

Revision Class

Astrostatistics (Part III)

14 May 2020 @ 3:30pm BST

Remote

Future

- basic online open book pass/fail exam with straightforward questions (3 of 3) in June
- another one in September?
- Conventional classed (distinction/merit) exam when possible

Course Outline

- Astronomy Foundations: Luminosity, flux, magnitudes, distance modulus, spectra, redshifts, light curves (time series)
- Probability Foundations: random variables; joint, conditional, marginal pdfs, limit theorems (LLN, CLT)
- Frequentist properties of estimators: expectation/bias, variance, MSE
- Properties of Maximum Likelihood, Fisher Information, Cramér-Rao bound
- **Application:** Estimating the Hubble constant using Cepheid and supernova data
- Properties of 1D and Multivariate Gaussian random variables and distributions

Course Outline

- Latent Variable and Forward Modelling to construct a complex likelihood function:
- Linear Regression with heteroskedastic (x,y) -measurement errors and intrinsic dispersion
- **Application:** quasar analysis
- Bayesian Inference: Likelihood, Prior, Posterior
- **Application:** Estimating distances from parallax
- Methods of Bayesian computation
- Approximating posterior expectations using Monte Carlo
- Importance Sampling
- **Application:** Estimating the mass of the Milky Way Galaxy

Course Outline

- Markov Chain Monte Carlo (MCMC) methods
- Metropolis and Metropolis-Hastings algorithms
- Gibbs sampling for multi-parameter problems
- Metropolis-within-Gibbs algorithms
- Convergence in theory: stationary distributions and detailed balance
- Convergence and mixing in practice: Multiple Chains/Gelman-Rubin diagnostics, autocorrelation, effective sample size
- **Application:** Estimation of cosmological parameter using supernovae as standard candles

Course Outline

- Gaussian Processes: modelling and applications, kernels and hyperparameters, marginal likelihood and predictive distributions
- **Application:** Supernova and quasar time delay estimation
- Hierarchical Bayesian / Multi-level models
 - Describing data, latent variables and hyperparameters within a joint probabilistic framework
 - Probabilistic Graphical Models (PGM) to depict conditional independence structure
 - Shrinkage in Population analyses
- **Application:** Estimating the intrinsic absolute magnitude distributions of supernovae
- Bayesian Model Comparison: Evaluating the Evidence / Marginal Likelihood
 - Evidence and the prior predictive density in data-space
 - Laplace Approximations

Example Sheet 4: Bayesian Model Comparison 3

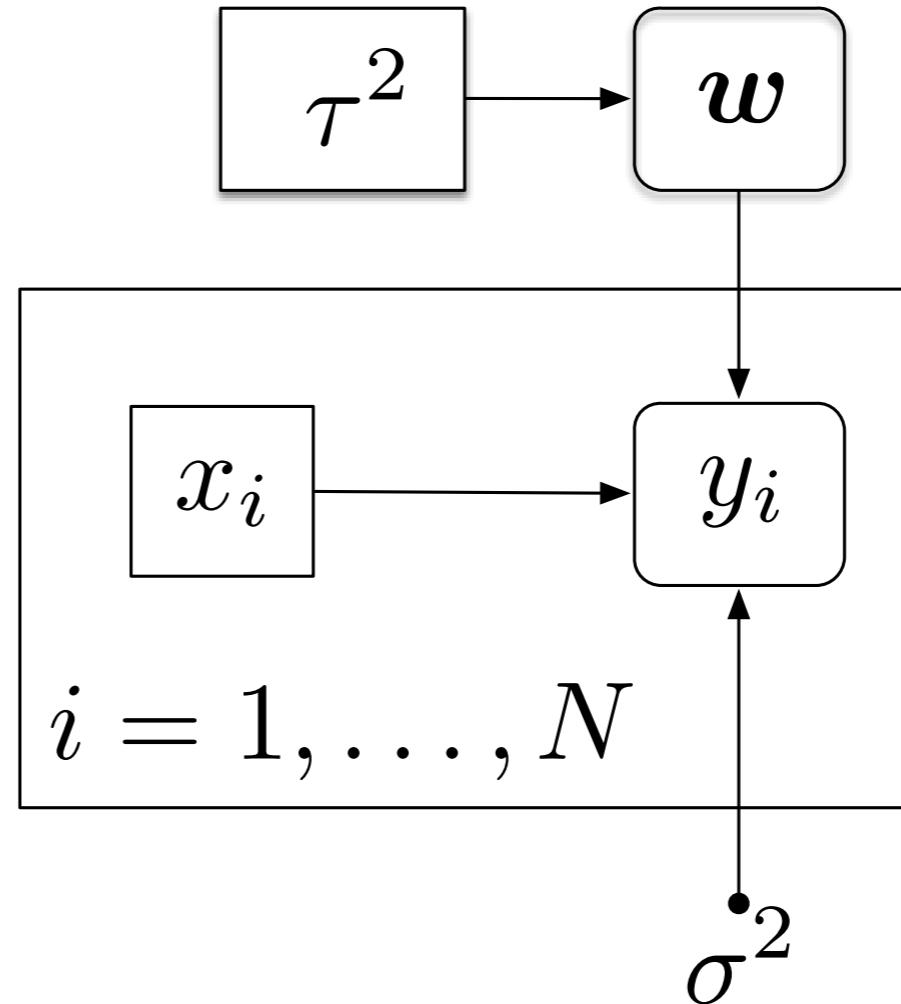
3. Consider Bayesian model selection over p -order polynomials. Given known covariates \mathbf{x} , the model for the data \mathbf{y} under hypothesis H_p is:

