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: expection/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

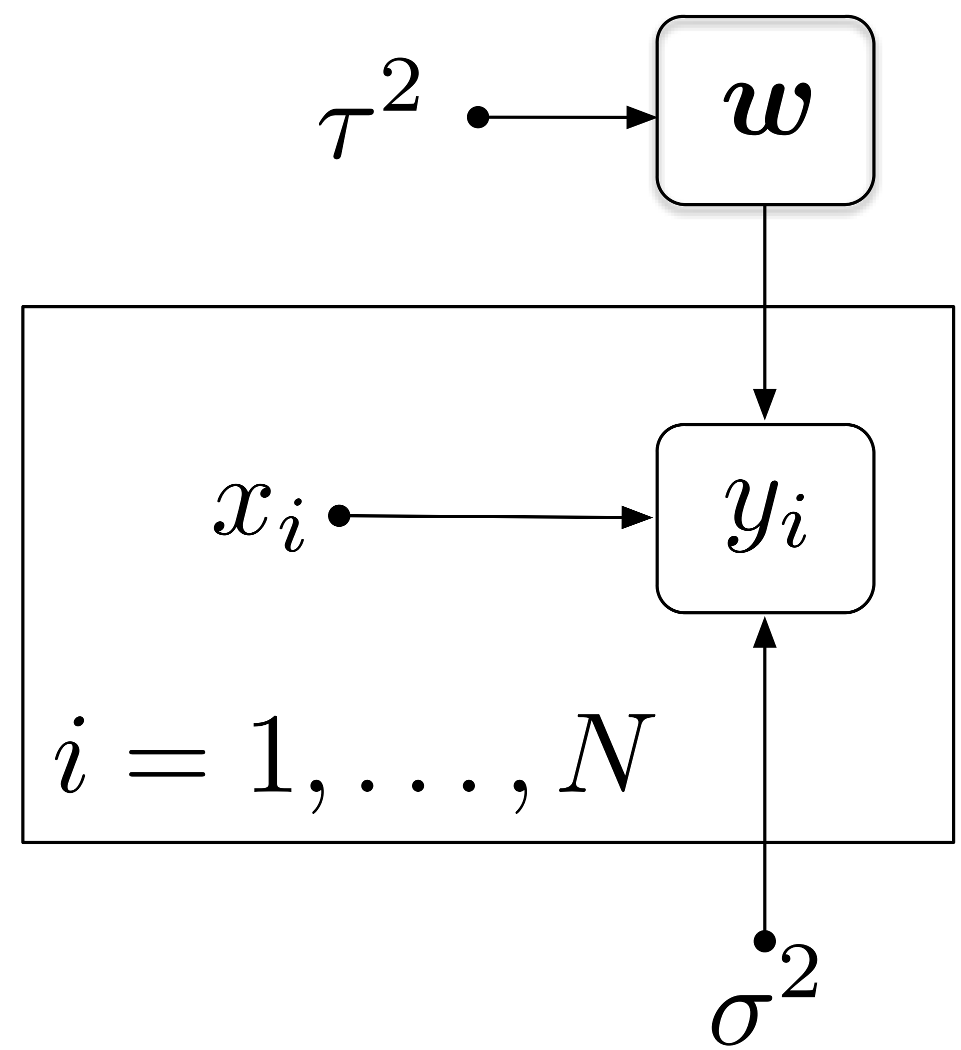
3. Consider Bayesian model selection over p-order polynomials. Given known covariates $\boldsymbol{x}$, the model for the data $\boldsymbol{y}$ under hypothesis H_p is:

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

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

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

Let $\boldsymbol{w} = (w_0, \ldots, w_p)^T$



Example Sheet 4: Bayesian Model Comparison 3

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

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

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

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

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

Posterior:

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

Use result from OLS

$$\propto N(\boldsymbol{y}|\, \boldsymbol{D}\,\boldsymbol{w}, \sigma^2 \boldsymbol{I}_N) \times N(\boldsymbol{w}|\, \boldsymbol{0}, \tau^2 \boldsymbol{I}_{p+1})$$

