Example Sheet 2 Solutions
Example Class: Thursday, 27 Feb 2020, 3:30pm, MR13

Part III Astrostatistics

Includes solutions to problems 1, 2 & 3.

1 Warm-Ups

1.1 Product of Gaussian densities

Read the "multivariate_gaussian_notes.pdf" posted on the course website. Prove that the product of m multivariate Gaussian densities in random d-dimensional vector $\boldsymbol{x}$:

$$\prod_{i=1}^{m} N(\boldsymbol{x}|\,\boldsymbol{\mu}_i, \boldsymbol{C}_i) \tag{1}$$

is proportional to a single Gaussian density in $\boldsymbol{x}$. Here the $\{\boldsymbol{\mu}_i, \boldsymbol{C}_i\}$ are m pairs of constant mean d-vectors and $d \times d$ covariance matrices. Find the mean and covariance matrix of the single resulting Gaussian. Simplify for the case of $d = 1$ and $m = 2$.

Solution: First, we derive a vectorial "complete the square" lemma:

$$\begin{aligned}\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} + \boldsymbol{x}^T\boldsymbol{b} + c = (\boldsymbol{x}-\boldsymbol{d})^T\boldsymbol{A}(\boldsymbol{x}-\boldsymbol{d}) + e &= \boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} - \boldsymbol{d}^T\boldsymbol{A}\boldsymbol{x} - \boldsymbol{x}^T\boldsymbol{A}\boldsymbol{d} + \boldsymbol{d}^T\boldsymbol{A}\boldsymbol{d} + e \\ &= \boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} - 2\boldsymbol{d}^T\boldsymbol{A}\boldsymbol{x} + \boldsymbol{d}^T\boldsymbol{A}\boldsymbol{d} + e\end{aligned}$$

Therefore, $\boldsymbol{d} = -\frac{1}{2}\boldsymbol{A}^{-1}\boldsymbol{b}$ and $e = c - \boldsymbol{d}^T\boldsymbol{A}\,\boldsymbol{d} = c - \frac{1}{4}\boldsymbol{b}^T\,\boldsymbol{A}^{-1}\boldsymbol{b}$.

Now, by definition, the density $I(x) \equiv \prod_{i=1}^m N(\boldsymbol{x}|\,\boldsymbol{\mu}_i, \boldsymbol{C}_i)$, is

$$I(\boldsymbol{x}) = \prod_{i=1}^{m} |2\pi\boldsymbol{C}_i|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_i)^T\boldsymbol{C}_i^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_i)\right]$$

$$\begin{aligned}-2\log I(\boldsymbol{x}) &= \textbf{const} + \sum_{i=1}^{m} \boldsymbol{x}^T\boldsymbol{C}_i^{-1}\boldsymbol{x} - \boldsymbol{x}^T\boldsymbol{C}_i^{-1}\boldsymbol{\mu}_i - \boldsymbol{\mu}_i^T\boldsymbol{C}_i^{-1}\boldsymbol{x} + \boldsymbol{\mu}_i^T\boldsymbol{C}_i^{-1}\,\boldsymbol{\mu}_i \\ &= \textbf{const} + \boldsymbol{x}^T\left[\sum_{i=1}^{m}\boldsymbol{C}_i^{-1}\right]\boldsymbol{x} - 2\boldsymbol{x}^T\sum_{i=1}^{m}\boldsymbol{C}_i^{T}\boldsymbol{\mu}_i\end{aligned}$$

Let $\boldsymbol{A} = \sum_{i=1}^m \boldsymbol{C}_i^{-1}$ and $\boldsymbol{b} = -2\sum_{i=1}^m \boldsymbol{C}_i^{-1}\,\boldsymbol{\mu}_i$. We have

$$I(\boldsymbol{x}) \propto \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_x)^T\boldsymbol{C}_x^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_x)\right] \propto N(\boldsymbol{x}|\,\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x)$$

where the resulting precision matrix is the sum of the individual precision matrices,

$$\boldsymbol{\Sigma}_x^{-1} \equiv \sum_{i=1}^{m}\boldsymbol{C}_i^{-1}$$

and the resulting mean is the precision-weighted mean of the individual means,

$$\boldsymbol{\mu}_x \equiv \boldsymbol{\Sigma}_x \sum_{i=1}^{m} \boldsymbol{C}_i^{-1} \boldsymbol{\mu}_i = \left[\sum_{i=1}^{m} \boldsymbol{C}_i^{-1} \right]^{-1} \sum_{i=1}^{m} \boldsymbol{C}_i^{-1} \boldsymbol{\mu}_i$$

In the case of $d = 1$ and $m = 2$, let μ_1, μ_2 and σ_1^2, σ_2^2 be the individual means and variances. The product of two scalar Gaussian densities is proportional to a single Gaussian density.

$$I(x) \propto N(x|\,\mu_x, \sigma_x^2) \tag{2}$$

The resulting precision (inverse variance) is given by the sum of the individual precisions:

$$\sigma_x^{-2} = \sigma_1^{-2} + \sigma_2^{-2}, \tag{3}$$

and the resulting mean is the precision-weighted mean of the individual means:

$$\mu_x = \sigma_x^2 \left[\sigma_1^{-2} \mu_1 + \sigma_2^{-2} \mu_2 \right] = \frac{\sigma_1^{-2} \mu_1 + \sigma_2^{-2} \mu_2}{\sigma_1^{-2} + \sigma_2^{-2}}. \tag{4}$$

1.2 Sum of Gaussian random variables

Suppose x, y, are independent univariate Gaussian random variables. The marginal distributions are given by:

$$x \sim N(\mu_x, \sigma_x^2) \tag{5}$$

$$y \sim N(\mu_y, \sigma_y^2) \tag{6}$$

Derive the probability density of $z = x + y$. (Hint: look up *characteristic function*).

Solution: The characteristic function of the Gaussian random variable x is given by the Fourier transform of its probability density:

$$\phi_x(t) = \mathbb{E}[e^{itx}] = \int_{-\infty}^{\infty} e^{itx} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(x-\mu_x^2)/\sigma_x^2} \, dx. \tag{7}$$

Let's evaluate this by defining $n = (x - \mu)/\sigma$, which is a standard normal random variable.

$$\phi_x(t) = e^{it\mu_x} \int_{-\infty}^{\infty} e^{itn\sigma} \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}n^2) \, dn = e^{it\mu_x} I(t\sigma) \tag{8}$$

Now consider the integral:

$$\begin{aligned} I(k) &\equiv \int_{-\infty}^{\infty} e^{ink} \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}n^2) \, dn \\ &= \int_{-\infty}^{\infty} \left[\cos(nk) + i \sin(nk) \right] \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}n^2) \, dn \\ &= \int_{-\infty}^{\infty} \cos(nk) \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}n^2) \, dn \end{aligned} \tag{9}$$

where the last line follows from the oddness of the sine function. To find $I(k)$ we can take the derivative:

$$\begin{aligned} I'(k) &= \int_{-\infty}^{\infty} -n\sin(kn)\frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}n^2)\,dn \\ &= -\frac{1}{\sqrt{2\pi}}\sin(kn)\exp(-\frac{1}{2}n^2)\Big|_{-\infty}^{\infty} - \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp(-\frac{1}{2}n^2)k\cos(kn)\,dn \\ &= -k\,I(k) \end{aligned} \tag{10}$$