$$y_i \sim N \left(\sum_{j=0}^p w_j x_i^j, \sigma^2 \right) \quad (7)$$

with priors $w_j \sim N(0, \tau^2)$ for $j = 0, \dots, p$ with known τ^2 and σ^2 , and $i = 1, \dots, N$.

- (a) Draw a probabilistic graphical model depicting the joint distribution of the data \mathbf{y} and parameters \mathbf{w} , given all known quantities.

Let $\mathbf{w} = (w_1, \dots, w_p)^T$



Example Sheet 4: Bayesian Model Comparison 3

3. Consider Bayesian model selection over p -order polynomials. Given known covariates \mathbf{x} , the model for the data \mathbf{y} under hypothesis H_p is:

$$y_i \sim N\left(\sum_{j=0}^p w_j x_i^j, \sigma^2\right) \quad (7)$$

with priors $w_j \sim N(0, \tau^2)$ for $j = 0, \dots, p$ with known τ^2 and σ^2 , and $i = 1, \dots, N$.

- (b) Derive a general expression for the posterior distribution of the polynomial coefficients $P(\mathbf{w} | \mathbf{y}, \mathbf{x}; H_p)$ and the evidence for hypothesis H_p for arbitrary $p \geq 0$.

For the p th-order polynomial model H_p , let $\mathbf{w} = (w_0, \dots, w_p)^T$ and let the design matrix \mathbf{D} have rows $\mathbf{D}_i = (1, x, x^2, \dots, x^p)$. The posterior of the coefficients \mathbf{w} is

Posterior:

$$P(\mathbf{w} | \mathcal{D}) \propto \left[\prod_{i=1}^N N\left(y_i \middle| \sum_{j=0}^p w_j x_i^j, \sigma^2\right) \right] \times \left[\prod_{j=0}^p N(w_j | 0, \tau^2) \right]$$

Use result from OLS

Product of MV Gaussians

$$\begin{aligned} & \propto N(\mathbf{y} | \mathbf{D}\mathbf{w}, \sigma^2 \mathbf{I}_N) \times N(\mathbf{w} | \mathbf{0}, \tau^2 \mathbf{I}_{p+1}) \\ & \propto N(\mathbf{w} | (\mathbf{D}^T \mathbf{D})^{-1} \mathbf{D}^T \mathbf{y}, \sigma^2 (\mathbf{D}^T \mathbf{D})^{-1}) \times N(\mathbf{w} | \mathbf{0}, \tau^2 \mathbf{I}_{p+1}) \\ & = N(\mathbf{w} | \hat{\mathbf{w}}, \mathbf{V}_{\mathbf{w}}) \end{aligned}$$

where $\mathbf{V}_{\mathbf{w}}^{-1} \equiv \sigma^{-2} \mathbf{D}^T \mathbf{D} + \tau^{-2} \mathbf{I}_{p+1}$ and $\hat{\mathbf{w}} = \mathbf{V}_{\mathbf{w}} [\sigma^{-2} \mathbf{D}^T \mathbf{y}]$.

Example Sheet 4: Bayesian Model Comparison 3

3. Consider Bayesian model selection over p -order polynomials. Given known covariates \mathbf{x} , the model for the data \mathbf{y} under hypothesis H_p is:

$$y_i \sim N\left(\sum_{j=0}^p w_j x_i^j, \sigma^2\right) \quad (7)$$

with priors $w_j \sim N(0, \tau^2)$ for $j = 0, \dots, p$ with known τ^2 and σ^2 , and $i = 1, \dots, N$.

- (b) Derive a general expression for the posterior distribution of the polynomial coefficients $P(\mathbf{w} | \mathbf{y}, \mathbf{x}; H_p)$ and the evidence for hypothesis H_p for arbitrary $p \geq 0$.

