Example Class 4

Astrostatistics (Part III)

Future

- Revision Class scheduled for Thu, 14 May @ 3:30pm (online, remote)

- basic online open book pass/fail exam with straightforward questions in June

- another one in September?

- Conventional classed (distinction/merit) exam when possible

Example Sheet 4

1 Probabilistic Graphical Models and Gibbs Sampling

Consider linear regression of the quasars' X-ray spectral indices vs. bolometric luminosities in the presence of heteroskedastic measurement error in both quantities and intrinsic dispersion. Consider the probabilistic generative model described in class:

$$\xi_i | \, \mu, \tau^2 \sim N(\mu, \tau^2) \tag{1}$$

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

$$x_i | \xi_i \sim N(\xi_i, \sigma_{x,i}^2) \tag{3}$$

$$y_i | \eta_i \sim N(\eta_i, \sigma_{y,i}^2) \tag{4}$$

The astronomer observes values $\mathcal{D} = \{x_i, y_i\}$, which are noisy measurements of the true luminosity ξ_i and the true spectral index η_i of each quasar. The measurement errors are independent and heterskedastic with known variances $\{\sigma_{x,i}^2, \sigma_{y,i}^2\}$, for $i = 1, \ldots, N$ independent quasars.

1 Probabilistic Graphical Models and Gibbs Sampling

Consider linear regression of the quasars' X-ray spectral indices vs. bolometric luminosities in the presence of heteroskedastic measurement error in both quantities and intrinsic dispersion. Consider the probabilistic generative model described in class:

$$\xi_i|\,\mu,\tau^2 \sim N(\mu,\tau^2) \tag{1}$$

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

$$x_i|\xi_i \sim N(\xi_i,\sigma_{x,i}^2) \tag{3}$$

$$y_i|\eta_i \sim N(\eta_i,\sigma_{y,i}^2) \tag{4}$$

The astronomer observes values $\mathcal{D}=\{x_i,y_i\}$, which are noisy measurements of the true luminosity ξ_i and the true spectral index η_i of each quasar. The measurement errors are independent and heterskedastic with 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.

<u>Strategy:</u> Factor joint pdf into conditional and marginal distributions and exploit conditional independence structure of model assumptions.

Solution:

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

Consider the probabilistic generative model described in class:

$$\xi_i|\,\mu,\tau^2 \sim N(\mu,\tau^2) \tag{1}$$

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

$$x_i|\xi_i \sim N(\xi_i,\sigma^2_{x,i}) \tag{3}$$

$$y_i|\eta_i \sim N(\eta_i,\sigma^2_{y,i}) \tag{4}$$

The astronomer observes values $\mathcal{D}=\{x_i,y_i\}$, which are noisy measurements of the true luminosity ξ_i and the true spectral index η_i of each quasar. The measurement errors are independent and heterskedastic with known variances $\{\sigma^2_{x,i},\sigma^2_{y,i}\}$, for $i=1,\ldots,N$ independent quasars.

2. Adopt "non-informative" hyperpriors on the hyperparameters: flat improper priors for each of $P(\alpha)$, $P(\beta)$, $P(\mu)$, and flat positive improper priors for each of $P(\tau^2)$ and $P(\sigma^2)$. 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:

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

Utilise conditional independence of individual quasars

$$\text{Let } H(z)=\begin{cases}1, & z>0\\ 0, & z\le 0\end{cases} \qquad P(\tau^2)\propto H(z) \qquad P(\sigma^2)\propto H(z)$$

Consider the probabilistic generative model described in class:

$$\xi_i|\,\mu,\tau^2 \sim N(\mu,\tau^2) \tag{1}$$

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

$$x_i|\xi_i \sim N(\xi_i,\sigma^2_{x,i}) \tag{3}$$

$$y_i|\eta_i \sim N(\eta_i,\sigma^2_{y,i}) \tag{4}$$

The astronomer observes values $\mathcal{D}=\{x_i,y_i\}$, which are noisy measurements of the true luminosity ξ_i and the true spectral index η_i of each quasar. The measurement errors are independent and heterskedastic with known variances $\{\sigma^2_{x,i},\sigma^2_{y,i}\}$, for $i=1,\ldots,N$ independent quasars.