Solving the ordinary differential equation, we find $I(k) = \exp(-\frac{1}{2}k^2)$, where the integration constant is fixed by the normalisation of the Gaussian $I(0) = 1$. Therefore, the characteristic function of a Gaussian random variable x with mean μ_x and variance σ_x^2 is:

$$\phi_x(t) = e^{it\mu_x}\, e^{-t^2\sigma_x^2/2} \tag{11}$$

A similar expression can be written for y. The characteristic function of $z = x + y$ is

$$\begin{aligned} \phi_z(t) &= \mathbb{E}[e^{it(x+y)}] = \mathbb{E}[e^{itx}]\mathbb{E}[e^{ity}] = \phi_x(t)\phi_y(t) \\ &= e^{it(\mu_x+\mu_y)}\, e^{t^2(\sigma_x^2+\sigma_y^2)/2} = e^{it\mu_z} e^{t^2\sigma_z^2/2}. \end{aligned} \tag{12}$$

We recognise this as the characteristic function of a Gaussian random variable with mean $\mu_z = \mu_x + \mu_y$ and variance $\sigma_z^2 = \sigma_x^2 + \sigma_y^2$. Therefore,

$$P(z) = \frac{1}{\sigma_z\sqrt{2\pi}} e^{-\frac{1}{2}(z-\mu_z)^2/\sigma_z^2} \tag{13}$$

(One can perform the inverse Fourier transform of $\phi_z(t)$ to verify this). This result can also be obtained by performing the integral:

$$P(z) = \int P(x)\,P(y = z - x)\,dx = \frac{1}{2\pi}\int e^{-\frac{1}{2}(x-\mu_x)^2/\sigma_x^2}\, e^{-\frac{1}{2}(z-x-\mu_y)^2/\sigma_x^2}\,dx \tag{14}$$

and simplifying. This is saying that the probability density of the value z is the probability of the value x and the probability that $y = z - x$, marginalised over all possible values of x. Note that this takes the form of a convolution of the densities in x and y, so by the convolution theorem, the Fourier transform of their convolution is the product of the individual Fourier transforms, as shown above.

1.3 Bayesian Inference for Gaussian data with unknown mean and variance

Data $\{y_i\}$ are iid from Gaussian distribution with unknown population mean μ and variance σ^2:

$$y_i \overset{iid}{\sim} N(\mu, \sigma^2) \tag{15}$$

for $i = 1, \dots, N$.

1. Derive the likelihood function $P(\boldsymbol{y}|\,\mu, \sigma^2)$, expressed in terms of the sufficient statistics: the sample mean,

$$\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i \tag{16}$$

and sample variance:

$$s^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{y})^2. \tag{17}$$

Solution:

$$\begin{aligned}
P(\boldsymbol{y}|\,\mu,\sigma^2) &= \prod_{i=1}^{N} N(y_i|\,\mu,\sigma^2) = \prod_{i=1}^{N}(2\pi\sigma^2)^{-1/2}\, e^{-(y_i-\mu)^2/2\sigma^2} \\
&= (2\pi\sigma^2)^{-N/2}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i-\mu)^2\right) \\
&= (2\pi\sigma^2)^{-N/2}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i-\bar{y}+\bar{y}-\mu)^2\right) \\
&= (2\pi\sigma^2)^{-N/2}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{N}\left[(y_i-\bar{y})^2 + 2(\bar{y}-\mu)(y_i-\bar{y}) + (\bar{y}-\mu)^2\right]\right) \\
&= (2\pi\sigma^2)^{-N/2}\exp\left(-\frac{1}{2\sigma^2}\left[\left(\sum_{i=1}^{N}(y_i-\bar{y})^2\right) + N(\bar{y}-\mu)^2\right]\right) \\
&= (2\pi\sigma^2)^{-N/2}\exp\left(-\frac{(N-1)\,s^2}{2\sigma^2}\right)\exp\left(-\frac{N}{2\sigma^2}(\bar{y}-\mu)^2\right).
\end{aligned}$$

where $s^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i-\bar{y})^2$ is the sample variance.

2. Adopt a "non-informative" improper prior density $P(\mu,\sigma^2) \propto \sigma^{-2}$ for $\sigma^2 > 0$. Derive the posterior density $P(\mu,\sigma^2|\,\boldsymbol{y})$. Show that:

$$P(\mu|\,\sigma^2,\boldsymbol{y}) = N(\mu|\,\bar{y},\sigma^2/n) \tag{18}$$

and

$$P(\sigma^2|\,\boldsymbol{y}) = \text{Inv-}\chi^2(\sigma^2|n-1,s^2) \tag{19}$$

where the scaled inverse χ^2 distribution has an unnormalised density:

$$\text{Inv-}\chi^2(\theta\,|\,n-1,s^2) \propto \theta^{(-\nu/2+1)}\exp(-\nu s^2/(2\theta)). \tag{20}$$

Solution: For $\sigma^2 > 0$, we have the posterior

$$\begin{aligned}
P(\mu,\sigma^2|\,\boldsymbol{y}) &\propto P(\boldsymbol{y}|\,\mu,\sigma^2)\times\sigma^{-2} \\
&\propto (\sigma^2)^{-N/2-1}\exp\left(-\frac{(N-1)\,s^2}{2\sigma^2}\right)\exp\left(-\frac{N}{2\sigma^2}(\bar{y}-\mu)^2\right).
\end{aligned}$$

By inspection, we see that for any fixed σ^2, the conditional posterior depends on μ only through the factor

$$P(\mu|\,\sigma^2,\boldsymbol{y}) \propto \exp\left(-\frac{N}{2\sigma^2}(\bar{y}-\mu)^2\right).$$

The normalisation constant can be inferred by requiring a Gaussian integral to integrate to unity. Finally,

$$P(\mu|\,\sigma^2,\boldsymbol{y}) = N(\mu|\,\bar{y},\sigma^2/N)$$

The marginal posterior of σ^2 can be computed by integrating out μ from the joint posterior.

$$\begin{aligned} P(\sigma^2|\,\boldsymbol{y}) &\propto (\sigma^2)^{-N/2-1} \exp\left(-\frac{(N-1)\,s^2}{2\sigma^2}\right) \int \exp\left(-\frac{N}{2\sigma^2}(\bar{y}-\mu)^2\right) d\mu \\ &\propto (\sigma^2)^{-N/2-1} \exp\left(-\frac{(N-1)\,s^2}{2\sigma^2}\right) (2\pi\sigma^2/N)^{1/2} \\ &\propto (\sigma^2)^{-N/2-1/2} \exp\left(-\frac{(N-1)\,s^2}{2\sigma^2}\right) \\ &= \textbf{Inv-}\chi^2(\sigma^2\,|\,n-1, s^2) \end{aligned}$$

where the normalisation constant can be inferred to be the same as that normalising the inverse χ^2 distribution by the fact that they expression have the same functional form with respect to $\theta = \sigma^2$.