The evidence Z_p is

Using MV Gaussian marginalisation

$$\begin{aligned} Z_p &= \int N(\mathbf{y} | \mathbf{D}\mathbf{w}, \sigma^2 \mathbf{I}_N) \times N(\mathbf{w} | \mathbf{0}, \tau^2 \mathbf{I}_{p+1}) d\mathbf{w} \\ &= N(\mathbf{y} | \mathbf{0}, \sigma^2 \mathbf{I}_N + \tau^2 \mathbf{D}\mathbf{D}^T). \end{aligned}$$

Example Sheet 4: Bayesian Model Comparison 3

3. Consider Bayesian model selection over p -order polynomials. Given known covariates \mathbf{x} , the model for the data \mathbf{y} under hypothesis H_p is:

$$y_i \sim N\left(\sum_{j=0}^p w_j x_i^j, \sigma^2\right) \quad (7)$$

with priors $w_j \sim N(0, \tau^2)$ for $j = 0, \dots, p$ with known τ^2 and σ^2 , and $i = 1, \dots, N$.

- (c) Analyse the following dataset:

$$\mathbf{x} = (-2.0, -1.5, 1.0, -0.5, 0, 0.5, 1.0, 1.5, 2.0) \quad (27)$$

$$\mathbf{y} = (-5.76, -1.54, -0.55, 0.29, -0.90, -1.21, -0.95, 2.80, 6.43). \quad (28)$$

Assume $\sigma = 1$ and $\tau = 1$. Compute the evidence for each hypothesis H_p for $p = 0, \dots, 8$. Which model has the greatest evidence? For the model with the greatest evidence, compute the posterior distribution of the parameters \mathbf{w} and plot the fitted model against the data.

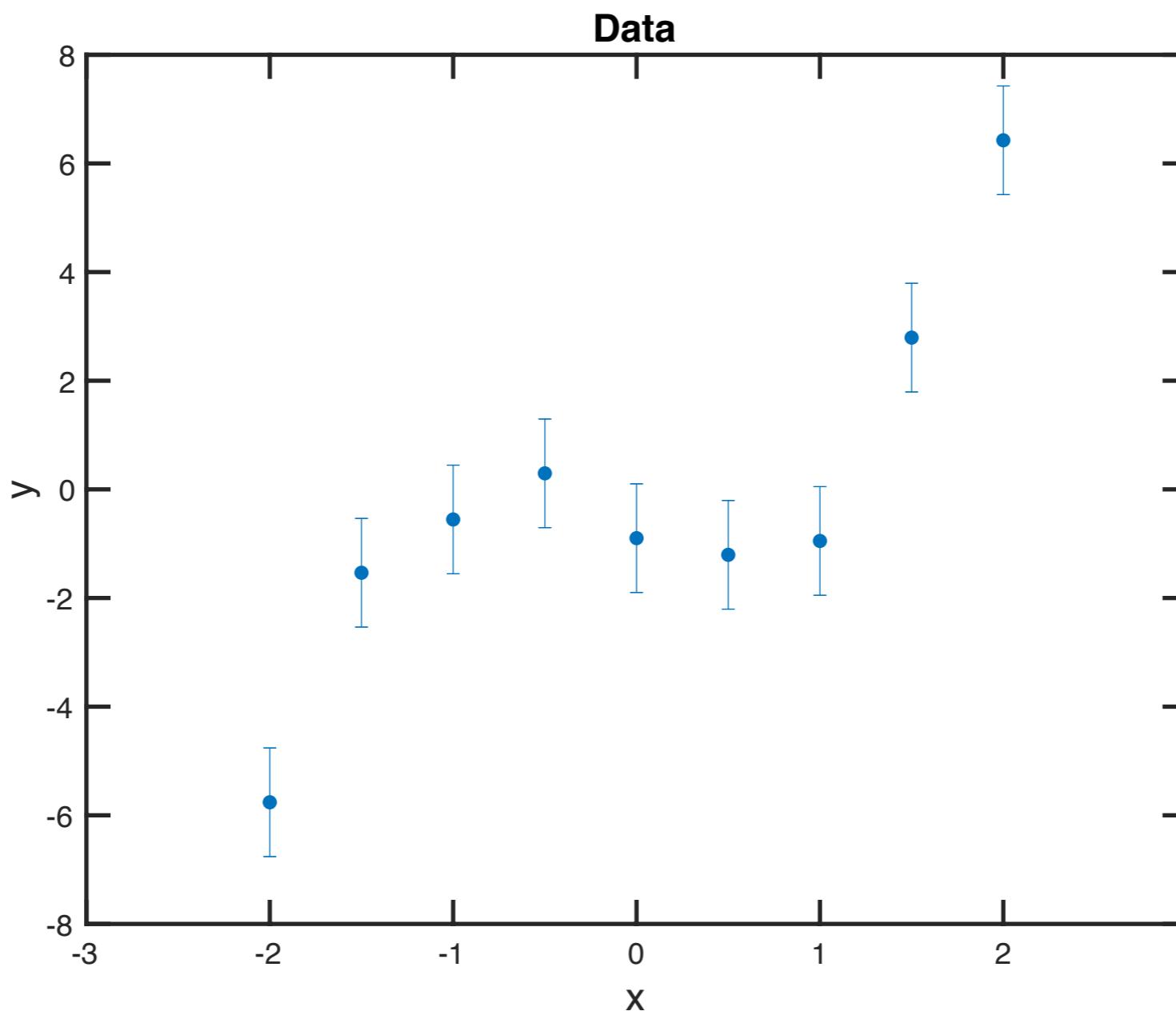
Example Sheet 4: Bayesian Model Comparison 3

