Part III Astrostatistics:
Solutions for Example Sheet 4
Example Class: Thursday, 26 Apr 2018, 1:00pm, MR5

1 Combining Uncertain Estimates

Suppose two Type Ia supernovae are observed in a single external galaxy. Since the distance from Earth to the galaxy is so much greater than the size of the galaxy, we can assume that the two supernovae, and the galaxy, are all at nearly the same distance from Earth. Analysis of the supernovae yield unbiased distance estimates $\hat{d}_1$ and $\hat{d}_2$, which have Gaussian error with known standard deviations σ_1 and σ_2.

1. Consider all estimators that are linear combinations of the two distances: $\hat{d} = \alpha_1\hat{d}_1 + \alpha_2\hat{d}_2$. What relation between the coefficients is required of the subset of unbiased estimators?

   Solution: The expectation of these linear estimators is:

   $$\mathbb{E}[\hat{d}] = \alpha_1\mathbb{E}[\hat{d}_1] + \alpha_2\mathbb{E}[\hat{d}_2] = (\alpha_1 + \alpha_2)d \tag{1}$$

   since the estimators are unbiased: $\mathbb{E}[\hat{d}_i] = d$. Therefore the linear combination $\hat{d}$ is unbiased only when $\alpha_1 + \alpha_2 = 1$. Let us now define $\alpha = \alpha_1 = 1 - \alpha_2$.

2. Find the unbiased linear estimator with minimum variance. What is the variance of this estimator?

   Solution: The variance of a random variable is the covariance of it with itself, and the covariance is a bilinear operator in each of its arguments (you can take out multiplicative constants and sums).

   $$\begin{aligned}\mathbf{Var}[\hat{d}] = \mathbf{Cov}[\hat{d},\hat{d}] &= \mathbf{Cov}[\alpha\hat{d}_1 + (1-\alpha)\hat{d}_2, \alpha\hat{d}_1 + (1-\alpha)\hat{d}_2] \\ &= \alpha^2\,\mathbf{Var}[\hat{d}_1] + (1-\alpha)^2\,\mathbf{Var}[\hat{d}_2] = \alpha^2\sigma_1^2 + (1-\alpha)^2\sigma_2^2\end{aligned} \tag{2}$$

   where we have assumed that the measurements of each supernova are independent so their covariance is zero: $\mathbf{Cov}[\hat{d}_1, \hat{d}_2] = 0$. Let us denote this expression $V(\alpha)$ as a function of α. The conditions for a minimum are that $V'(\alpha) = 0$ and $V''(\alpha) > 0$. We find:

   $$V'(\alpha) = 2\alpha\sigma_1^2 - 2(1-\alpha)\sigma_2^2 = 0 \tag{3}$$

   which gives us

   $$\alpha_{\mathbf{min}} = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} = \frac{\sigma_1^{-2}}{\sigma_1^{-2} + \sigma_2^{-2}} \tag{4}$$

   and

   $$\alpha_{2,\mathbf{min}} = 1 - \alpha_{\mathbf{min}} = \frac{\sigma_2^{-2}}{\sigma_1^{-2} + \sigma_2^{-2}} \tag{5}$$

Thus, the optimal coefficient of $\hat{d}_i$ is proportional to its inverse variance, or precision, σ_i^{-2}. The second derivative is

$$V''(\alpha) = 2(\sigma_1^2 + \sigma_2^2) > 0. \tag{6}$$

The minimal variance is then

$$V(\alpha_{\mathbf{min}}) = \frac{1}{\sigma_1^{-2} + \sigma_2^{-2}}. \tag{7}$$

The inverse of the variance is called the precision, so the maximal precision is simply the sum of the individual precisions.

$$V(\alpha_{\mathbf{min}})^{-1} = \sigma_1^{-2} + \sigma_2^{-2}. \tag{8}$$

3. Suppose the standard deviations are the same $\sigma_1 = \sigma_2 = \sigma$. What is the minimum variance of the combined estimate?

   Solution: Plugging in, we find $V(\alpha_{\mathbf{min}}) = \sigma^2/2$. Thus the standard deviation is reduced by square root of 2.

4. Suppose because of observational systematic errors, the distance errors are correlated with correlation coefficient ρ, such that $|\rho| < 1$. What is the minimum variance unbiased linear estimator in this case?

   Solution: Again we need $\alpha = \alpha_1 = 1 - \alpha_2$. The covariance $\mathbf{Cov}[\hat{d}_1, \hat{d}_2] = \sigma_1\sigma_2\rho$. Now,

$$V(\alpha) = \mathbf{Var}[\hat{d}] = \mathbf{Cov}[\hat{d}_1, \hat{d}_2] = \alpha^2\sigma_1^2 + (1-\alpha)^2\sigma_2^2 + 2\alpha(1-\alpha)\sigma_1\sigma_2\rho \tag{9}$$

   Solving $V'(\alpha) = 0$, we find

$$\alpha_{\mathbf{min}} = \frac{\sigma_1^{-2} - \sigma_1^{-1}\sigma_2^{-1}\rho}{\sigma_1^{-2} + \sigma_2^{-2} - 2\sigma_1^{-1}\sigma_2^{-1}\rho} \tag{10}$$

   We verify:

$$\begin{aligned} V''(\alpha) &= 2\sigma_1^2 + 2\sigma_2^2 - 4\sigma_1\sigma_2\rho \\ &> 2\sigma_1^2 + 2\sigma_2^2 - 4\sigma_1\sigma_2 \\ &> 2(\sigma_1 - \sigma_2)^2 > 0 \end{aligned} \tag{11}$$

   So this is a minimum.

2 Gaussian Processes as Infinite Basis Expansions

Functions drawn from a Gaussian process prior often have an equivalent description as arising from a linear combination of an infinite set of basis functions. Consider a finite set of $J > 2$ basis functions with a Gaussian shape centred at values c_i,

$$\phi_i(x) = \exp\left[-\frac{(x - c_i)^2}{l^2}\right] \tag{12}$$

defined on the real line $x \in \mathbb{R}$. The centres span a distance $c_J - c_1 = h$, and the centres are spaced so that $\Delta c = c_{i+1} - c_i = h/(J-1)$. Suppose a function is formed as a linear combination of these functions:

$$f(x) = \sum_{i=1}^{J} w_i \, \phi_i(x). \tag{13}$$

Suppose we put a Gaussian prior on the coefficients, $w_i \sim N(0, \sigma^2 h/J)$.