3. Show that the marginal $P(\mu|\,\boldsymbol{y})$ is a t-distribution and derive its parameters. A t-random variable has unnormalised density:

$$t_\nu(\theta|\,\mu,\sigma^2) \propto \left[1 + \frac{1}{\nu}\left(\frac{\theta-\mu}{\sigma}\right)^2\right]^{-(\nu+1)/2}. \tag{21}$$

Solution: The marginal posterior of μ can be obtained by integration:

$$P(\mu|\,\boldsymbol{y}) \propto \int_0^\infty (\sigma^2)^{-N/2-1} \exp\left(-\frac{(N-1)\,s^2}{2\sigma^2}\right) \exp\left(-\frac{N}{2\sigma^2}(\bar{y}-\mu)^2\right) d\sigma^2.$$

By changing variables to $z = A/2\sigma^2$, where $A = (n-1)s^2 + n(\mu-\bar{y})^2$, we find

$$\begin{aligned} P(\mu|\,\boldsymbol{y}) &\propto \int_0^\infty \left(\frac{A}{2z}\right)^{-n/2-1} \exp(-z)\frac{A}{2z^2}\,dz \\ &\propto A^{-n/2} \int_0^\infty z^{(n-2)/2} \exp(-z)\,dz \end{aligned}$$

The integral is a dimensionless gamma function, which we do not need to compute as it is a constant. Now, we have

$$\begin{aligned} P(\mu|\,\boldsymbol{y}) &\propto A^{-n/2} \\ &\propto \left[(n-1)s^2 + n(\mu-\bar{y})^2\right]^{-n/2} \propto \left[1 + \frac{n(\mu-\bar{y})^2}{(n-1)s^2}\right]^{-n/2} \\ &= t_{n-1}(\mu|\,\bar{y}, s^2/n). \end{aligned}$$

Therefore the marginal is a $t-$distribution with those parameters.

4. Suppose the sufficient statistics of the data are $\bar{y} = 0$ and $s^2 = 1$. Plot the marginal posterior density $P(\mu|\,\boldsymbol{y})$ for $N = 2, 5, 10, 30$, and compare against a Gaussian density with mean $\bar{y}$ and variance s^2/N.

Solution: See plot below. One has to be mindful of the proper normalisations of each probability density. As $N \to \infty$, the Gaussian distribution closely approximates the Student t distribution.

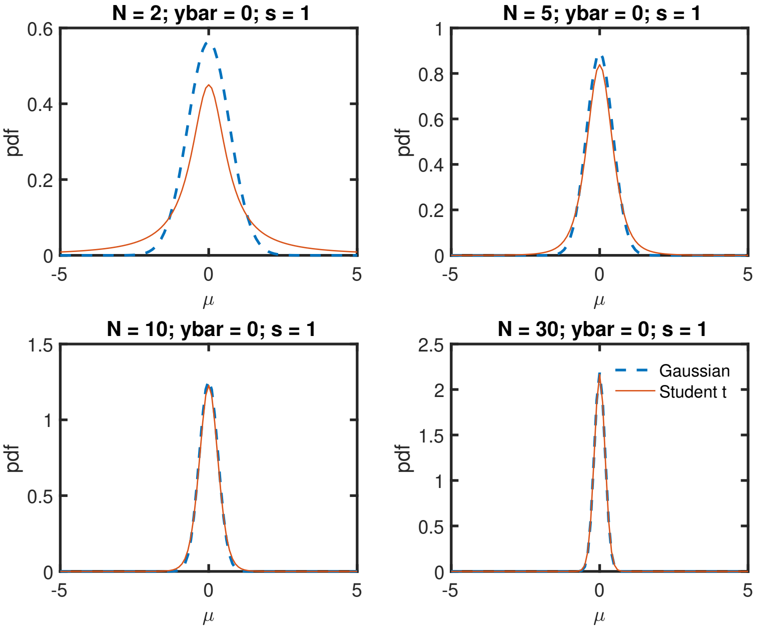

Figure 1: **Comparison between Gaussian and t-distributions.**

2 Linear Regression with heteroskedastic $(x, y)-$measurement error and intrinsic dispersion: Quasar X-ray Spectral Slopes vs. Eddington Ratios

In class we examined 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:

$$\xi_i \sim N(\mu, \tau^2) \tag{22}$$

$$\eta_i|\xi_i \sim N(\alpha + \beta\xi_i, \sigma^2) \tag{23}$$

$$x_i|\xi_i \sim N(\xi_i, \sigma^2_{x,i}) \tag{24}$$

$$y_i|\eta_i \sim N(\eta_i, \sigma^2_{y,i}) \tag{25}$$

The astronomer measures values $\mathcal{D} = \{x_i, y_i\}$ with known measurement error variances $\{\sigma^2_{x,i}, \sigma^2_{y,i}\}$, for $i = 1, \ldots, N$ 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:

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

which comes from expanding the joint into conditionals and marginals and using the modeling assumptions, Eq. 1-4.

2. Derive the observed data likelihood function for all the quasars:

$$L(\alpha, \beta, \sigma^2, \mu, \tau^2) = \prod_{i=1}^{N} P(x_i, y_i |\, \alpha, \beta, \sigma^2, \mu, \tau^2). \tag{26}$$

Show all steps and maximally simplify.

Solution: We could marginalise out the latent coordinates (ξ_i, η_i) analytically:

$$P(x_i, y_i | \alpha, \beta, \sigma^2, \mu, \sigma^2, \tau^2) = \int\int d\eta_i \, d\xi_i P(x_i, y_i, \xi_i, \eta_i |\, \alpha, \beta, \sigma^2, \mu, \tau^2),$$

but that would be tedious and not very insightful. Instead, we can utilise the properties of multivariate Gaussian random vectors. We use the fact that a marginal distribution of multivariate Gaussian vector $\boldsymbol{V}$,

$$\boldsymbol{V} \sim N(\boldsymbol{V}_0, \boldsymbol{\Sigma}_V)$$

and a conditional distribution of $\boldsymbol{U}|\boldsymbol{V}$:

$$\boldsymbol{U}|\boldsymbol{V} \sim N(\boldsymbol{U}_0 + \boldsymbol{X}\boldsymbol{V}, \boldsymbol{\Sigma}_{U|V})$$

for some matrix $\boldsymbol{X}$ of the appropriate dimensionality, are equivalent to a joint distribution that is also multivariate Gaussian

$$\begin{pmatrix} \boldsymbol{U} \\ \boldsymbol{V} \end{pmatrix} \sim N\left(\begin{pmatrix} \boldsymbol{U}_0 + \boldsymbol{X}\boldsymbol{V}_0 \\ \boldsymbol{V}_0 \end{pmatrix}, \begin{pmatrix} \boldsymbol{X}\boldsymbol{\Sigma}_V\boldsymbol{X}^T + \boldsymbol{\Sigma}_{U|V} & \boldsymbol{X}\boldsymbol{\Sigma}_V \\ \boldsymbol{\Sigma}_V\boldsymbol{X}^T & \boldsymbol{\Sigma}_V \end{pmatrix} \right).$$