(c) Analyse the following dataset:

$$\mathbf{x} = (-2.0, -1.5, 1.0, -0.5, 0, 0.5, 1.0, 1.5, 2.0) \quad (27)$$

$$\mathbf{y} = (-5.76, -1.54, -0.55, 0.29, -0.90, -1.21, -0.95, 2.80, 6.43). \quad (28)$$

Assume $\sigma = 1$ and $\tau = 1$. Compute the evidence for each hypothesis H_p for $p = 0, \dots, 8$. Which model has the greatest evidence? For the model with the greatest evidence, compute the posterior distribution of the parameters \mathbf{w} and plot the fitted model against the data.



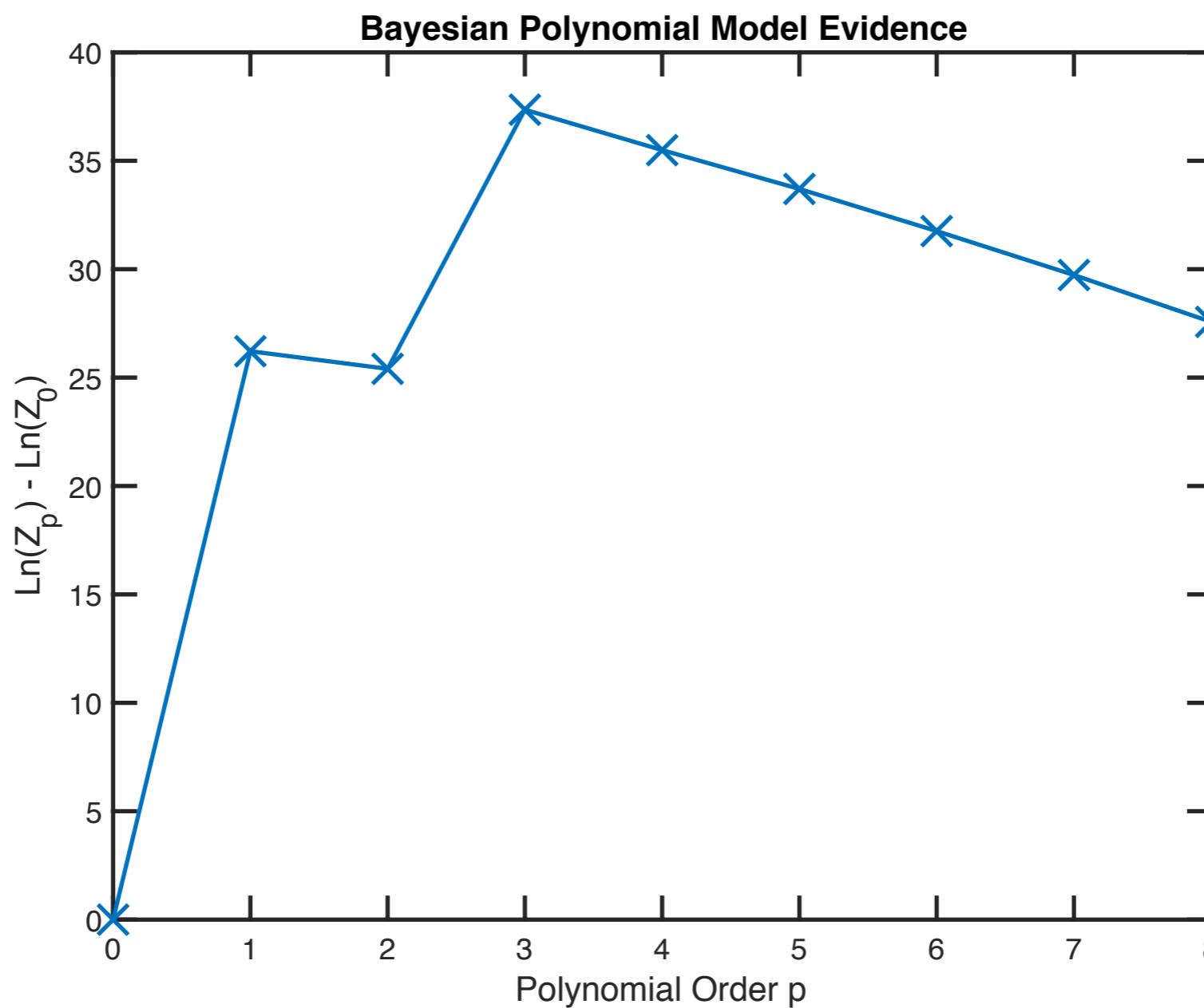
Example Sheet 4: Bayesian Model Comparison 3