1. What is the mean $\mathbb{E}[f(x)]$ and the covariance function $k(x, x') = \mathrm{Cov}[f(x), f(x')]$?

   Solution: The expectation is

$$\mathbb{E}[f(x)] = \sum_{i=1}^{J} \phi_i(x)\mathbb{E}(w_i) = 0 \tag{14}$$

   The kernel is

$$\begin{aligned} k(x, x') = \mathbf{Cov}[f(x), f(x')] &= \mathbf{Cov}\left[\sum_{i=1}^{J} w_i\phi_i(x), \sum_{j=1}^{J} w_i\phi_i(x')\right] \\ &= \sum_{i=1}^{J}\sum_{j=1}^{J} \phi_i(x)\phi_j(x')\mathbf{Cov}[w_i, w_j] \\ &= \sum_{i=1}^{J}\sum_{j=1}^{J} \phi_i(x)\phi_j(x')\,\delta_{ij}\,\sigma^2 h/J \\ &= \sum_{i=1}^{J} \phi_i(x)\phi_i(x')\sigma^2 h/J \\ &= \sigma^2 \sum_{i=1} \phi_i(x)\phi_i(x')\frac{J-1}{J}\Delta c \end{aligned} \tag{15}$$

   where δ_{ij} is a Kronecker delta function.

2. Derive the kernel function $k(x, x')$ in the limit of an infinite number of basis functions spanning the real line: $J \to \infty$ and $c_1 \to -\infty$, $h \to \infty$.

   Solution: In the limit of $J \to \infty$, this Riemann sum becomes the integral

$$\begin{aligned} k(x, x') &= \sigma^2 \int_{c_1}^{c_J = c_1 + h} \phi_i(x)\phi_i(x')\, dc \\ &= \sigma^2 \int_{c_1}^{c_1+h} e^{-(x-c)^2/l^2} e^{-(x'-c)^2/l^2}\, dc \end{aligned} \tag{16}$$

   Now letting the basis span the real line, $c_1, h \to \infty$, we have

$$k(x, x') = \sigma^2 \int_{-\infty}^{+\infty} e^{-(x-c)^2/l^2} e^{-(x'-c)^2/l^2}\, dc \tag{17}$$

   Noting that

$$e^{-(x-c)^2/l^2} = \frac{l}{\sqrt{2}}\sqrt{2\pi}N(x|c, l^2/2) = \frac{l}{\sqrt{2}}\sqrt{2\pi}N(c|x, l^2/2), \tag{18}$$

we find

$$\begin{aligned} k(x,x') &= \sigma^2 \left(\frac{l}{\sqrt{2}}\sqrt{2\pi}\right)^2 \int_{-\infty}^{+\infty} N(x|\, c, l^2/2) N(c|\, x', l^2/2)\, dc \\ &= \sigma^2 \left(\frac{l}{\sqrt{2}}\sqrt{2\pi}\right)^2 N(x|\, x', l^2) \end{aligned} \tag{19}$$

We recognised the integral from previous Gaussian marginalisation examples. Finally,

$$k(x,x') = \frac{l\sigma^2}{2}\sqrt{2\pi} e^{-(x-x')^2/2l^2} \tag{20}$$

Therefore, the squared exponential kernel generates functions that are linear combinations of Gaussian functions of width l, distributed densely along the real line.

3. What is the variance of the resulting Gaussian process at any x?

   Solution: $\mathbf{Var}(f(x)) = k(x,x) = \sqrt{\frac{\pi}{2}} l\sigma^2$.

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

$$k(\boldsymbol{x}, \boldsymbol{x}') = A^2 \exp\left(-\frac{|\boldsymbol{x} - \boldsymbol{x}'|^2}{2l^2}\right). \tag{21}$$

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

$$\boldsymbol{u}(t) = \left(r\sin\frac{2\pi t}{T}, r\cos\frac{2\pi t}{T}\right). \tag{22}$$

1. Derive the covariance $k(t,t')$ between $g(t)$ and $g(t')$. Show that the Gaussian process on the circle is stationary.

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

$$k(t,t') = \mathbf{Cov}[g(t), g(t')] = \mathbf{Cov}[f(\boldsymbol{u}(t)), f(\boldsymbol{u}(t'))] = A^2 \exp\left(-\frac{|\boldsymbol{u}(t) - \boldsymbol{u}(t')|^2}{2l^2}\right) \tag{23}$$

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

$$\begin{aligned} |\boldsymbol{u} - \boldsymbol{u}'|^2 &= r^2(\sin\theta - \sin\theta')^2 + r^2(\cos\theta - \cos\theta')^2 \\ &= r^2\left[2\sin\left(\frac{\theta-\theta'}{2}\right)\cos\left(\frac{\theta+\theta'}{2}\right)\right]^2 + r^2\left[-2\sin\left(\frac{\theta+\theta'}{2}\right)\sin\left(\frac{\theta-\theta'}{2}\right)\right]^2 \\ &= r^2\sin^2\left(\frac{\theta-\theta'}{2}\right)\left[4\left(\cos^2\left(\frac{\theta+\theta'}{2}\right) + \sin^2\left(\frac{\theta+\theta'}{2}\right)\right)\right] \\ &= 4r^2\sin^2\left(\frac{\theta-\theta'}{2}\right) \end{aligned} \tag{24}$$

Therefore the covariance on the circle is:

$$k(t,t') = A^2 \exp\left(-\frac{2r^2}{l^2}\sin^2\left(\pi(t-t')/T\right)\right) \tag{25}$$

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 $t, t' \to t+c, t'+c$.

2. What is the period of functions drawn from this GP? Justify.

   Solution: Since $\sin^2\theta$ repeats every π radians, the maximal covariance is when $t = t' + k\pi$ where k is any integer.

3. (*non-exam*). Draw random functions $g(t) \sim \mathcal{GP}(\mathbf{0}, k(t,t'))$ to verify their period and explore their behaviour as you vary $\tilde{l} = l/r$ from small to large.

   Solution:

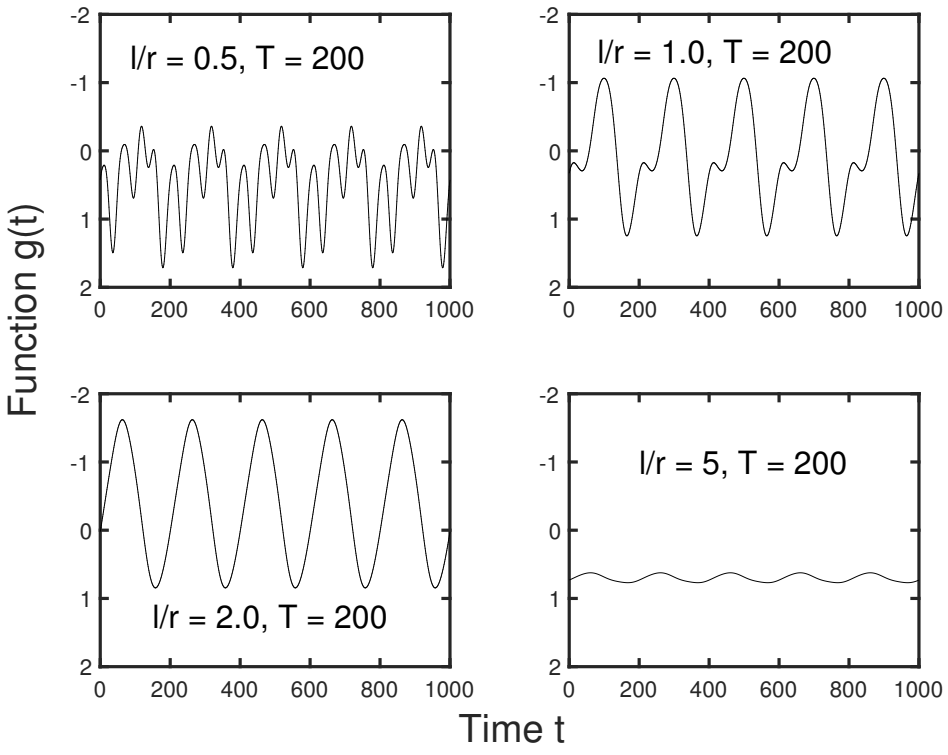

Figure 1: Random functions $g(t)$ drawn from the GP with a periodic kernel with various scale hyperparameters.