Then marginally, $\boldsymbol{U} \sim N(\boldsymbol{U}_0 + \boldsymbol{X}\boldsymbol{V}_0, \boldsymbol{X}\boldsymbol{\Sigma}_V\boldsymbol{X}^T + \boldsymbol{\Sigma}_{U|V})$ is also multivariate Gaussian. We can apply this to Eqs. 1 & 2 to get the joint distribution of (η_i, ξ_i),

$$\begin{pmatrix} \eta_i \\ \xi_i \end{pmatrix} \sim N\left(\begin{pmatrix} \alpha + \beta\mu \\ \mu \end{pmatrix}, \begin{pmatrix} \beta^2\tau^2 + \sigma^2 & \beta\tau^2 \\ \beta\tau^2 & \tau^2 \end{pmatrix} \right).$$

We also note that Eqs 3 & 4 can be described jointly as a multivariate Gaussian vector:

$$\begin{pmatrix} y_i \\ x_i \end{pmatrix} \Big| \begin{pmatrix} \eta_i \\ \xi_i \end{pmatrix} \sim N\left(\begin{pmatrix} \eta_i \\ \xi_i \end{pmatrix}, \begin{pmatrix} \sigma^2_{y,i} & 0 \\ 0 & \sigma^2_{x,i} \end{pmatrix} \right)$$

We can apply the same lemma again (with $\boldsymbol{X} = I$) to obtain the marginal density of $\boldsymbol{U} = (y_i, x_i)^T$, integrating out the latent variables $\boldsymbol{V} = (\eta_i, \xi)^T$:

$$\begin{pmatrix} y_i \\ x_i \end{pmatrix} \sim N\left(\begin{pmatrix} \alpha + \beta\mu \\ \mu \end{pmatrix}, \begin{pmatrix} \beta^2\tau^2 + \sigma^2 + \sigma^2_{y,i} & \beta\tau^2 \\ \beta\tau^2 & \tau^2 + \sigma^2_{x,i} \end{pmatrix} \right) \equiv N(\boldsymbol{\zeta}, \boldsymbol{V}_i).$$

Hence the joint sampling distribution for $\boldsymbol{z}_i \equiv (y_i, x_i)^T$ is $P(y_i, x_i | \alpha, \beta, \sigma^2, \mu, \tau^2) = N(\boldsymbol{z}_i |\, \boldsymbol{\zeta}, \boldsymbol{V}_i)$, so $\boldsymbol{z}_i$ is also a multivariate Gaussian vector. The likelihood function of the parameters is then

$$L(\alpha, \beta, \sigma^2, \mu, \tau^2) = \prod_{i=1}^{N} N(\boldsymbol{z}_i |\, \boldsymbol{\zeta}, \boldsymbol{V}_i),$$

assuming the data from each quasar are independently sampled.

3. Write a code to find the maximum likelihood estimate, if given $\{x_i, y_i\}$ and their known measurement variances for $i = 1 \dots N$ quasars. Find an approximate 68% confidence interval for each parameter using the observed Fisher information. (Use a generic optimisation library or toolbox to numerically minimise a given function, e.g. scipy.optimize in Python, fmincon in Matlab, or optim in R, or equivalent).

   Solution: see code. We minimise $-\log L(\alpha, \beta, \sigma^2, \mu, \tau^2)$ and use its Hessian at the minimum as the observed Fisher information matrix.

4. Using the dataset provided online ("quasar_data.txt"), find the maximum likelihood estimates (MLEs) of the parameters $\alpha, \beta, \sigma^2, \mu, \tau^2$, and their uncertainties.

   Solution: see code. We find $\hat{\alpha} = 3.05 \pm 0.52$, $\hat{\beta} = 1.10 \pm 0.62$, $\widehat{\sigma^2} = 0.45 \pm 0.073$, $\hat{\mu} = -0.84 \pm 0.08$ and $\widehat{\tau^2} = 0.55 \pm 0.06$.

5. Suppose the distribution of the latent (true) independent variables $\{\xi_i\}$ is much wider than their individual uncertainties $\sigma_{x,i}$. If $\tau \gg \max(\sigma_{x,i})$, show that Eq. 26 factors

$$L(\alpha, \beta, \sigma^2, \mu, \tau^2) \approx L_1(\alpha, \beta, \sigma^2) \times L_2(\mu, \tau^2) \tag{27}$$

   so that the estimation of the regression parameters $(\alpha, \beta, \sigma^2)$ decouples from the estimation of the latent distribution of the independent variables. Find $L_1(\alpha, \beta, \sigma^2)$ and $L_2(\mu, \tau^2)$. What are the maximum likelihood estimators for μ, τ^2 ?

   Solution: Let $\boldsymbol{\theta} = (\alpha, \beta, \sigma^2)$ be the parameters of the regression model, and $\boldsymbol{\phi} = (\mu, \tau^2)$ be the parameters of the population of the latent independent variable. Noting that the joint distribution of $P(y_i, x_i|\boldsymbol{\theta}, \boldsymbol{\phi})$ is Gaussian, we can decompose it into the product of the conditional $P(y_i|x_i, \boldsymbol{\theta}, \boldsymbol{\phi})$ and the marginal $P(x_i|\boldsymbol{\theta}, \boldsymbol{\phi})$. We can read from the solution to part 2, the marginal

$$P(x_i|\boldsymbol{\theta}, \boldsymbol{\phi}) = N(x_i|\, \mu, \tau^2 + \sigma^2_{x,i}).$$

   From the properties of the multivariate Gaussian, the conditional density derived from the solution to part 2 is:

$$P(y_i|\, x_i; \boldsymbol{\theta}, \boldsymbol{\phi}) = N(y_i|\, \mathbb{E}[y_i|\, x_i; \boldsymbol{\theta}, \boldsymbol{\phi}], \mathbf{Var}[y_i|\, x_i; \boldsymbol{\theta}, \boldsymbol{\phi}])$$

   in which the conditional expectation is:

$$\begin{aligned}\mathbb{E}[y_i|\, x_i; \boldsymbol{\theta}, \boldsymbol{\phi}] &= \alpha + \beta\mu + \frac{\beta\tau^2}{\tau^2 + \sigma^2_{x,i}}(x_i - \mu) \\ &= \alpha + \frac{\beta\tau^2}{\tau^2 + \sigma^2_{x,i}} x_i + \frac{\beta\sigma^2_{x,i}}{\tau^2 + \sigma^2_{x,i}} \mu\end{aligned}$$

$$\mathbf{Var}[y_i|\, x_i; \boldsymbol{\theta}, \boldsymbol{\phi}] = \beta^2\tau^2 + \sigma^2 + \sigma^2_{y,i} - \frac{(\beta\tau^2)^2}{\tau^2 + \sigma^2_{x,i}}$$