Product of MV Gaussians

$$\propto N(\boldsymbol{w}|\, (\boldsymbol{D}^T\boldsymbol{D})^{-1}\boldsymbol{D}^T\boldsymbol{y}, \sigma^2(\boldsymbol{D}^T\boldsymbol{D})^{-1}) \times N(\boldsymbol{w}|\, \boldsymbol{0}, \tau^2 \boldsymbol{I}_{p+1})$$

$$= N(\boldsymbol{w}|\, \hat{\boldsymbol{w}}, \boldsymbol{V}_{\boldsymbol{w}})$$

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

Example Sheet 4: Bayesian Model Comparison 3

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

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

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

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

The evidence Z_p is

Using MV Gaussian marginalisation

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

Example Sheet 4: Bayesian Model Comparison 3

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

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

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

(c) Analyse the following dataset:

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

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

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

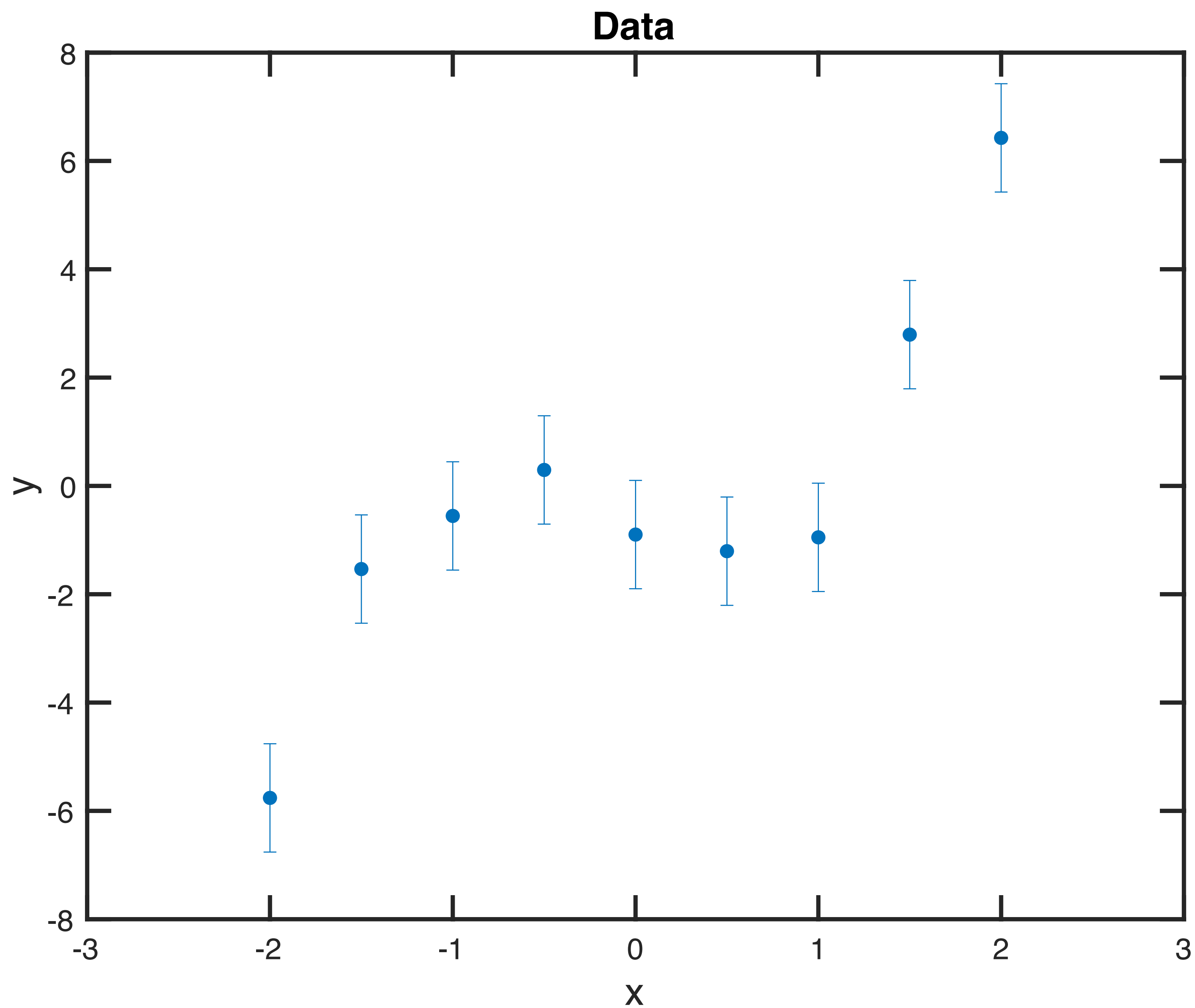
Example Sheet 4: Bayesian Model Comparison 3