(c) Analyse the following dataset:

$$\mathbf{x} = (-2.0, -1.5, 1.0, -0.5, 0, 0.5, 1.0, 1.5, 2.0) \quad (27)$$

$$\mathbf{y} = (-5.76, -1.54, -0.55, 0.29, -0.90, -1.21, -0.95, 2.80, 6.43). \quad (28)$$

Assume $\sigma = 1$ and $\tau = 1$. Compute the evidence for each hypothesis H_p for $p = 0, \dots, 8$. Which model has the greatest evidence? For the model with the greatest evidence, compute the posterior distribution of the parameters \mathbf{w} and plot the fitted model against the data.



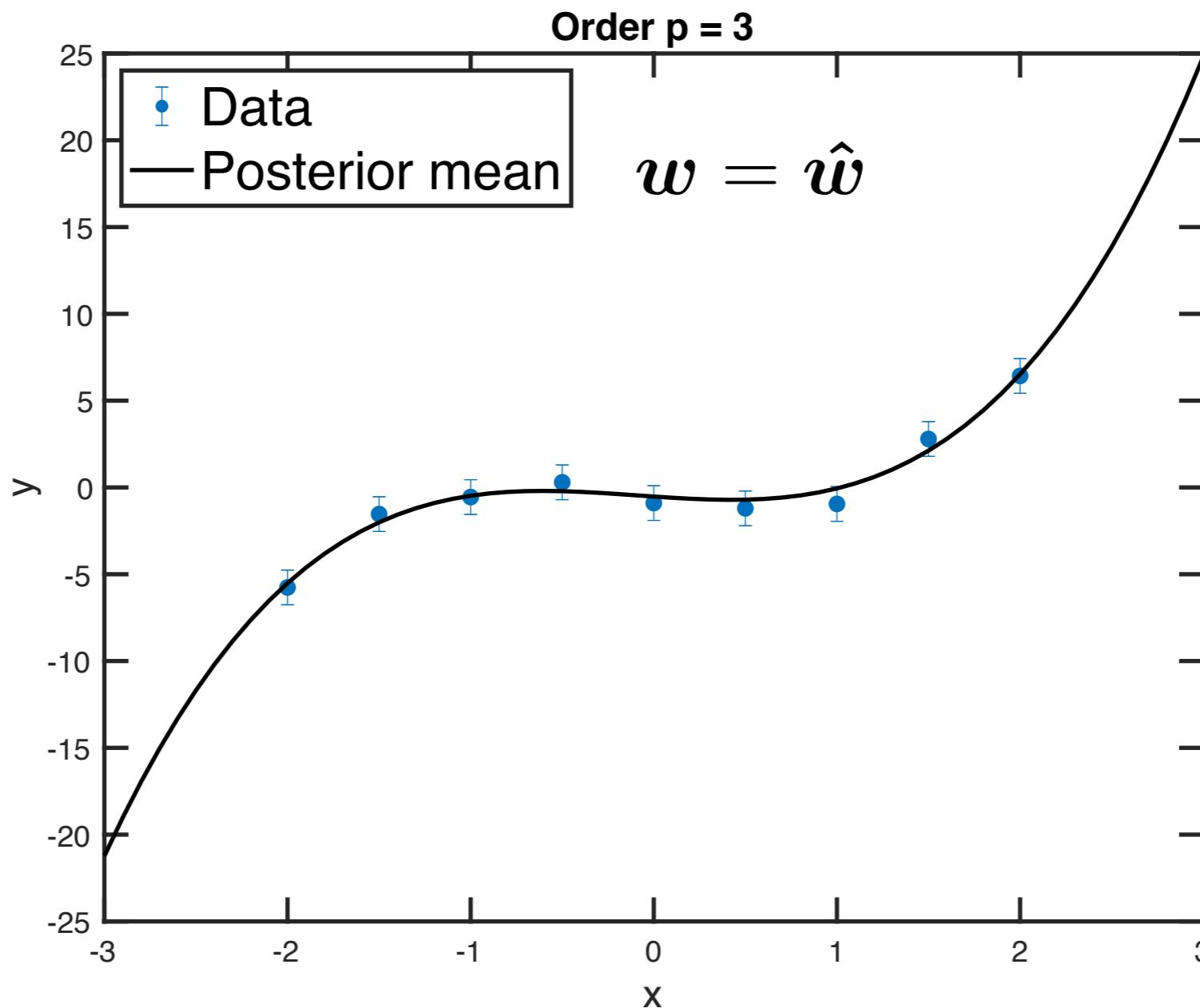
Example Sheet 4: Bayesian Model Comparison 3