   In the limit $\tau \gg \max(\sigma_{x,i})$, $\mathbb{E}[y_i|\, x_i; \boldsymbol{\theta}, \boldsymbol{\phi}] \to \alpha + \beta x_i$,

$$\beta^2\tau^2 - \frac{(\beta\tau^2)^2}{\tau^2 + \sigma^2_{x,i}} \approx \beta^2\tau^2 - \frac{\beta^2\tau^2}{1 + \sigma^2_{x,i}/\tau^2} \approx \beta^2\tau^2[1 - (1 - \sigma^2_{x,i}/\tau^2)] \approx \beta^2\sigma^2_{x,i}$$

Therefore, Var$[y_i|\, x_i; \boldsymbol{\theta}, \boldsymbol{\phi}] \to \sigma^2 + \sigma_{y,i}^2 + \beta^2\sigma_{x,i}^2$. The likelihood function is then approximately

$$L(\alpha, \beta, \sigma^2, \mu, \tau^2) \approx \prod_{i=1}^{N} N(y_i|\, \alpha + \beta x_i, \sigma^2 + \sigma_{y,i}^2 + \beta^2\sigma_{x,i}^2) \times \prod_{i=1}^{N} N(x_i|\, \mu, \tau^2)$$

Note that only the first product depends on the regression parameters α, β, σ^2.

$$L_1(\alpha, \beta, \sigma^2) = \prod_{i=1}^{N} N(y_i|\, \alpha + \beta x_i, \sigma^2 + \sigma_{y,i}^2 + \beta^2\sigma_{x,i}^2)$$

$$L_2(\mu, \tau^2) = \prod_{i=1}^{N} N(x_i|\, \mu, \tau^2)$$

The MLE for μ, τ^2 are obtained via L_2 alone: $\hat{\mu} = \bar{x}$, $\widehat{\tau^2} = N^{-1}\sum_{i=1}^{N}(x_i - \bar{x})^2$.

6. Compare your MLE for β using Eq 26 against what you get using ordinary least squares (OLS), minimum χ^2, FITEXY modified χ^2 methods, and the MLE in the non-informative $\tau \gg \max(\sigma_{x,i})$ limit.

   (a) Ordinary Least Squares minimises the residual sum of squares (RSS) with respect to the parameters:

   $$\text{RSS} = \sum_{i=1}^{n} (y_i - \alpha - \beta\, x_i)^2 \,. \tag{28}$$

   Estimate the uncertainty (variance) of this $\hat{\beta}_{\text{OLS}}$.

   Solution: This can be solved by direct numerical minimisation of the above. However, the solution can be written in closed form:

   $$\textbf{RSS} = (\boldsymbol{y} - \boldsymbol{X}b)^T(\boldsymbol{y} - \boldsymbol{X}b)$$

   where $\boldsymbol{y}$ has elements y_i, $\boldsymbol{X}$ is a matrix with the ith row being $X_i = (1, x_i)$ and parameter vector $b = (\alpha, \beta)^T$. This ordinary least squares problem has a linear algebra solution:

   $$\hat{\boldsymbol{b}} = \arg\min_{b} \textbf{RSS} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$$

   An unbiased estimate of the residual variance is $\widehat{s^2} = \textbf{RSS}(\hat{\boldsymbol{b}})/(N-2)$. The covariance matrix of $\hat{\boldsymbol{b}}$ is Var$[\hat{\boldsymbol{b}}] = \widehat{s^2}(\boldsymbol{X}^T\boldsymbol{X})^{-1}$.

   We find $\hat{\alpha} = 2.63 \pm 0.10$ and $\hat{\beta} = 0.55 \pm 0.089$.

   (b) Minimum χ^2 or Weighted Least Squares minimises the following with respect to the parameters:

   $$\chi^2 = \sum_{i=1}^{N} \frac{(y_i - \alpha - \beta\, x_i)^2}{\sigma_{y,i}^2}. \tag{29}$$

   Estimate the uncertainty (variance) of this $\hat{\beta}$.

   Solution: This can be solved by direct numerical minimisation of the above. However, the solution can be written in closed form:

   $$\chi^2 = (\boldsymbol{y} - \boldsymbol{X}b)^T\boldsymbol{W}^{-1}(\boldsymbol{y} - \boldsymbol{X}b)$$

where $\boldsymbol{y}$ has elements y_i, $\boldsymbol{X}$ is a matrix with the ith row being $X_i = (1, x_i)$ and parameter vector $b = (\alpha, \beta)^T$. The weight matrix is diagonal with elements $W_{ii} = \sigma^2_{y,i}$. This weighted least squares problem has a linear algebra solution:

$$\hat{\boldsymbol{b}} = \arg\min_b \chi^2 = (\boldsymbol{X}^T \boldsymbol{W}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{W}^{-1} \boldsymbol{Y}$$

The variance is Var$[\hat{\boldsymbol{b}}] = (\boldsymbol{X}^T \boldsymbol{W}^{-1} \boldsymbol{X})^{-1}$.

We find $\hat{\alpha} = 2.52 \pm 0.10$ and $\hat{\beta} = 0.51 \pm 0.089$.

(c) The FITEXY methods (Press et al. *Numerical Recipes in C*) minimise an "effective" χ^2 statistic that takes in account x-measurement errors

$$\chi^2_{EXY} = \sum_{i=1}^{N} \frac{(y_i - \alpha - \beta\, x_i)^2}{\sigma^2_{y,i} + \beta^2 \sigma^2_{x,i}}. \tag{30}$$

Solution: We find $\hat{\alpha} = 3.13$ and $\hat{\beta} = 1.33$.

(d) The maximum likelihood solution assuming a non-informative distribution on the $\{\xi_i\}$ is obtained by minimising $-\log L_1(\alpha, \beta, \sigma^2)$, which you derived in part (5) above. Estimate the uncertainty (variance) of this $\hat{\beta}$.

Solution: We find $\hat{\alpha} = 2.63 \pm 0.12$ and $\hat{\beta} = 0.57 \pm 0.007$.

7. State and employ appropriate non-informative priors on the parameters $\alpha, \beta, \sigma^2, \mu, \tau^2$ defined in Eqs. 22 - 26. Construct and implement a MCMC algorithm to sample from the posterior probability density:

$$P(\alpha, \beta, \sigma^2, \mu, \tau^2 | \mathcal{D}) \propto L(\alpha, \beta, \sigma^2, \mu, \tau^2) \times P(\alpha, \beta, \sigma^2, \mu, \tau^2) \tag{31}$$

Run 4 independent chains, initialise appropriately at different starting points, to diagnose convergence using the Gelman-Rubin ratio. Remove "burn-in" and use the combined chains to compute the marginal distributions of the parameters, and compare against the point estimates you obtained with the other methods.