(c) Analyse the following dataset:

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

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

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



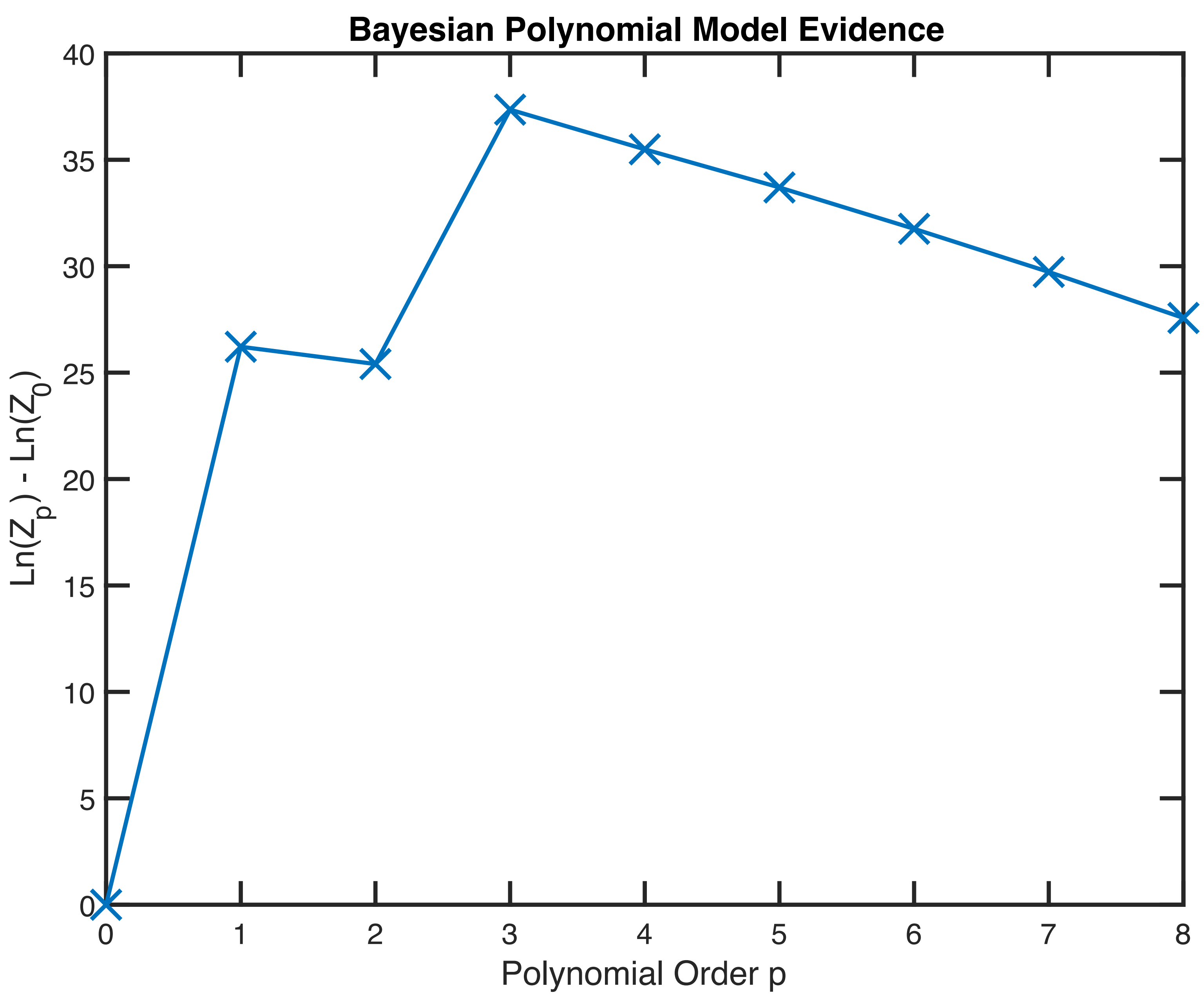
Example Sheet 4: Bayesian Model Comparison 3