2. Adopt "non-informative" hyperpriors on the hyperparameters: flat improper priors for each of $P(\alpha)$, $P(\beta)$, $P(\mu)$, and flat positive improper priors for each of $P(\tau^2)$ and $P(\sigma^2)$. Write down the full joint distribution of all data $\mathcal{D}$, latent variables $\{\xi_i,\eta_i\}$, and hyperparamters $\alpha,\beta,\sigma^2,\mu,\tau^2$.

$$P(\{x_i,y_i\},\{\xi_i,\eta_i\},\alpha,\beta,\sigma^2,\mu,\tau^2) \propto$$
$$\Big[\prod_{i=1}^{N} N(x_i|\,\xi_i,\sigma^2_{x,i})N(y_i,|\eta_i,\sigma^2_{y,i})N(\eta_i|\,\alpha+\beta\xi_i,\sigma^2)N(\xi_i|\,\mu,\tau^2)\Big]$$
$$\times\, H(\sigma^2)\,H(\tau^2)$$

Consider the probabilistic generative model described in class:

$$\xi_i|\,\mu,\tau^2 \sim N(\mu,\tau^2) \tag{1}$$

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

$$x_i|\xi_i \sim N(\xi_i,\sigma^2_{x,i}) \tag{3}$$

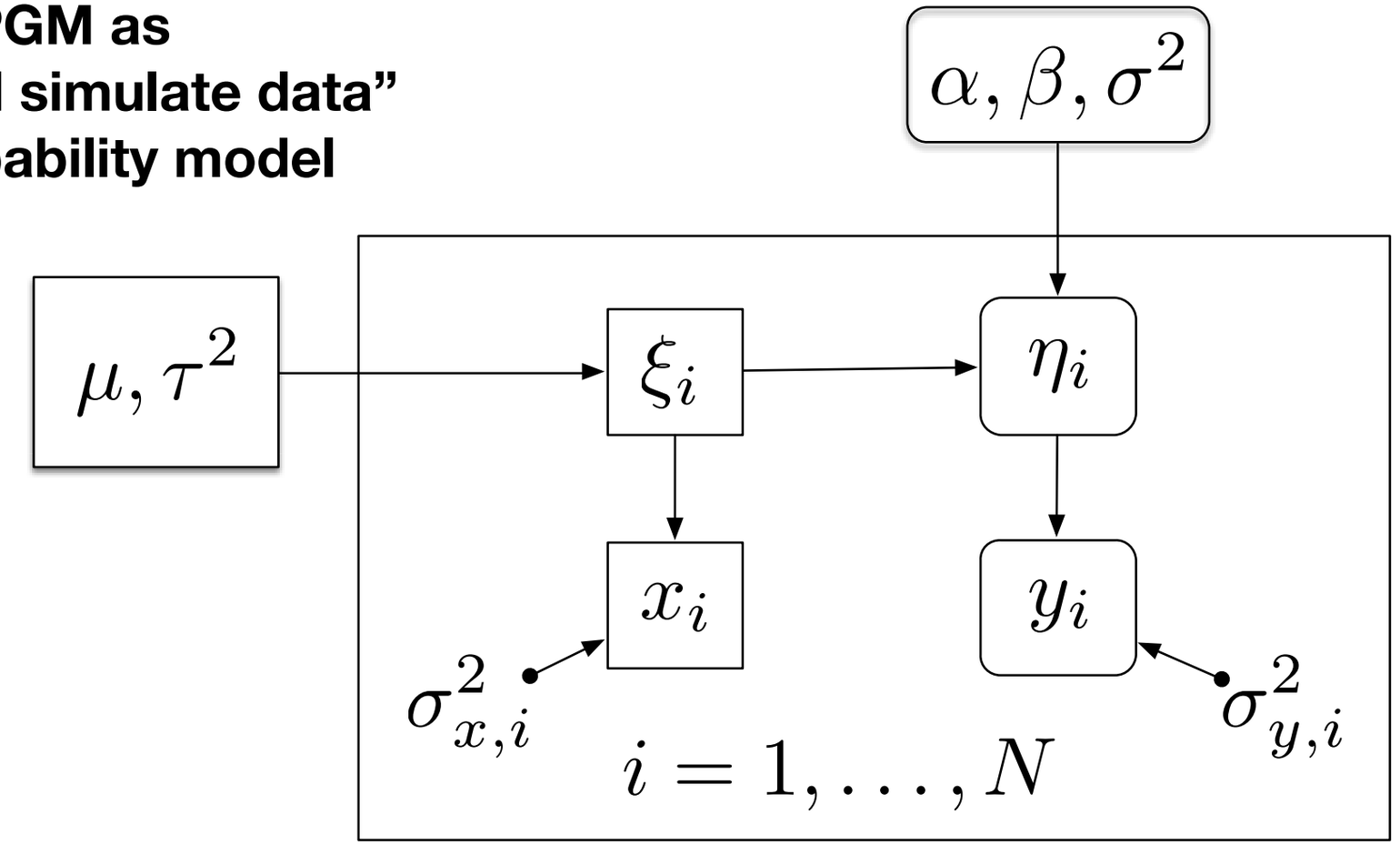
$$y_i|\eta_i \sim N(\eta_i,\sigma^2_{y,i}) \tag{4}$$