Solution: See code: We implemented a Metropolis algorithm with 5-dimensional proposals. We find $\hat{\alpha} = 3.04 \pm 0.153$ and $\hat{\beta} = 1.087 \pm 0.171$. See Figures 2, 3, 4.

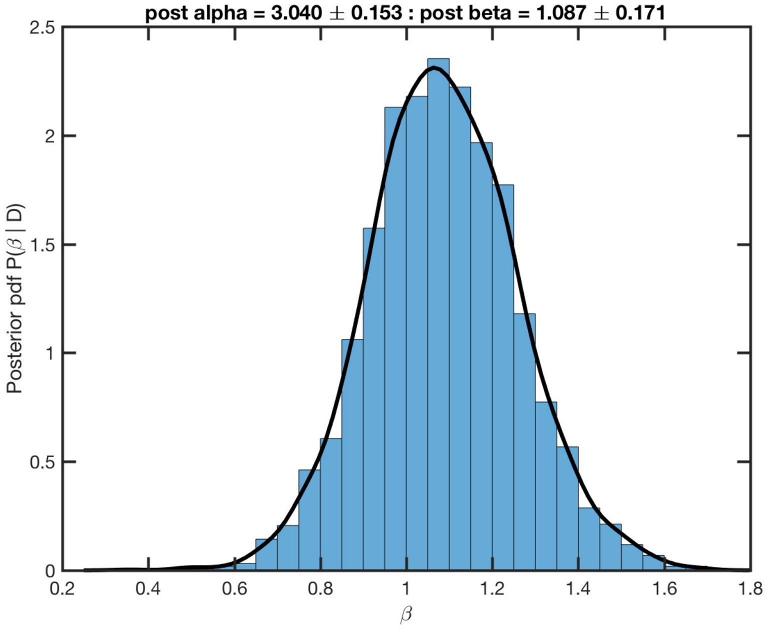

Figure 2: **Marginal Posterior of β**

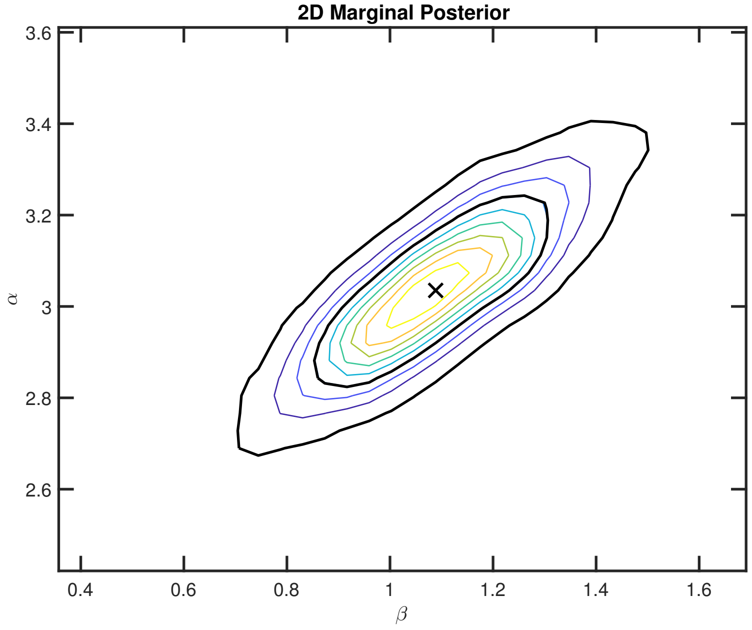

Figure 3: **Joint Posterior of α, β**

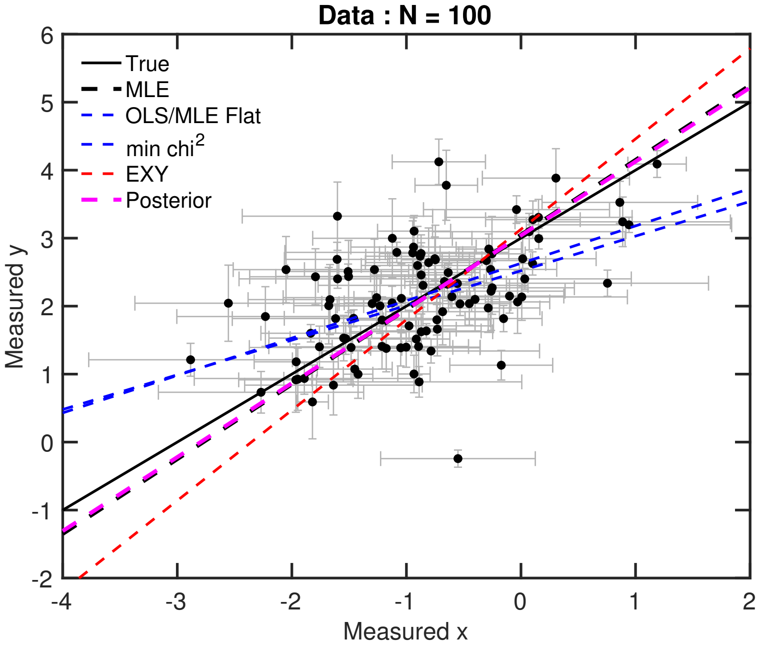

Figure 4: **Comparison of the different estimates β.**

3 Importance Sampling for Bayesian Estimates of the Milky Way Mass using Angular Momentum Measurements

Look up the paper Patel et al. 2017, "Orbits of massive satellite galaxies – II. Bayesian estimates of the Milky Way and Andromeda masses using high-precision astrometry and cosmological simulations." *Monthly Notices of the Royal Astronomical Society*, 468, 3428. Use the measurements in Table 1 and the online data from the Illustris simulation to estimate the Milky Way mass using angular momentum j and the rotational velocity $v_{\max}$ of the Large Magellanic Cloud (LMC). In this context, the Milky Way is the central (host) galaxy of the system, and the LMC is a "satellite" galaxy. We wish to infer the log of the Milky Way mass, $m = \log_{10} M$.

1. Let $\boldsymbol{x} = (v_{\max}, j)$ be the latent parameters, and let $\boldsymbol{d} = (v_{\max}^{\text{obs}}, j^{\text{obs}})$ be their measured values, with uncertainties shown in Table 1. Write down the likelihood function $P(\boldsymbol{d}|\,\boldsymbol{x})$, assuming Gaussian measurement errors.

   Solution:

   $$P(\boldsymbol{d}|\boldsymbol{x}) = N(v_{\mathbf{max}}^{\mathbf{obs}}|\,v_{\mathbf{max}}, \sigma_v^2) \times N(j^{\mathbf{obs}}|\,j, \sigma_j^2)$$

   where σ_v and σ_k are the standard deviations of measurement errors associated with $v_{\mathbf{max}}^{\mathbf{obs}}$ and $j^{\mathbf{obs}}$, listed in Table 1.

2. The Illustris simulation implicitly encodes a joint distribution between these latent dynamical parameters of satellites and the $\log_{10}$ masses of central (or host) galaxies, $P(\boldsymbol{x}, \log_{10} M)$. Assuming this exists, write down an expression for the normalised posterior probability density of the Milky Way $\log_{10}$ mass.