(c) Analyse the following dataset:

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

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

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



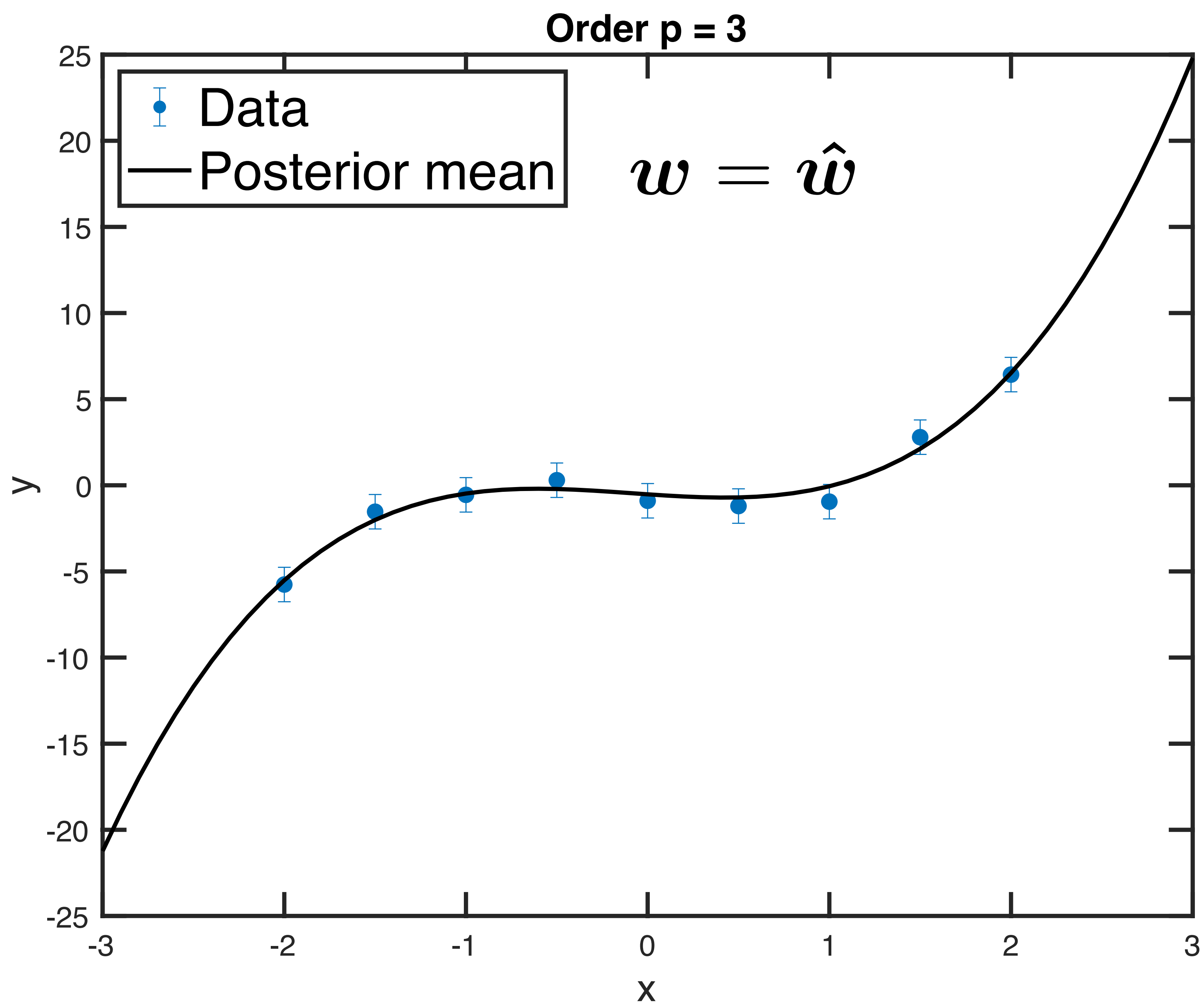
Example Sheet 4: Bayesian Model Comparison 3