(c) Analyse the following dataset:

$$\mathbf{x} = (-2.0, -1.5, 1.0, -0.5, 0, 0.5, 1.0, 1.5, 2.0) \quad (27)$$

$$\mathbf{y} = (-5.76, -1.54, -0.55, 0.29, -0.90, -1.21, -0.95, 2.80, 6.43). \quad (28)$$

Assume $\sigma = 1$ and $\tau = 1$. Compute the evidence for each hypothesis H_p for $p = 0, \dots, 8$. Which model has the greatest evidence? For the model with the greatest evidence, compute the posterior distribution of the parameters \mathbf{w} and plot the fitted model against the data.



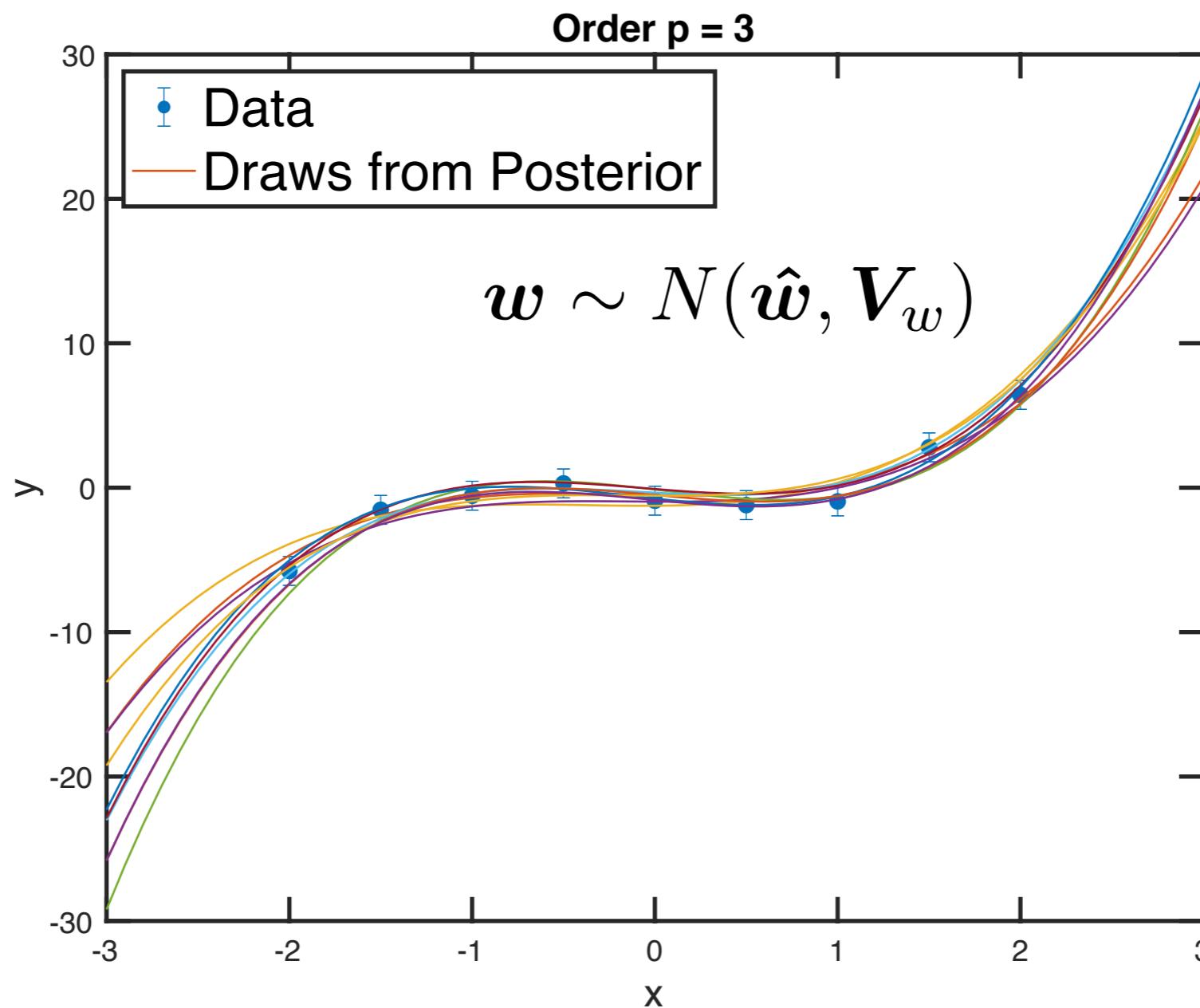
Example Sheet 4: Bayesian Model Comparison 3