   Solution: The joint posterior density of the latent parameters $\boldsymbol{x}$ and $m \equiv \log_{10} M$ is given by Bayes' Theorem:

   $$P(\boldsymbol{x}, m|\boldsymbol{d}) = \frac{P(\boldsymbol{d}|\boldsymbol{x}, m)P(\boldsymbol{x}, m)}{\int P(\boldsymbol{d}|\boldsymbol{x}, m)P(\boldsymbol{x}, m)\, d\boldsymbol{x}\, dm}$$

   but since $\boldsymbol{d}$ is conditionally independent of m, given $\boldsymbol{x}$, $P(\boldsymbol{d}|\boldsymbol{x}, m) = P(\boldsymbol{d}|\boldsymbol{x})$. Then marginalising over the latent variables $\boldsymbol{x}$, we have

   $$P(m|\boldsymbol{d}) = \frac{\int P(\boldsymbol{d}|\,\boldsymbol{x})P(\boldsymbol{x}, m)\, d\boldsymbol{x}}{\int P(\boldsymbol{d}|\,\boldsymbol{x})P(\boldsymbol{x}, m)\, d\boldsymbol{x}\, dm}$$

3. Write down an expression for the posterior mean estimate of the $\log_{10}$ MW mass in terms of integrals involving the likelihood and prior.

   Solution:

   $$\mathbb{E}[m|\boldsymbol{d}] = \int m\, P(m|\,\boldsymbol{d})\, dm = \frac{\int m\, P(\boldsymbol{d}|\,\boldsymbol{x})P(\boldsymbol{x}, m)\, d\boldsymbol{x}\, dm}{\int P(\boldsymbol{d}|\,\boldsymbol{x})P(\boldsymbol{x}, m)\, d\boldsymbol{x}\, dm}$$

4. Using an arbitrary importance sampling distribution $Q(\boldsymbol{x}, \log_{10} M)$ from which we can easily draw samples, rewrite this expression in terms of expectations with respect to Q.

   Solution: We can multiply the top and bottom integrands by $1 = Q(\boldsymbol{x}, m)/Q(\boldsymbol{x}, m)$.

   $$\mathbb{E}[m|\boldsymbol{d}] = \frac{\int m\, P(\boldsymbol{d}|\,\boldsymbol{x}) \frac{P(\boldsymbol{x},m)}{Q(\boldsymbol{x},m)}\, Q(\boldsymbol{x}, m)\, d\boldsymbol{x}\, dm}{\int P(\boldsymbol{d}|\,\boldsymbol{x}) \frac{P(\boldsymbol{x},m)}{Q(\boldsymbol{x},m)}\, Q(\boldsymbol{x}, m)\, d\boldsymbol{x}\, dm}$$

5. Rewrite this expression now assuming now that the importance sampling distribution is the same as the prior $Q(\boldsymbol{x}, \log_{10} M) = P(\boldsymbol{x}, \log_{10} M)$. Approximate this expression with weighted sums over the prior samples, suitable for the Monte Carlo method, and derive the importance weights.

   Solution: In general, if we have K samples from the importance sampling function, $\boldsymbol{\theta}_i \sim Q(\boldsymbol{\theta})$, then we can approximate expectations with respect to $P(\boldsymbol{\theta})$ using the Monte Carlo sum:

$$\mathbb{E}[f(\boldsymbol{\theta})] = \int f(\boldsymbol{\theta})\, P(\boldsymbol{\theta})\, d\boldsymbol{\theta} = \int f(\boldsymbol{\theta}) \frac{P(\boldsymbol{\theta})}{Q(\boldsymbol{\theta})} Q(\boldsymbol{\theta})\, d\boldsymbol{\theta} \approx \frac{1}{K} \sum_{i=1}^{K} f(\boldsymbol{\theta}_i) w_i$$

   in which the importance weights are $w_i = P(\boldsymbol{\theta}_i)/Q(\boldsymbol{\theta}_i)$ evaluated over the samples $\boldsymbol{\theta}_i$. Applying this we have:

$$\mathbb{E}[m|\boldsymbol{d}] \approx \frac{\frac{1}{K}\sum_{i=1}^{K} m_i\, P(\boldsymbol{d}|\, \boldsymbol{x}_i) \frac{P(\boldsymbol{x}_i, m)}{Q(\boldsymbol{x}_i, m_i)}}{\frac{1}{K}\sum_{i=1}^{K} P(\boldsymbol{d}|\, \boldsymbol{x}_i) \frac{P(\boldsymbol{x}_i, m)}{Q(\boldsymbol{x}_i, m_i)}}$$

   However, since we are using the Illustris simulation both as the prior and importance sampling distribution $P(\boldsymbol{x}, m) = Q(\boldsymbol{x}, m)$, and this simplifies to:

$$\mathbb{E}[m|\boldsymbol{d}] \approx \frac{\sum_{i=1}^{K} m_i\, P(\boldsymbol{d}|\, \boldsymbol{x}_i)}{\sum_{i=1}^{K} P(\boldsymbol{d}|\, \boldsymbol{x}_i)} = \sum_{i=1}^{K} m_i\, w_i$$

   where the importance weights are

$$w_i = \frac{P(\boldsymbol{d}|\, \boldsymbol{x}_i)}{\sum_{i=1}^{K} P(\boldsymbol{d}|\, \boldsymbol{x}_i)}$$

   which are already normalised ($\sum_{i=1}^{K} w_i = 1$).

6. Use the Illustris host-satellite data in the online file "Patel17b_Illustris_Data_KM.txt" as samples from the prior. Use the columns labelled "MVIR", "SATVMAX" and "SATJMAG". Compute the importance weights, and estimate the posterior mean and standard deviation of the $\log_{10}$ MW mass, given the LMC data $\boldsymbol{d}$. Also compute an effective sample size using Eq. B2 in the paper, and compare against the number of samples from the prior.

   Solution: see code. The posterior mean and std deviation of $m = \log_{10} M$ are 12.01 and 0.30. The effective sample size is 2766.

7. Using the bandwidth Eq. B1, create a weighted KDE representation of the posterior distribution $P(\log_{10} M|\boldsymbol{d})$. Plot it over a KDE representation of the marginal prior $P(\log_{10} M)$.

   Solution: see code and Figure 5.

8. Prove that the optimal importance function for approximating the posterior mean of $m = \log_{10} M$, in the sense of minimum variance, is

$$Q^*(m) = \frac{|m|\, P(m|\, \boldsymbol{d})}{\int |m|\, P(m|\, \boldsymbol{d})\, dm} \tag{32}$$

   (You may use Jensen's Inequality: $\mathbb{E}[g(X)] \geq g(\mathbb{E}[X])$ where g is a convex function and X is a generic random variable.) Compare $Q^*(m)$ to your marginal importance function you have plotted, the marginal prior $P(m)$.