(c) Analyse the following dataset:

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

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

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



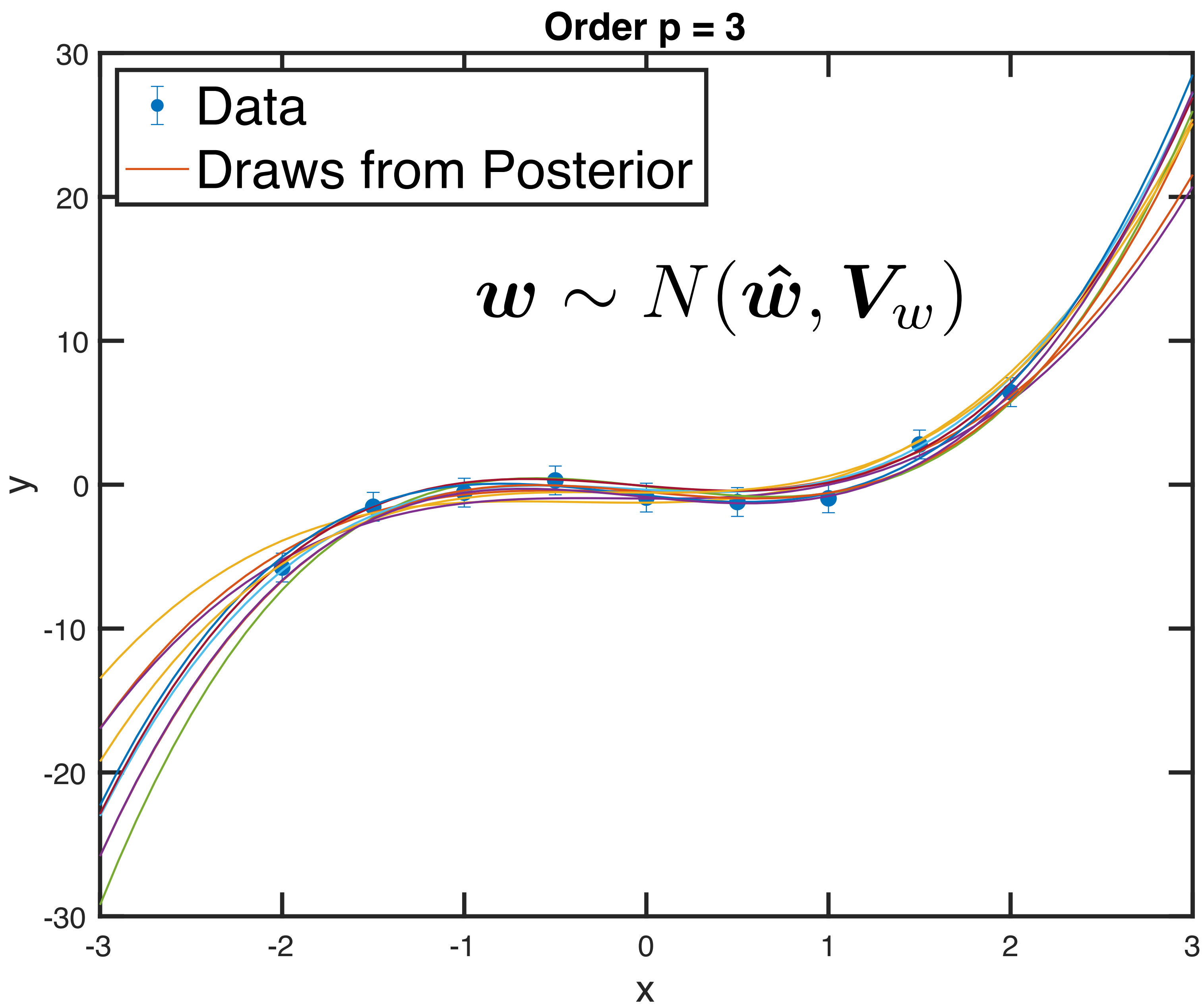
Example Sheet 4: Bayesian Model Comparison 3