(c) Analyse the following dataset:

$$\mathbf{x} = (-2.0, -1.5, 1.0, -0.5, 0, 0.5, 1.0, 1.5, 2.0) \quad (27)$$

$$\mathbf{y} = (-5.76, -1.54, -0.55, 0.29, -0.90, -1.21, -0.95, 2.80, 6.43). \quad (28)$$

Assume $\sigma = 1$ and $\tau = 1$. Compute the evidence for each hypothesis H_p for $p = 0, \dots, 8$. Which model has the greatest evidence? For the model with the greatest evidence, compute the posterior distribution of the parameters \mathbf{w} and plot the fitted model against the data.



Example Sheet 4: Problem 3

3 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 ,

$$\phi_i(x) = \exp\left[-\frac{(x - c_i)^2}{l^2}\right] \quad (10)$$

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:

$$f(x) = \sum_{i=1}^J w_i \phi_i(x). \quad (11)$$

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 $k(x, x') = \text{Cov}[f(x), f(x')] ?$

Solution: The expectation is

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

3 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 ,

$$\phi_i(x) = \exp\left[-\frac{(x - c_i)^2}{l^2}\right] \quad (10)$$

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:

$$f(x) = \sum_{i=1}^J w_i \phi_i(x). \quad (11)$$

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

The kernel is

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

where δ_{ij} is a Kronecker delta function.

3 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 ,

$$\phi_i(x) = \exp\left[-\frac{(x - c_i)^2}{l^2}\right] \quad (10)$$

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:

$$f(x) = \sum_{i=1}^J w_i \phi_i(x). \quad (11)$$

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

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

$$k(x, x') = \sigma^2 \sum_{i=1}^J \phi_i(x) \phi_i(x') \frac{J-1}{J} \Delta c$$

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

$$\begin{aligned} k(x, x') &= \sigma^2 \int_{c_1}^{c_J=c_1+h} \phi_i(x) \phi_i(x') dc \\ &= \sigma^2 \int_{c_1}^{c_1+h} e^{-(x-c)^2/l^2} e^{-(x'-c)^2/l^2} dc \end{aligned} \quad (10)$$

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

$$k(x, x') = \sigma^2 \int_{-\infty}^{+\infty} e^{-(x-c)^2/l^2} e^{-(x'-c)^2/l^2} dc \quad (11)$$

3 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 ,

$$\phi_i(x) = \exp\left[-\frac{(x - c_i)^2}{l^2}\right] \quad (10)$$

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:

$$f(x) = \sum_{i=1}^J w_i \phi_i(x). \quad (11)$$

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

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

Noting that

$$e^{-(x-c)^2/l^2} = \frac{l}{\sqrt{2}} \sqrt{2\pi} N(x|c, l^2/2) = \frac{l}{\sqrt{2}} \sqrt{2\pi} N(c|x, l^2/2), \quad (12)$$

we find

$$\begin{aligned} k(x, x') &= \sigma^2 \left(\frac{l}{\sqrt{2}} \sqrt{2\pi} \right)^2 \int_{-\infty}^{+\infty} N(x|c, l^2/2) N(c|x', l^2/2) dc \\ &= \sigma^2 \left(\frac{l}{\sqrt{2}} \sqrt{2\pi} \right)^2 N(x|x', l^2) \end{aligned} \quad (13)$$

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

$$k(x, x') = \frac{l\sigma^2}{2} \sqrt{2\pi} e^{-(x-x')^2/2l^2} \quad (14)$$

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

3 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 ,

$$\phi_i(x) = \exp\left[-\frac{(x - c_i)^2}{l^2}\right] \quad (10)$$

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:

$$f(x) = \sum_{i=1}^J w_i \phi_i(x). \quad (11)$$

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

Kernel (Covariance) Function:

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

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

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

Example Sheet 3, Problem 2

2 Supernova Cosmology

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

$$M_s \sim N(M_0, \sigma_{\text{int}}^2) \quad (4)$$

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

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

where the true distance modulus at the observed redshift z_s is

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

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

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

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

2 Supernova Cosmology

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

$$M_s \sim N(M_0, \sigma_{\text{int}}^2) \quad (4)$$

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

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

where the true distance modulus at the observed redshift z_s is

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

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

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

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

- Derive likelihood function for the sample of N supernovae:

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

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

Combining the above equations, we have:

$$m_s = M_0 - 5 \log h + f(z_s; w, \Omega_M) + \epsilon_{\text{int}}^s = \mathcal{M} + f(z_s; w, \Omega_M) + \epsilon_{\text{int}}^s$$

where $\epsilon_{\text{int}}^s \sim N(0, \sigma_{\text{int}}^2)$.