4 Probabilistic Graphical Model and Gibbs Sampling for the Normal-Normal Hierarchical Model

Consider our Normal-Normal hierarchical Bayesian model for supernova magnitudes. The true, latent absolute magnitudes are drawn from a Gaussian distribution with the population mean μ and variance τ^2 as hyperparameters:

$$M_s|\,\mu,\tau^2 \sim N(\mu,\tau^2). \tag{26}$$

Each absolute magnitude is observed to yield the data D_s with measurement error with known variance σ_s^2.

$$D_s|\,M_s \sim N(M_s,\sigma_s^2) \tag{27}$$

Suppose we observe $s = 1,\ldots,N_{\rm SN}$ independent supernovae.

1. For a single supernova s, write down the joint probability density of the datum D_s and the latent variable M_s, conditional on the hyperparameters μ,τ^2.

   Solution:

   $$P(D_s,M_s|\,\mu,\tau^2) = P(D_s|\,M_s)P(M_s|\,\mu,\tau^2) = N(D_s|\,M_s,\sigma_s^2)N(M_s|\,\mu,\tau^2) \tag{28}$$

2. For all the $N_{\rm SN}$ supernovae, write down the joint probability density of all the data $\mathcal{D} = \{D_s\}$ and the latent variables $\{M_s\}$ given the hyperparameters.

   Solution:

   $$P(\{D_s,M_s\}|\,\mu,\tau^2) = \prod_{s=1}^{N} P(D_s,M_s|\,\mu,\tau^2) = \prod_{s=1}^{N} N(D_s|M_s,\sigma_s^2)N(M_s|\,\mu,\tau^2) \tag{29}$$

3. Adopt a "non-informative" hyperpriors on the hyperparameters $P(\mu,\tau^2) \propto 1$, $\tau^2 > 0$. Write down the joint probability density of all data $\mathcal{D}$, latent variables $\{M_s\}$, and hyperparameters μ,τ^2.

   Solution: If the prior is $P(\mu,\tau^2) \propto (\tau^2)^\alpha$, $\tau^2 > 0$ (in this case, $\alpha = 0$), then the joint density is:

   $$\begin{aligned} P(\{D_s,M_s\},\mu,\tau^2) &= P(\{D_s,M_s\}|\mu,\tau^2)P(\mu,\tau^2) \\ &\propto (\tau^2)^\alpha \prod_{s=1}^{N} N(D_s|\,M_s,\sigma^2)N(M_s|\,\mu,\tau^2) \end{aligned} \tag{30}$$

4. Draw a probabilistic graphical model or directed acyclic graph representing this joint probability density. **Solution: See Fig. 2 or Fig. 3.**

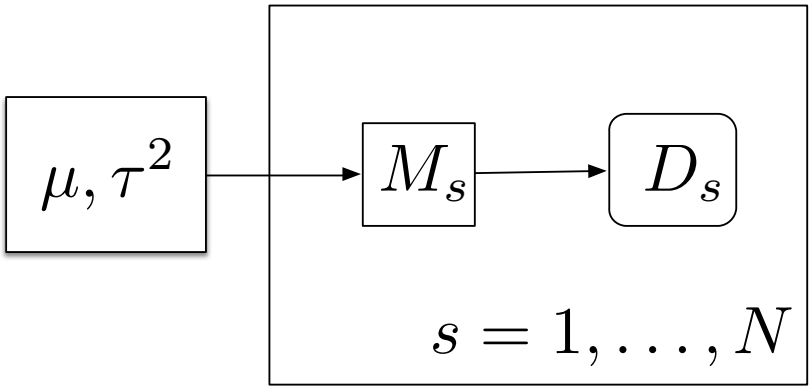

Figure 2: Probabilistic Graphical Model for Normal-Normal hierarchical Bayesian model.

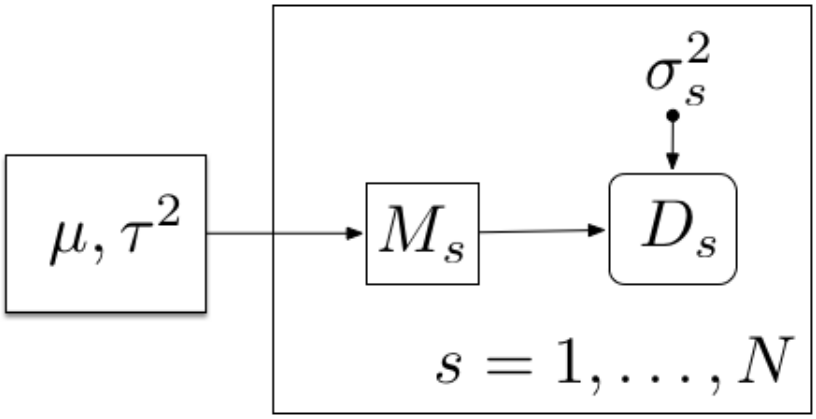

Figure 3: Probabilistic Graphical Model for Normal-Normal hierarchical Bayesian model, optionally including the constants $\{\sigma_s^2\}$.

5. Construct a Gibbs sampling algorithm to sample from the full posterior $P(\{M_s\}, \mu, \tau^2 | \mathcal{D})$ by deriving a complete set of tractable conditional posterior densities. You may choose to draw "blocks" or parameter subsets from their conditionals in a single step, rather than drawing a single parameter in each step. You may assume that you have access to algorithms that generate random draws from Gaussian distributions any mean and (positive) variance, and inverse-gamma distributions with shape parameter $a > 0$ and scale parameter $b > 0$:

$$\text{Inv-Gamma}(x|\, a, b) \propto x^{-(a+1)} \exp(-b/x), x > 0. \tag{31}$$

Solution: The posterior density of the unknowns given the knowns (observed data) is

$$\begin{aligned} P(\{M_s\}, \mu, \tau^2 | \mathcal{D}) &= P(\mathcal{D}, \{M_s\}, \mu, \tau^2)/P(\mathcal{D}) \\ &\propto \prod_{s=1}^{N} N(D_s|\, M_s, \sigma_s^2) N(M_s|\mu, \tau^2) \end{aligned} \tag{32}$$

There are many ways to construct a valid Gibbs sampler. We need only sample a complete set of conditional posterior distributions, so that every parameter is updated in a cycle. A natural way to construct a Gibbs sampler for this model is to alternately sample from $\mu, \tau^2 | \{M_s\}, \mathcal{D}$ and then $M_s | \mu, \tau^2, \mathcal{D}$ for all s.