The astronomer observes values $\mathcal{D}=\{x_i,y_i\}$, which are noisy measurements of the true luminosity ξ_i and the true spectral index η_i of each quasar. The measurement errors are independent and heterskedastic with known variances $\{\sigma^2_{x,i},\sigma^2_{y,i}\}$, for $i=1,\ldots,N$ independent quasars.

3. Draw a probabilistic graphical model to represent this joint distribution.

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

Draw PGM as
"how you would simulate data"
from the probability model



$$P(\{\xi_i, \eta_i\}, \alpha, \beta, \sigma^2, \mu, \tau^2 | \mathcal{D}) \propto$$
$$\left[\prod_{i=1}^{N} N(x_i | \xi_i, \sigma_{x,i}^2) N(y_i, | \eta_i, \sigma_{y,i}^2) N(\eta_i | \alpha + \beta\xi_i, \sigma^2) N(\xi_i | \mu, \tau^2)\right]$$
$$\times H(\sigma^2)\, H(\tau^2)$$

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 sequence of proposed moves that are always accepted. Specify the order in which you run through your sequence. You have access to algorithms that generate random draws from univariate and multivariate Gaussian distributions, and scaled inverse-χ^2 distributions.

- Derive a complete set of conditional posterior densities of each (subset of) parameters conditional on all the others (and the data)

- For hierarchical models, the conditional independence structure enables a natural strategy for Gibbs sampling in **two stages**:

  - For each individual i, sample the latent variables (ξ_i, η_i) given the current values of the hyperparameters $(\alpha, \beta, \sigma^2, \mu, \tau)$

  - Then sample the hyperparameters $(\alpha, \beta, \sigma^2, \mu, \tau)$ given all the latent variables for all i $(\boldsymbol{\xi}, \boldsymbol{\eta})$

$$P(\{\xi_i, \eta_i\}, \alpha, \beta, \sigma^2, \mu, \tau^2 | \mathcal{D}) \propto$$

$$\Big[\prod_{i=1}^{N} N(x_i | \xi_i, \sigma_{x,i}^2) N(y_i, | \eta_i, \sigma_{y,i}^2) N(\eta_i | \alpha + \beta\xi_i, \sigma^2) N(\xi_i | \mu, \tau^2)\Big]$$

$$\times H(\sigma^2)\, H(\tau^2)$$

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 sequence of proposed moves that are always accepted. Specify the order in which you run through your sequence. You have access to algorithms that generate random draws from univariate and multivariate Gaussian distributions, and scaled inverse-χ^2 distributions.