2 Supernova Cosmology

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

$$M_s \sim N(M_0, \sigma_{\text{int}}^2) \quad (4)$$

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

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

where the true distance modulus at the observed redshift z_s is

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

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

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

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

So the likelihood function for one SN is

$$P(m_s | z_s, \mathcal{M}, \sigma_{\text{int}}^2, w, \Omega_M) = N(m_s | \mathcal{M} + f(z_s; w, \Omega_M), \sigma_{\text{int}}^2)$$

The likelihood for all N independent SNe is:

$$L(\mathcal{M}, \sigma_{\text{int}}^2, w, \Omega_M) = \prod_{s=1}^N N(m_s | \mathcal{M} + f(z_s; w, \Omega_M), \sigma_{\text{int}}^2)$$

So the likelihood function for one SN is

$$P(m_s | z_s, \mathcal{M}, \sigma_{\text{int}}^2, w, \Omega_M) = N(m_s | \mathcal{M} + f(z_s; w, \Omega_M), \sigma_{\text{int}}^2)$$

The likelihood for all N independent SNe is:

$$L(\mathcal{M}, \sigma_{\text{int}}^2, w, \Omega_M) = \prod_{s=1}^N N(m_s | \mathcal{M} + f(z_s; w, \Omega_M), \sigma_{\text{int}}^2)$$

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

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

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

The unnormalised posterior is:

$$P(\mathcal{M}, \sigma_{\text{int}}^2, w, \Omega_M | \mathcal{D}) \propto L(\mathcal{M}, \sigma_{\text{int}}^2, w, \Omega_M) H(\Omega_M) H(\sigma_{\text{int}}^2).$$

So the likelihood function for one SN is

$$P(m_s|z_s, \mathcal{M}, \sigma_{\text{int}}^2, w, \Omega_M) = N(m_s|\mathcal{M} + f(z_s; w, \Omega_M), \sigma_{\text{int}}^2)$$

The likelihood for all N independent SNe is:

$$L(\mathcal{M}, \sigma_{\text{int}}^2, w, \Omega_M) = \prod_{s=1}^N N(m_s|\mathcal{M} + f(z_s; w, \Omega_M), \sigma_{\text{int}}^2)$$

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

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

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

The conditional $P(\mathcal{M} | \sigma_{\text{int}}^2, w, \Omega_M; \mathcal{D})$ is:

$$\begin{aligned} P(\mathcal{M} | \sigma_{\text{int}}^2, w, \Omega_M; \mathcal{D}) &\propto \prod_{s=1}^N N[m_s | \mathcal{M} + f(z_s; w, \Omega_M), \sigma_{\text{int}}^2] \\ &\propto \prod_{s=1}^N N[\mathcal{M} | m_s - f(z_s; w, \Omega_M), \sigma_{\text{int}}^2] \\ &= N(\mathcal{M} | \bar{M}, \sigma_{\text{int}}^2/N) \end{aligned}$$

where $\bar{M} \equiv \frac{1}{N} \sum_{s=1}^N m_s - f(z_s; w, \Omega_M)$.

Another derivation shows that $P(\sigma_{\text{int}}^2 | \mathcal{M}, w, \Omega_M; \mathcal{D})$ has an inverse χ^2 distribution.

Mixed Gibbs sampler

1. Draw and accept a new $\mathcal{M} \sim P(\mathcal{M} | \sigma_{\text{int}}^2, w, \Omega_M; \mathcal{D})$ **(Gaussian)**
2. Draw and accept a new $\sigma_{\text{int}}^2 \sim P(\sigma_{\text{int}}^2 | \mathcal{M}, w, \Omega_M; \mathcal{D})$ **(Inv- χ^2)**
3. Metropolis step: propose/accept/reject (w, Ω_M)
 $(w, \Omega_M)^* \sim N([w, \Omega_M], \Sigma_J)$

Accept (w^*, Ω_M^*) with probability $r = \frac{P(w^*, \Omega_M^* | \mathcal{M}, \sigma_{\text{int}}^2; \mathcal{D})}{P(w, \Omega_M | \mathcal{M}, \sigma_{\text{int}}^2; \mathcal{D})}$

Otherwise remain at (w, Ω_M)