(a) $\mu, \tau^2 \sim P(\mu, \tau^2 | \{M_s\}, \mathcal{D})$. Sample μ, τ^2 jointly conditional on (holding fixed) all the $\{M_s\}$ (and the data $\mathcal{D} = \{D_s\}$). To this, we derive:

$$P(\mu, \tau^2 | \{M_s\}, \mathcal{D}) = P(\mu, \tau^2 | \{M_s\}) \tag{33}$$

This is because, conditional on $\{M_s\}$, the posterior density of μ, τ^2 does not depend on $\mathcal{D}$. Put another way, μ, τ^2 are *conditionally independent* from $\mathcal{D}$ given $\{M_s\}$. This can be seen from the graph, or from just inspecting the joint posterior density above. Let $\bar{M} = \frac{1}{N} \sum_{s=1}^{N} M_s$ be the sample mean of the current values of the absolute magnitudes.

$$\begin{aligned} P(\mu, \tau^2 | \{M_s\}, \mathcal{D}) = P(\mu, \tau^2 | \{M_s\}) &\propto \prod_{s=1}^{N} N(M_s | \mu, \tau^2) \propto \prod_{s=1}^{N} (2\pi\tau^2)^{-1/2} e^{-(M_s-\mu)^2/2\tau^2} \\ &\propto \tau^{-N} \exp\left(-\frac{1}{2\tau^2} \sum_{s=1}^{N} (M_s - \mu)^2 \right) \\ &\propto \tau^{-N} \exp\left(-\frac{1}{2\tau^2} \sum_{s=1}^{N} (M_s - \bar{M} + \bar{M} - \mu)^2 \right) \\ &\propto \tau^{-N} \exp\left(-\frac{1}{2\tau^2} \sum_{s=1}^{N} \left[(M_s - \bar{M})^2 + 2(\bar{M} - \mu)(M_s - \bar{M}) + (\bar{M} - \mu)^2 \right] \right) \\ &\propto \tau^{-N} \exp\left(-\frac{1}{2\tau^2} \left[\left(\sum_{s=1}^{N} (M_s - \bar{M})^2 \right) + N(\bar{M} - \mu)^2 \right] \right) \\ &\propto \tau^{-N} \exp\left(-\frac{(N-1)S^2}{2\tau^2} \right) \exp\left(-\frac{N}{2\tau^2} (\bar{M} - \mu)^2 \right). \end{aligned} \tag{34}$$

where $S^2 = \frac{1}{N-1}\sum_{s=1}^{N}(M_s - \bar{M})^2$ is the sample variance of the current values of the absolute magnitudes $\{M_s\}$. We can decompose this joint density into the product of a conditional and a marginal:

$$P(\mu, \tau^2 | \{M_s\}) = P(\mu | \tau^2, \{M_s\}) \times P(\tau^2 | \{M_s\}). \tag{35}$$

A joint draw of μ, τ^2 from the joint can be accomplished by first drawing τ^2 from the marginal and then using that value to draw $\mu | \tau^2$ from the conditional. The conditional can be read off from the joint:

$$\begin{aligned} P(\mu | \tau^2, \{M_s\}) &\propto \exp\left(-\frac{N}{2\tau^2}(\bar{M} - \mu)^2\right) \\ &= N(\mu | \bar{M}, \tau^2/N) \end{aligned} \tag{36}$$

and the marginal can be found by marginalisation:

$$\begin{aligned} P(\tau^2 | \{M_s\}) &= \int P(\mu, \tau^2 | \{M_s\})\, d\mu \\ &\propto \tau^{-N} e^{-(N-2)S^2/2\tau^2} \int e^{-N(\bar{M}-\mu)^2/2\tau^2}\, d\mu \\ &\propto \tau^{-N} e^{-(N-2)S^2/2\tau^2} \frac{\tau}{\sqrt{N}} \\ &\propto (\tau^2)^{-(N-1)/2}\, e^{-(N-1)S^2/2\tau^2} \\ &= \textbf{Inv-Gamma}(\tau^2 | a = (N-3)/2, b = (N-1)S^2/2) \end{aligned} \tag{37}$$

Therefore we can generate a joint draw of $\mu, \tau^2 | \{M_s\}$ by drawing τ^2 from this inverse gamma distribution, and then drawing $\mu | \tau^2, \{M_s\}$ from the Gaussian.

(b) Next for $i = 1, \ldots, N$, we draw $M_i \sim P(M_i | \mu, \tau^2, \{M_s \setminus M_i\}, \mathcal{D})$. We note from the the conditional independence of the graph, or from examining the joint posterior that, conditional on μ, τ^2, the probability density of M_i is independent of any $M_{s \neq i}$ and any $D_{s \neq i}$.

$$\begin{aligned} P(M_i | \mu, \tau^2, \{M_s \setminus M_i\}, \mathcal{D}) &= P(M_i | \mu, \tau^2, D_s) \\ &\propto N(D_i | M_i, \sigma_i^2) \times N(M_i | \mu, \tau^2) \\ &= N(M_i | \tilde{\mu}_i, \tilde{\sigma}_i^2) \end{aligned} \tag{38}$$

where we recognise that the product of two Gaussian densities is also a Gaussian density with a mean equal to the precision-weighted mean, and a precision (inverse variance) equal to the sum of the precisions.

$$\tilde{\mu}_i = \frac{\sigma_i^{-2} D_i + \tau^{-2}\mu}{\sigma_i^{-2} + \tau^{-2}} \tag{39}$$

$$\tilde{\sigma}_i^{-2} = \tau^{-2} + \sigma_i^{-2} \tag{40}$$

Thus we can generate a draw for each M_i from these Gaussian densities.

6. Briefly describe how you would implement and run the Gibbs sampler, diagnose convergence, and analyse the resulting output.