(c) Analyse the following dataset:

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

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

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

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

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:

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

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

   Solution: The expectation is

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

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,

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

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:

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

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

The kernel is

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

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,

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

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:

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

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: $J \to \infty$ and $c_1 \to -\infty$, $h \to \infty$.

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

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

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

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

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

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,

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

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:

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

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: $J \to \infty$ and $c_1 \to -\infty$, $h \to \infty$.

Noting that

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

we find

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

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

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

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,

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

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:

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

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

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

$$M_s \sim N(M_0, \sigma^2_{\text{int}}) \tag{4}$$

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

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

where the true distance modulus at the observed redshift z_s is

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

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,

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

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

$$M_s \sim N(M_0, \sigma^2_{\text{int}}) \tag{4}$$

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

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

where the true distance modulus at the observed redshift z_s is

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

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,

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

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.

1. Derive likelihood function for the sample of N supernovae:

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

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:

$$m_s = M_0 - 5\log h + f(z; w, \Omega_M) + \epsilon^s_{\textbf{int}} = \mathcal{M} + f(z_s; w, \Omega_M) + \epsilon^s_{\textbf{int}}$$

where $\epsilon^s_{\textbf{int}} \sim N(0, \sigma^2_{\textbf{int}})$**.**

defining: $$f(z_s; w, \Omega_M) = 25 + 5\log_{10}\left[\frac{c}{(100\ \textbf{km s}^{-1}\ \textbf{Mpc}^{-1})}\tilde{d}(z_s; w, \Omega_M)\ \textbf{Mpc}^{-1}\right]$$

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

$$M_s \sim N(M_0, \sigma^2_{\text{int}}) \tag{4}$$

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

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

where the true distance modulus at the observed redshift z_s is

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

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,

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

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

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

The likelihood for all N independent SNe is:

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

So the likelihood function for one SN is

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

The likelihood for all N independent SNe is:

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

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

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

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

The unnormalised posterior is:

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

So the likelihood function for one SN is

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

The likelihood for all N independent SNe is:

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

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

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

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

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

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

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

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

Mixed Gibbs sampler

1. Draw and accept a new $\mathcal{M} \sim P(\mathcal{M}|\,\sigma^2_{\rm int}, w, \Omega_M; \mathcal{D})$ **(Gaussian)**

2. Draw and accept a new $\sigma^2_{\rm int} \sim P(\sigma^2_{\rm int}|\,\mathcal{M}, w, \Omega_M; \mathcal{D})$ **(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** $r = \dfrac{P(w^*, \Omega_M^*|\mathcal{M}, \sigma^2_{\rm int}; \mathcal{D})}{P(w, \Omega_M|\mathcal{M}, \sigma^2_{\rm int}; \mathcal{D})}$

Otherwise remain at (w, Ω_M)