- Derive a complete set of conditional posterior densities of each (subset of) parameters conditional on all the others (and the data)

- For hierarchical models, the conditional independence structure enables a natural strategy for Gibbs sampling in **two stages**:

  - For each individual i, sample the latent variables given the current values of the hyperparameters

$$P(\xi_i, \eta_i | \alpha, \beta, \sigma^2, \mu, \tau; \mathcal{D})$$

  - Then sample the hyperparameterss given all the latent variables for all i

$$P(\alpha, \beta, \sigma^2, \mu, \tau | \boldsymbol{\xi}, \boldsymbol{\eta}; \mathcal{D})$$

Each Stage can then be further broken up by inspecting conditional structure

$$P(\{\xi_i, \eta_i\}, \alpha, \beta, \sigma^2, \mu, \tau^2 | \mathcal{D}) \propto$$

$$\Big[\prod_{i=1}^{N} N(x_i | \xi_i, \sigma_{x,i}^2) N(y_i, | \eta_i, \sigma_{y,i}^2) N(\eta_i | \alpha + \beta\xi_i, \sigma^2) N(\xi_i | \mu, \tau^2)\Big]$$

$$\times H(\sigma^2)\, H(\tau^2)$$

Drawing new latent variables, conditional on hyperparameters:

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

$$\begin{aligned} P(\xi_i, \eta_i | \alpha, \beta, \sigma^2, \mu, \tau^2) &\propto N(x_i | \xi_i, \sigma_{x,i}^2) N(y_i, | \eta_i, \sigma_{y,i}^2) N(\eta_i | \alpha + \beta\xi_i, \sigma^2) N(\xi_i | \mu, \tau^2) \\ &\propto N(\boldsymbol{\phi}_i | \boldsymbol{z}_i, \boldsymbol{\Sigma}_i) \times N(\boldsymbol{\phi}_i | \boldsymbol{\gamma}, \boldsymbol{C}) \end{aligned} \tag{10}$$

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

$$\boldsymbol{C} \equiv \begin{pmatrix} \beta^2\tau^2 + \sigma^2 & \beta\tau^2 \\ \beta\tau^2 & \tau^2 \end{pmatrix} \tag{11}$$

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

$$P(\{\xi_i, \eta_i\}, \alpha, \beta, \sigma^2, \mu, \tau^2 | \mathcal{D}) \propto$$

$$\Big[\prod_{i=1}^{N} N(x_i | \xi_i, \sigma_{x,i}^2) N(y_i, | \eta_i, \sigma_{y,i}^2) N(\eta_i | \alpha + \beta\xi_i, \sigma^2) N(\xi_i | \mu, \tau^2)\Big]$$

$$\times H(\sigma^2)\, H(\tau^2)$$

Drawing new latent variables, conditional on hyperparameters:

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

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

$$\boldsymbol{\phi}_i \equiv \begin{pmatrix} \eta_i \\ \xi_i \end{pmatrix}$$

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

$$\bar{\boldsymbol{\phi}}_i \equiv (\boldsymbol{C}^{-1} + \boldsymbol{\Sigma}^{-1})^{-1} (\boldsymbol{C}^{-1}\boldsymbol{\gamma} + \boldsymbol{\Sigma}_i^{-1}\boldsymbol{z}_i) \tag{13}$$

$$\boldsymbol{V}_\phi^i = (\boldsymbol{C}^{-1} + \boldsymbol{\Sigma}_i^{-1})^{-1} \tag{14}$$

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

$$\begin{pmatrix} \eta_i \\ \xi_i \end{pmatrix} \sim N(\bar{\boldsymbol{\phi}}_i, \boldsymbol{V}_\phi^i) \tag{15}$$

for each quasar i.