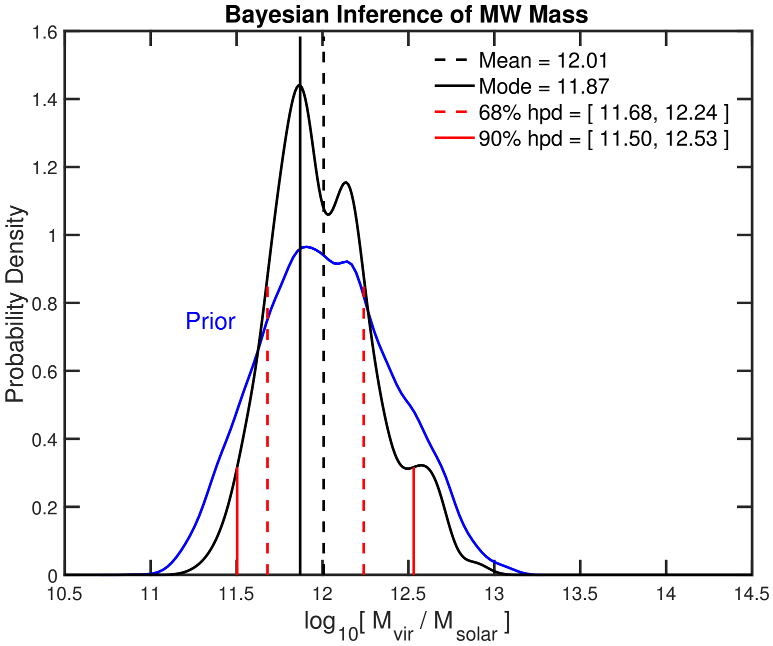

Figure 5: **Importance-weighted KDE of the posterior log mass of the Milky Way Galaxy.**

Solution: We wish to estimate the integral

$$I = \int m\, P(m|\, \boldsymbol{d})\, dm = \int m\, \frac{P(m|\, \boldsymbol{d})}{Q(m)} Q(m)\, dm = \mathbb{E}_Q[Y]$$

where $Y(m) = m\, P(m|\boldsymbol{d})/Q(m)$**. We can construct the estimator**

$$\hat{I} = \frac{1}{N}\sum_{i=1}^{N} Y_i = \frac{1}{N}\sum_{i=1}^{N} \frac{m_i\, P(m_i|\, \boldsymbol{d})}{Q(m_i)}$$

from i.i.d draws $m_i \sim Q(m)$**. The variance of this estimator is** $\mathbf{Var}[\hat{I}] = \frac{1}{N}\mathbf{Var}[Y]$**.**

$$\begin{aligned}
\mathbf{Var}_Q[Y] &= \mathbb{E}_Q[Y^2] - (\mathbb{E}_Q[Y])^2 \\
&= \int Y^2(m)\, Q(m)\, dm - \left(\int Y(m)\, Q(m)\, dm\right)^2 \\
&= \int \frac{m^2 P^2(m|\, \boldsymbol{d})}{Q^2(m)}\, Q(m)\, dm - (m\, P(m|\, \boldsymbol{d})\, dm)^2 \\
&\geq \left(\int \frac{|m|\, P(m|\, \boldsymbol{d})}{Q(m)} Q(m)\, dm\right)^2 - (m\, P(m|\, \boldsymbol{d})\, dm)^2 \\
&\geq \left(\int |m|\, P(m|\, \boldsymbol{d})\, dm\right)^2 - (m\, P(m|\, \boldsymbol{d})\, dm)^2
\end{aligned}$$

which establishes a lower bound on the variance. However, $Q^*(m)$ **achieves this**

bound:

$$\mathbf{Var}_{Q^*}[Y] = (|m|\, P(m|\boldsymbol{d})\, dm)^2 - (m\, P(m|\,\boldsymbol{d})\, dm)^2$$

which proves the claim.

4 Bayesian Inference for the Hubble Constant

Consider the Hubble constant estimation problem from Example Sheet 1. Assume the measurement errors are heteroskedastic (of both the measured magnitudes and the Cepheid distance estimates), and all errors are independent, unless stated otherwise.

1. Write down the likelihood function for the parameters M_0, σ_{int}, and $\theta = 5\log_{10} h$, where $h = H_0/(100 \text{ km s}^{-1} \text{ Mpc}^{-1})$.

2. Adopting flat, improper priors on M_0 and θ, and a positive flat improper prior on $\sigma_{\text{int}} > 0$, write down the posterior density

$$P(M_0, \theta, \sigma_{\text{int}} | \, \mathcal{D}_K, \mathcal{D}_N) \tag{33}$$

   where $\mathcal{D}_K = \{\hat{m}_k, \hat{\mu}_{C,k}\}$ is the data of the calibrators and $\mathcal{D}_N = \{\hat{m}_i, \hat{z}_i\}$ is the data of the Hubble flow sample.

3. From the joint posterior, derive useful forms for the following:

$$P(M_0 | \, \theta, \sigma_{\text{int}}, \mathcal{D}_K, \mathcal{D}_N) \tag{34}$$

$$P(\theta | \, M_0, \sigma_{\text{int}}, \mathcal{D}_K, \mathcal{D}_N) \tag{35}$$

$$P(\theta | \, \sigma_{\text{int}}, \mathcal{D}_K, \mathcal{D}_N) \tag{36}$$

4. Assuming $\sigma_{\text{int}} = 0.12$, construct and implement an algorithm to sample the posterior

$$P(h | \, \sigma_{\text{int}} = 0.12, \mathcal{D}_K, \mathcal{D}_N). \tag{37}$$

   Apply this to the data from Dhawan et al. 2018, available in the machine-readable tables provided online. Plot a histogram of the posterior samples of h, and estimate the posterior mean and standard deviation of h. Express the uncertainty as a percentage fractional standard deviation.

5. Now with σ_{int} unknown, construct and implement an MCMC algorithm to generate samples from the marginal posterior:

$$P(h, \sigma_{\text{int}} | \, \mathcal{D}_K, \mathcal{D}_N) \tag{38}$$

   Describe how you initialise your chains, tune any proposals necessary, and assess convergence. Make a scatter plot of the samples from the joint distribution, plot marginal posterior histograms of the parameters, and compute the posterior mean and fractional standard deviation of h.

6. Suppose the astronomer now realises that there is a systematic error in measuring the magnitudes of the Hubble Flow sample. For only the Hubble flow supernovae, the measured magnitude is now related to the true magnitude by

$$\hat{m}_i | \, m_i = m_i + \epsilon_{m,i} + \xi \tag{39}$$

   The photometric errors $\epsilon_{m,i} \sim N(0, \sigma^2_{m,i})$ are still mutually independent. The systematic error is ξ, which affects all the Hubble flow supernovae equally. The true value of the systematic error ξ is unknown, but we have some prior knowledge: $\xi \sim N(0, \sigma^2_\xi)$, where the variance σ^2_ξ is known.

   Modify the likelihood, posterior, and sampler appropriately, and compute the posterior mean and fractional standard deviation of h for $\sigma_\xi = 0.01, 0.02$, and 0.05.