Solution: We run a Gibbs sampler in the parameter space of $\boldsymbol{\theta} = (M_1, \ldots, M_N, \mu, \tau^2)$ to generate draws from the posterior, Eq. 32. First, we initialise the chain by choosing a starting point $\boldsymbol{\theta}_0$. We can randomise the starting values by choosing the absolute magnitudes near the data, $M_s \sim N(D_s, \sigma_s^2)$, and choosing $\mu = \bar{M}$ and $\tau^2 = S^2$. Then we run the chain by alternating draws of μ, τ^2 from Step (a) above and then M_i from Step (b) above for all $i = 1, \ldots, N$. Then we record the current state of the chain $\boldsymbol{\theta}_i = (M_1, \ldots, M_N, \mu, \tau^2)$. We run the sampler for a larger number M of cycles. We run 4 independent chain from different randomised starting points $\boldsymbol{\theta}_0$. To asses convergence, we compute Gelman-Rubin ratios to compare the variance between the chains to the variance within the chains to see if they are mixing and converging to the same distribution. We aim for a G-R ratio < 1.05. We may judge the amount of "burn-in" samples to be removed from the beginning of the chain by requiring the G-R ratio of the remaining chains to be satisfactory. We also assess the autocorrelation time scale of the chain by computing the autocorrelations of the chain in each parameter. We can choose a "thinning" factor by finding the maximum lag n among the parameters after which the chain values are approximately uncorrelated. We then removed the burn-in and thin the chain by retaining every nth step in the chain. We can concatenate the multiple chains now to compute histograms, scatter and contour plots to represent the posterior inferences of the parameters of interest, as well as numerical posterior summaries, such as posterior mean and standard deviations of the parameters.

7. (*non-exam*) Implement your Gibbs sampler in code and apply it to analyse the data from Example Sheet 1, Problem 1.

   Solution: see code.

5 PGM and Gibbs Sampling for the Hierarchical Linear Regression Model:

Consider the problem of linear regression of the quasar X-ray spectral index vs. bolometric luminosity in the presence of measurement error in both quantities and intrinsic dispersion. (Regression is also described in Feigelson & Babu, Chapter 7, Ivezic et al., Chapter 8, and Kelly et al. 2007, The Astrophysical Journal, 665, 1506). Consider the probabilistic generative model described in class:

$$\xi_i | \, \mu, \tau^2 \sim N(\mu, \tau^2) \tag{41}$$

$$\eta_i | \, \xi_i; \alpha, \beta, \sigma^2 \sim N(\alpha + \beta \xi_i, \sigma^2) \tag{42}$$

$$x_i | \xi_i \sim N(\xi_i, \sigma_{x,i}^2) \tag{43}$$

$$y_i | \eta_i \sim N(\eta_i, \sigma_{y,i}^2) \tag{44}$$

The astronomer observes values $\mathcal{D} = \{x_i, y_i\}$ with heteroskedastic measurement error of known variances $\{\sigma_{x,i}^2, \sigma_{y,i}^2\}$, for $i = 1, \ldots, N$ independent quasars.

1. Write down the joint distribution $P(x_i, y_i, \xi_i, \eta_i | \, \alpha, \beta, \sigma^2, \mu, \tau^2)$ for a single quasar.

Solution:

$$\begin{aligned} P(x_i, y_i, \xi_i, \eta_i |\, \alpha, \beta, \sigma^2, \mu, \tau^2) &= P(x_i, y_i |\, \xi_i, \eta_i) P(\xi_i, \eta_i |\, \alpha, \beta, \sigma^2, \mu, \tau^2) \\ &= P(x_i |\, \xi_i) P(y_i |\, \eta_i) P(\eta_i |\, \xi_i; \alpha, \beta, \sigma^2) P(\xi_i |\, \mu, \tau^2) \\ &= N(x_i |\, \xi_i, \sigma_{x,i}^2) N(y_i |\, \eta_i, \sigma_{y,i}^2) N(\eta_i |\, \alpha + \beta\xi_i, \sigma^2) N(\xi_i | \mu, \tau^2) \end{aligned} \tag{45}$$

2. Adopt "non-informative" hyperpriors on the hyperparameters $P(\alpha, \beta, \sigma^2) \propto 1, \sigma^2 > 0$ and $P(\mu, \tau^2) \propto 1, \tau^2 > 0$. Write down the full joint distribution of all data $\mathcal{D}$, latent variables $\{\xi_i, \eta_i\}$, and hyperparamters $\alpha, \beta, \sigma^2, \mu, \tau^2$.

   Solution: We assume independent, improper flat priors on α, β, μ: $P(\alpha) \propto 1$, $P(\beta) \propto 1$, $P(\mu) \propto 1$. We assume flat positive priors on σ^2, τ^2: $P(\sigma^2) \propto 1, \sigma^2 > 0$, $P(\tau^2) \propto 1, \tau^2 > 0$. Then the joint distribution:

$$\begin{aligned} P(\{x_i, y_i\}, \{\xi_i, \eta_i\}, \alpha, \beta, \sigma^2, \mu, \tau^2 = \Big[\prod_{i=1}^{N} & N(x_i |\, \xi_i, \sigma_{x,i}^2) N(y_i, |\eta_i, \sigma_{y,i}^2) \\ & N(\eta_i |\, \alpha + \beta\xi_i, \sigma^2) N(\xi_i |\, \mu, \tau^2) \Big] \\ & \times P(\alpha) P(\beta) P(\sigma^2) P(\mu) P(\tau^2) \end{aligned} \tag{46}$$

   On $\tau^2, \sigma^2 > 0$, this is:

$$\begin{aligned} P(\{x_i, y_i\}, \{\xi_i, \eta_i\}, \alpha, \beta, \sigma^2, \mu, \tau^2 \propto \Big[\prod_{i=1}^{N} & N(x_i |\, \xi_i, \sigma_{x,i}^2) N(y_i, |\eta_i, \sigma_{y,i}^2) \\ & N(\eta_i |\, \alpha + \beta\xi_i, \sigma^2) N(\xi_i |\, \mu, \tau^2) \Big] \end{aligned} \tag{47}$$

   and zero otherwise.

3. Draw a probabilistic graphical model / directed acyclic graph to represent this joint distribution. **Solution: See Fig. 4**

4. Construct a Gibbs sampler for the posterior $P(\{\xi_i, \eta_i\}, \alpha, \beta, \sigma^2, \mu, \tau^2 |\, \mathcal{D})$ by deriving a complete set of conditional posterior densities. You may choose to draw "blocks" or parameter subsets from their conditionals in a single step, rather than drawing a single parameter in each step. You may assume that you have access to algorithms that generate random draws from Gaussian distributions any mean and (positive) variance, and inverse-gamma distributions with shape parameter $a > 0$ and scale parameter $b > 0$:

$$\text{Inv-Gamma}(x |\, a, b) \propto x^{-(a+1)} \exp(-b/x), x > 0. \tag{48}$$

   Solution: The full posterior (for $\tau^2, \sigma^2 > 0$) is given by:

$$\begin{aligned} P(\{\xi_i, \eta_i\}, \alpha, \beta, \sigma^2, \mu, \tau^2 |\, \{x_i, y_i\}) \propto \Big[\prod_{i=1}^{N} & N(x_i |\, \xi_i, \sigma_{x,i}^2) N(y_i, |\eta_i, \sigma_{y,i}^2) \\ & N(\eta_i |\, \alpha + \beta\xi_i, \sigma^2) N(\xi_i |\, \mu, \tau^2) \Big] \end{aligned} \tag{49}$$

