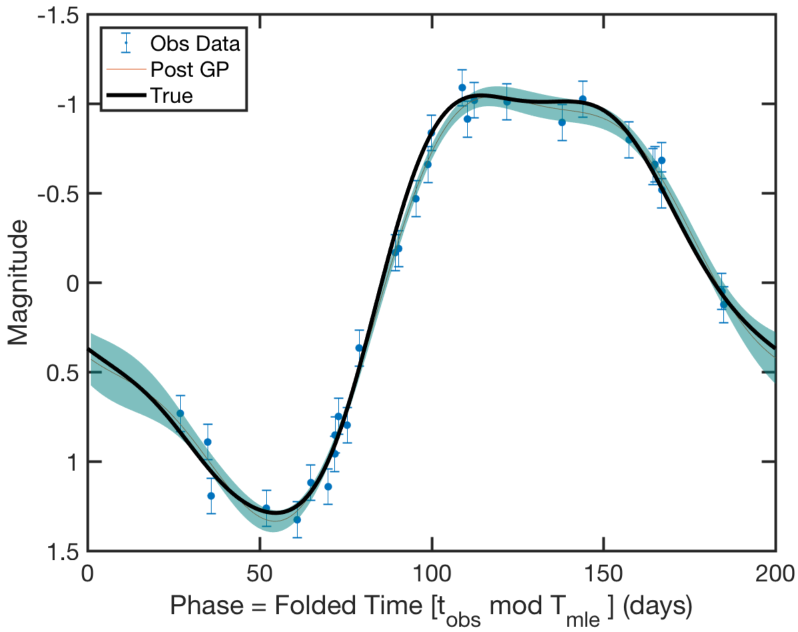

Figure 7: **Phase curve of GP posterior predictive mean and standard deviation vs. data phased by the estimated period. Phase = time mod period. Plotted against the true phase curve.**

3 Doubly Lensed Quasar Time Delay Estimation

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

$$df(t) = \tau^{-1}[c - f(t)]\, dt + \sigma dW_t \tag{9}$$

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

$$f(t) \sim \mathcal{GP}(m(t), k(t, t')\,) \tag{10}$$

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

$$\mathrm{Cov}[f(t), f(t')] = k(t, t') = A^2 \exp(-|t - t'|/\tau) \tag{11}$$

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.