$$P(\{\xi_i, \eta_i\}, \alpha, \beta, \sigma^2, \mu, \tau^2 | \, \mathcal{D}) \propto$$

$$\Big[\prod_{i=1}^{N} N(x_i | \, \xi_i, \sigma_{x,i}^2) N(y_i, | \eta_i, \sigma_{y,i}^2) N(\eta_i | \, \alpha + \beta\xi_i, \sigma^2) N(\xi_i | \, \mu, \tau^2)\Big]$$

$$\times \, H(\sigma^2) \, H(\tau^2)$$

Drawing new hyperparameters, conditional on the updated latent variables.
Break up into subsets:

Update Sequentially:

$$\text{(b) } \sigma^2 \sim P(\sigma^2 | \, \ldots; \mathcal{D}) = P(\sigma^2 | \, \alpha, \beta, \{\eta_i, \xi_i\})$$

$$\text{(c) } \alpha \sim P(\alpha | \, \ldots; \mathcal{D}) = P(\alpha | \, \beta, \sigma^2, \{\eta_i, \xi_i\})$$

$$\text{(d) } \beta \sim P(\beta | \, \ldots; \mathcal{D}) = P(\beta | \, \alpha, \sigma^2, \{\eta_i, \xi_i\})$$

$$\text{(e) } \mu, \tau^2 \sim P(\mu, \tau^2 | \, \ldots; \mathcal{D}) = P(\mu | \, \tau^2 \{\xi_i\}) \, P(\tau^2 | \{\xi_i\})$$

(draw a block of two parameters)

$$P(\{\xi_i, \eta_i\}, \alpha, \beta, \sigma^2, \mu, \tau^2 | \, \mathcal{D}) \propto$$
$$\Big[\prod_{i=1}^{N} N(x_i | \, \xi_i, \sigma_{x,i}^2) N(y_i, | \eta_i, \sigma_{y,i}^2) N(\eta_i | \, \alpha + \beta\xi_i, \sigma^2) N(\xi_i | \, \mu, \tau^2)\Big]$$
$$\times \, H(\sigma^2) \, H(\tau^2)$$

Update:

$$\text{(b) } \sigma^2 \sim P(\sigma^2 | \, \ldots; \mathcal{D}) = P(\sigma^2 | \, \alpha, \beta, \{\eta_i, \xi_i\})$$

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

$$\begin{aligned} P(\sigma^2 | \, \ldots \mathcal{D}) &\propto \prod_{i=1}^{N} N(\eta_i | \, \alpha + \beta\xi_i, \sigma^2) \\ &\propto \sigma^{-N} \prod_{i=1}^{N} \exp\left(-\frac{1}{2}(\eta_i - \alpha - \beta\xi_i)^2/\sigma^2\right) \\ &\propto (\sigma^2)^{-N/2} \exp\left(-\frac{1}{2}\mathbf{SSR}/\sigma^2\right) \\ &= \textbf{Inv-}\chi^2(\sigma^2 | \, (N-2), \mathbf{SSR}/(N-2)) \end{aligned} \tag{16}$$

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 scaled inverse χ^2 distribution.

$$P(\{\xi_i, \eta_i\}, \alpha, \beta, \sigma^2, \mu, \tau^2 | \mathcal{D}) \propto$$
$$\Big[\prod_{i=1}^{N} N(x_i | \xi_i, \sigma_{x,i}^2) N(y_i, | \eta_i, \sigma_{y,i}^2) N(\eta_i | \alpha + \beta\xi_i, \sigma^2) N(\xi_i | \mu, \tau^2)\Big]$$
$$\times H(\sigma^2)\, H(\tau^2)$$

Update:

$$\text{(c) } \alpha \sim P(\alpha | \ldots; \mathcal{D}) = P(\alpha | \beta, \sigma^2, \{\eta_i, \xi_i\})$$