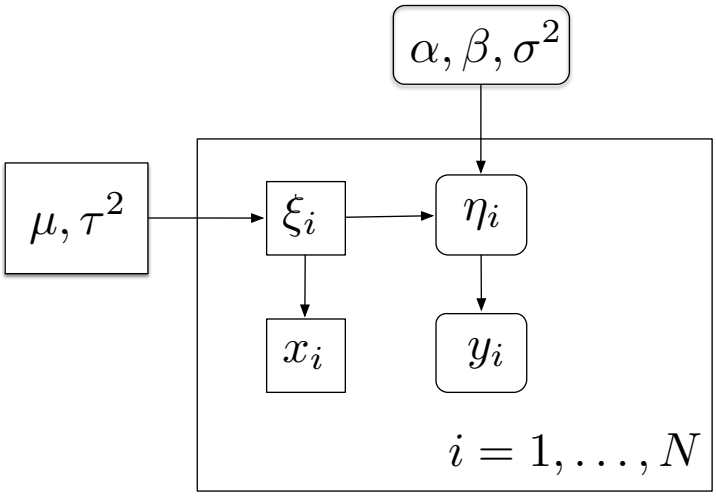

Figure 4: Probabilistic Graphical Model for hierarchical linear regression.

A Gibbs sampler is constructed by cycling through and drawing from a full set of conditional posteriors of subsets of parameters given the remaining parameters and data $\mathcal{D} = \{x_i, y_i\}$, such that each parameter is sampled and updated at least once in each cycle. We will derive one valid Gibbs sampler, but other constructions are possible that will differ in computational speed, convergence to the posterior, and difficulty in derivation. Our Gibbs sampler proceeds by sampling from $N+4$ conditional posteriors: $P(\xi_i, \eta_i | \ldots, \mathcal{D})$, $P(\sigma^2 | \ldots, \mathcal{D})$, $P(\alpha | \ldots, \mathcal{D})$, $P(\beta | \ldots, \mathcal{D})$, $P(\mu, \tau^2 | \ldots, \mathcal{D})$, where $\ldots$ indicates all other parameters in the full posterior that are not explicitly denoted. These conditional posteriors are derived from the full posterior above, by recognising the dependence on the parameters being updated in each step.

(a) For each individual quasar $i = 1, \ldots, N$, we update/sample the latent variables ξ_i, η_i from

$$\begin{aligned} P(\xi_i, \eta_i | \alpha, \beta, \sigma^2, \mu, \tau^2) &\propto N(x_i | \xi_i, \sigma_{x,i}^2) N(y_i, | \eta_i, \sigma_{y,i}^2) N(\eta_i | \alpha + \beta \xi_i, \sigma^2) N(\xi_i | \mu, \tau^2) \\ &\propto N(\boldsymbol{\phi}_i | \boldsymbol{\gamma}, \boldsymbol{C}) \times N(\boldsymbol{\phi}_i | \boldsymbol{z}_i, \boldsymbol{\Sigma}_i) \end{aligned} \tag{50}$$

where the second line is obtained using the properties of the multivariate Gaussian and $\boldsymbol{\phi}_i \equiv (\eta_i, \xi_i)^T$, $\boldsymbol{\gamma} \equiv (\alpha + \beta\mu, \mu)^T$ and

$$\boldsymbol{C} \equiv \begin{pmatrix} \beta^2\tau^2 + \sigma^2 & \beta\tau^2 \\ \beta\tau^2 & \tau^2 \end{pmatrix} \tag{51}$$

$$\boldsymbol{\Sigma}_i \equiv \begin{pmatrix} \sigma_{y,i}^2 & 0 \\ 0 & \sigma_{x,i}^2 \end{pmatrix} \tag{52}$$

The product of two Gaussian densities is proportional to a single Gaussian density, whose mean is the precision-weighted average of the individual means, and whose precision (matrix) is the sum of the individual precision

(matrices). Precision is the inverse of the variance, and a precision matrix is the inverse of a covariance matrix. Therefore let

$$\bar{\boldsymbol{\phi}}_i \equiv (\boldsymbol{C}^{-1} + \boldsymbol{\Sigma}^{-1})^{-1}(\boldsymbol{C}^{-1}\boldsymbol{\gamma} + \boldsymbol{\Sigma}_i^{-1}\boldsymbol{z}_i) \tag{53}$$

$$\boldsymbol{V}_\phi^i = (\boldsymbol{C}^{-1} + \boldsymbol{\Sigma}_i^{-1})^{-1} \tag{54}$$

Thus $P(\boldsymbol{\phi}|\ldots,\mathcal{D}) = N(\boldsymbol{\phi}|\,\bar{\boldsymbol{\phi}}, \boldsymbol{V}_\phi^i)$. Therefore we make the following draw

$$\begin{pmatrix}\eta_i \\ \xi_i\end{pmatrix} \sim N(\bar{\boldsymbol{\phi}}_i, \boldsymbol{V}_\phi^i) \tag{55}$$

for each quasar i.

(b) Next we update σ^2 from $P(\sigma^2|\ldots\mathcal{D})$.

$$\begin{aligned} P(\sigma^2|\ldots\mathcal{D}) &\propto \prod_{i=1}^N N(\eta_i|\,\alpha+\beta\xi_i,\sigma^2) \\ &\propto \sigma^{-N}\prod_{i=1}^N \exp\left(-\frac{1}{2}(\eta_i-\alpha-\beta\xi_i)^2/\sigma^2\right) \\ &\propto (\sigma^2)^{-N/2}\exp\left(-\frac{1}{2}\mathbf{SSR}/\sigma^2\right) \\ &= \textbf{Inv-Gamma}(\sigma^2|\,a=(N-2)/2, b=\mathbf{SSR}/2) \end{aligned} \tag{56}$$

where $\mathbf{SSR} \equiv \sum_{i=1}^N(\eta_i-\alpha-\beta\xi_i)^2$. Thus, we draw a new σ^2 from the above inverse gamma distribution.

(c) Next we update α from $P(\alpha|\ldots\mathcal{D})$.

$$\begin{aligned} P(\alpha|\ldots\mathcal{D}) &\propto \prod_{i=1}^N N(\eta_i|\,\alpha+\beta\xi,\sigma^2) \\ &\propto \prod_{i=1}^N \exp\left(-\frac{1}{2}(\eta_i-\alpha-\beta\xi_i)^2/\sigma^2\right) \\ &\propto \prod_{i=1}^N \exp\left(-\frac{1}{2}(\alpha+\beta\xi_i-\eta_i)^2/\sigma^2\right) \\ &\propto \prod_{i=1}^N N(\alpha|\,\eta_i-\beta\xi_i,\sigma^2). \end{aligned} \tag{57}$$

Now we use the fact that the product of N Gaussian densities (in α) is proportional to a single Gaussian density (in α). The resulting mean is the precision-weighted average of the individual means, and the resulting precision (inverse variance) is the sum of the individual precisions. In the case, all the precision weights are equal (to σ^{-2}).

$$P(\alpha|\ldots\mathcal{D}) = N\left(\alpha\middle|\, N^{-1}\sum_{i=1}^N(\eta_i-\beta\xi_i), \sigma^2/N\right) \tag{58}$$

Thus, we draw a new α from this Gaussian distribution.

(d) Next we update β from $P(\beta|\ldots\mathcal{D})$.

$$\begin{aligned} P(\beta|\ldots\mathcal{D}) &\propto \prod_{i=1}^{N} N(\eta_i|\,\alpha+\beta\xi_i,\sigma^2) \\ &\propto \prod_{i=1}^{N} \exp\left(-\frac{1}{2}(\eta_i-\alpha-\beta\xi_i)^2/\sigma^2\right) \\ &\propto \prod_{i=1}^{N} \exp\left(-\frac{1}{2}(\beta+\alpha/\xi_i-\eta_i/\xi_i)^2/(\sigma/\xi_i)^2\right) \\ &\propto \prod_{i=1}^{N} N(\beta|\,(\eta_i-\alpha)/\xi_i,\tau_i^2\equiv\sigma^2/\xi_i^2) \end{aligned} \tag{59}$$

We can apply the "product-of-Gaussians" rule again to derive that this conditional posterior density of β is $P(\beta|\ldots\mathcal{D}) = N(\beta|\,\bar{\beta},V_\beta)$ where

$$\bar{\beta} = \frac{\sum_{i=1}^{N}\tau_i^{-2}(\eta_i-\alpha)/\xi_i}{\sum_{i=1}^{N}\tau_i^{-2}} \tag{60}$$

$$V_\beta = \left(\sum_{i=1}^{N}\tau_i^{-2}\right)^{-1} \tag{61}$$

Thus, we draw a new β from the above Gaussian density.

(e) Finally, we update μ,τ^2 from $P(\mu,\tau^2|\,\ldots,\mathcal{D}) = P(\mu,\tau^2|\,\{\xi_i\})$. This is essentially the posterior of the mean and variance of a Gaussian given draws from that Gaussian. We can use the same approach as in Problem 4.5 above in this example sheet.

$$P(\mu,\tau^2|\,\{\xi_i\}) \propto \prod_{i=1}^{N} N(\xi_i|\,\mu,\tau^2) \tag{62}$$

Using the same derivation, we can factor

$$\begin{aligned} P(\mu,\tau^2|\,\{\xi_i\} &= P(\mu|\,\tau^2,\{\xi_i\}) \times P(\tau^2|\,\{\xi_i\}) \\ &= N(\mu|\,\bar{\xi},\tau^2/N) \times \textbf{Inv-Gamma}(\tau^2|\,a=(N-3)/2, b=(N-1)S_\xi^2/2) \end{aligned} \tag{63}$$

where $\bar{\xi} = N^{-1}\sum_{i=1}^{N}\xi_i$

$$S_\xi^2 = \frac{1}{N-1}\sum_{i=1}^{N}(\xi_i-\bar{\xi})^2. \tag{64}$$

Hence we generate a joint draw of μ,τ^2 from their conditional by first drawing τ^2 from the inverse-gamma and then $\mu|\,\tau^2,\bar{\xi}$ from the Gaussian.

5. Briefly describe how you would implement and run the Gibbs sampler, diagnose convergence, and analyse the resulting output.

Solution: First we would initialise each chain with randomised, but reasonable starting values. We can initialise the hyperparameters by finding the maximum likelihood of $L(\alpha,\beta,\sigma^2,\mu,\tau^2)$ derived in Example Sheet 2, Problem 1.2,

and computing the inverse of the observed Fisher information matrix. Then we can disperse initial starting points by generating values from a Gaussian with mean at the MLE, and with a covariance given by this inverse. Then in steps (a), we will update the $\{\xi_i, \eta_i\}$ values for all the N quasars. Then we update the hyperparameters in turn via steps (b)-(e). Since in each step, the proposal/draw is always accepted, the chain always moves in the full parameters space. After a full cycle of $N+4$ substeps, we repeat the cycles for a large number of M Gibbs cycles. To diagnose convergence, we run $4-8$ independent chains from different initial values. We can assess the convergence and mixing of the chains by comparing within-chain variance to between-chain variance using the Gelman-Rubin ratio. A ratio G-R $\lesssim 1.1$ indicates good mixing, and helps us ascertain the initial "burn-in" of each of the chains that needs to be discarded (the chains after removing the burn-in should have a G-R close to 1). We plot the sample autocorrelation function of the chains in each parameter-dimension to determine the thinning. We find the slowest-mixing parameter by finding the slowest-decaying autocorrelation function. We find the lag (number of Gibbs cycles) after which the samples in this parameter are close to uncorrelated, and choose this as the thinning factor t. We then thin the chains in all the parameters by only keeping every tth sample (where a sample is a $2N+5$ dimensional vector). We combine all the chains for analysis to, for example, compute sample posterior expectations and variances in each parameter, or compute histograms or 2D density plots of the joint posterior samples.

6. (*non-exam*) Implement your Gibbs sampler in code and apply it to analyse the data from Example Sheet 2, Problem 1.

6 Markov Chain Monte Carlo

Consider a general Bayesian inference of unknown parameters $\boldsymbol{\theta}$ from data $\boldsymbol{D}$. You have a likelihood function $L(\boldsymbol{\theta}) = P(\boldsymbol{D}|\,\boldsymbol{\theta})$ and a proper prior $P(\boldsymbol{\theta})$.

1. Construct an MCMC algorithm that samples from the posterior density $P(\boldsymbol{\theta}|\,\boldsymbol{D}) \propto L(\boldsymbol{\theta})P(\boldsymbol{\theta})$ in the long-run.

   Solution: Since we have no more specific information about the structure of the likelihood or prior, we can use a simple Metropolis algorithm to sample from the posterior $P(\boldsymbol{\theta}|\,\mathcal{D})$.

   (a) First, we initialise the chain to values $\boldsymbol{\theta}_0$.

   (b) We choose a proposal (or jumping) probability density $J(\boldsymbol{\theta}*|\,\boldsymbol{\theta})$, to propose the next position given the current position. For Metropolis, it should be symmetric between the current and proposed position $J(\boldsymbol{\theta}*|\,\boldsymbol{\theta}) = J(\boldsymbol{\theta}|\,\boldsymbol{\theta}*)$. A common choice is a multivariate Gaussian $J(\boldsymbol{\theta}*|\,\boldsymbol{\theta}) = N(\boldsymbol{\theta}^*|\,\boldsymbol{\theta}, \boldsymbol{\Sigma}_J)$ for some covariance matrix $\boldsymbol{\Sigma}_J$.

   (c) At step $t \geq 1$, we propose $\boldsymbol{\theta}^* \sim N(\boldsymbol{\theta}^*|\,\boldsymbol{\theta}_{t-1}, \boldsymbol{\Sigma}_J)$, and calculate the Metropolis ratio:

   $$r = P(\boldsymbol{\theta}*|\,\mathcal{D})/P(\boldsymbol{\theta}|\,\mathcal{D}) \tag{65}$$

   We accept the proposal with probability r: that is, after drawing $u \sim U(0,1)$, we move $\boldsymbol{\theta}_t = \boldsymbol{\theta}^*$ if $u < r$, otherwise we stay in the same position:

$\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1}$. We iterate this step many times until we reach some measure of convergence.

2. Show that the posterior $P(\boldsymbol{\theta}|\, \boldsymbol{D})$ is the stationary distribution of the chain. That is, if, at step $i-1$, $\boldsymbol{\theta}_{i-1} \sim P(\boldsymbol{\theta}|\, \boldsymbol{D})$ is drawn from the posterior density, then so will be $\boldsymbol{\theta}_i$ after a complete iteration of the chain.

Solution: Let us denote the posterior density as $\pi(\boldsymbol{\theta}) = P(\boldsymbol{\theta}|\, \mathcal{D})$. Suppose $\boldsymbol{\theta}^a$ and $\boldsymbol{\theta}^b$ are two points within the support of the posterior, and $\boldsymbol{\theta}^a \sim \pi(\boldsymbol{\theta}^a)$. Suppose the chain transitions at step t from $\boldsymbol{\theta}_{t-1} = \boldsymbol{\theta}^a$ to $\boldsymbol{\theta}_t = \boldsymbol{\theta}^b$ with transition density $T_t(\boldsymbol{\theta}^a \to \boldsymbol{\theta}^b)$ which is a conditional probability density $p_t(\boldsymbol{\theta}^b|\, \boldsymbol{\theta}^a)$. The probability density of the position at time t is

$$\begin{aligned} P(\boldsymbol{\theta}_t) &= \int P(\boldsymbol{\theta}_{t-1})\, T_t(\boldsymbol{\theta}^a \to \boldsymbol{\theta}^b) \\ &= \int \pi(\boldsymbol{\theta}^a)\, T_t(\boldsymbol{\theta}^a \to \boldsymbol{\theta}^b) \end{aligned} \tag{66}$$

The chain is stationary if after this transition, $P(\boldsymbol{\theta}_t) = P(\boldsymbol{\theta}^b) = \pi(\boldsymbol{\theta}^b)$. This can be shown if the chain respects detailed balance (DB): i.e. the probability of starting at $\boldsymbol{\theta}^a \sim \pi(\boldsymbol{\theta})$ and moving to $\boldsymbol{\theta}^b$ is the same as the probability of starting at $\boldsymbol{\theta}^b \sim \pi(\boldsymbol{\theta})$ and moving to $\boldsymbol{\theta}^a$:

$$\pi(\boldsymbol{\theta}^a)\, T_t(\boldsymbol{\theta}^a \to \boldsymbol{\theta}^b) = \pi(\boldsymbol{\theta}^b)\, T_t(\boldsymbol{\theta}^b \to \boldsymbol{\theta}^a) \tag{67}$$

If DB holds then:

$$\begin{aligned} P(\boldsymbol{\theta}_t) &= \int \pi(\boldsymbol{\theta}^a)\, T_t(\boldsymbol{\theta}^a \to \boldsymbol{\theta}^b) \\ &= \int \pi(\boldsymbol{\theta}^b)\, T_t(\boldsymbol{\theta}^b \to \boldsymbol{\theta}^a) \\ &= \pi(\boldsymbol{\theta}^b) \int p_t(\boldsymbol{\theta}^a|\, \boldsymbol{\theta}^b) = \pi(\boldsymbol{\theta}^b) \end{aligned} \tag{68}$$

Therefore if $\boldsymbol{\theta}_{t-1} = \boldsymbol{\theta}^a$ is a draw from the posterior density $\pi(\cdot)$, then so is $\boldsymbol{\theta}_t = \boldsymbol{\theta}^b$ after one transition. To demonstrate that Metropolis respects DB, let us assume, without loss of generality, that points $\boldsymbol{\theta}^a, \boldsymbol{\theta}^b$ are such that $\pi(\boldsymbol{\theta}^b) > \pi(\boldsymbol{\theta}^a)$. Because the transition probability is the proposal/jumping probability (J) times the probability of acceptance ($\min[r, 1]$) , then we have:

$$\pi(\boldsymbol{\theta}^a)\, T_t(\boldsymbol{\theta}^a \to \boldsymbol{\theta}^b) = \pi(\boldsymbol{\theta}^a) J(\boldsymbol{\theta}^b|\, \boldsymbol{\theta}^a) \tag{69}$$

and

$$\begin{aligned} \pi(\boldsymbol{\theta}^b)\, T_t(\boldsymbol{\theta}^b \to \boldsymbol{\theta}^a) &= \pi(\boldsymbol{\theta}^b) J(\boldsymbol{\theta}^a|\, \boldsymbol{\theta}^b) \frac{\pi(\boldsymbol{\theta}^a)}{\pi(\boldsymbol{\theta}^b)} \\ &= \pi(\boldsymbol{\theta}^a) J(\boldsymbol{\theta}^b|\, \boldsymbol{\theta}^a) \end{aligned} \tag{70}$$

due to the symmetry of the proposal distribution. Therefore the Metropolis algorithm respects detailed balance, and $\pi(\cdot)$ is the stationary distribution of the chain.

3. Describe how you would implement this algorithm, diagnose convergence, and prepare the output for analysis.

   Solution: First we might find the mode of the posterior using an optimisation algorithm to find the region of high probability density. We can take the inverse of the Fisher matrix at the mode $\boldsymbol{\theta}_{\text{mode}}$ to approximate the covariance matrix $\boldsymbol{\Sigma}$ of the posterior density. We then initialise a chain by choosing values from a Gaussian centred at the mode and with this covariance. We can then run the Metropolis algorithm as described above. We can tune the covariance matrix scale of the proposal/jumping kernel by $\boldsymbol{\Sigma}_J = \boldsymbol{\Sigma} \times c^2$, and tune the scalar c such that the acceptance rate (fraction of proposals that are accepted) is about 25-50% (a good rule of thumb if $c \approx 2.4/\sqrt{(d)}$ is the posterior is approximately a d-dimensional multivariate Gaussian probability density in the parameters.)

   To assess convergence, we run multiple independent chains from different initialisations. We can assess the mixing of the chains by comparing the between-chain variance to the within-chain variances using the Gelman-Rubin ratio. We run the chains suitably long enough such that the G-R of the cut chains are close to 1 after discarding an initial burn-in length. Since MCMC generates serially correlated samples, we thin the chains by a thinning factor determined by examining the sample autocorrelation functions of the chains in each parameter. For each parameter, we find the lag (in MCMC steps) after which the samples are close to uncorrelated. Then we take the maximum lag over all the parameters as the thinning factor t, and save only every tth sample (where each same is a vector) in the chain. Finally, we concatenate the thinned chains for analysis, to computer posterior expectations and produce visualisations of the posterior samples.