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

$$\begin{aligned} P(\alpha| \ldots \mathcal{D}) &\propto \prod_{i=1}^{N} N(\eta_i | \alpha + \beta\xi, \sigma^2) \\ &\propto \prod_{i=1}^{N} \exp\left(-\frac{1}{2}(\eta_i - \alpha - \beta\xi_i)^2/\sigma^2\right) \\ &\propto \prod_{i=1}^{N} \exp\left(-\frac{1}{2}(\alpha + \beta\xi_i - \eta_i)^2/\sigma^2\right) \\ &\propto \prod_{i=1}^{N} N(\alpha | \eta_i - \beta\xi_i, \sigma^2). \end{aligned} \tag{17}$$

$$P(\{\xi_i, \eta_i\}, \alpha, \beta, \sigma^2, \mu, \tau^2 | \mathcal{D}) \propto$$

$$\Big[\prod_{i=1}^{N} N(x_i | \xi_i, \sigma_{x,i}^2) N(y_i, | \eta_i, \sigma_{y,i}^2) N(\eta_i | \alpha + \beta\xi_i, \sigma^2) N(\xi_i | \mu, \tau^2)\Big]$$

$$\times H(\sigma^2)\, H(\tau^2)$$

$$\text{(c) } \alpha \sim P(\alpha | \ldots; \mathcal{D}) = P(\alpha | \beta, \sigma^2, \{\eta_i, \xi_i\})$$

$$\propto \prod_{i=1}^{N} N(\alpha | \eta_i - \beta\xi_i, \sigma^2)$$

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}).

$$P(\alpha | \ldots \mathcal{D}) = N\left(\alpha \middle| N^{-1} \sum_{i=1}^{N} (\eta_i - \beta\xi_i), \sigma^2/N\right) \tag{18}$$

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

$$P(\{\xi_i, \eta_i\}, \alpha, \beta, \sigma^2, \mu, \tau^2 | \, \mathcal{D}) \propto$$

$$\Big[\prod_{i=1}^{N} N(x_i | \, \xi_i, \sigma_{x,i}^2) N(y_i, | \eta_i, \sigma_{y,i}^2) N(\eta_i | \, \alpha + \beta\xi_i, \sigma^2) N(\xi_i | \, \mu, \tau^2)\Big]$$

$$\times \, H(\sigma^2) \, H(\tau^2)$$

Update: $\quad (\mathrm{d}) \; \beta \sim P(\beta | \; \ldots; \mathcal{D}) = P(\beta | \, \alpha, \sigma^2, \{\eta_i, \xi_i\})$

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

$$P(\beta | \; \ldots \mathcal{D}) \propto \prod_{i=1}^{N} N(\eta_i | \, \alpha + \beta\xi_i, \sigma^2)$$

$$\propto \prod_{i=1}^{N} \exp\left(-\frac{1}{2}(\eta_i - \alpha - \beta\xi_i)^2/\sigma^2\right)$$

$$\propto \prod_{i=1}^{N} \exp\left(-\frac{1}{2}(\beta + \alpha/\xi_i - \eta_i/\xi_i)^2/(\sigma/\xi_i)^2\right)$$

$$\propto \prod_{i=1}^{N} N(\beta | \, (\eta_i - \alpha)/\xi_i, \delta_i^2 \equiv \sigma^2/\xi_i^2)$$

$$P(\{\xi_i, \eta_i\}, \alpha, \beta, \sigma^2, \mu, \tau^2 | \mathcal{D}) \propto$$

$$\Big[\prod_{i=1}^{N} N(x_i | \xi_i, \sigma_{x,i}^2) N(y_i, | \eta_i, \sigma_{y,i}^2) N(\eta_i | \alpha + \beta\xi_i, \sigma^2) N(\xi_i | \mu, \tau^2)\Big]$$

$$\times H(\sigma^2)\, H(\tau^2)$$

$$\text{(d) } \beta \sim P(\beta | \ldots; \mathcal{D}) = P(\beta | \alpha, \sigma^2, \{\eta_i, \xi_i\})$$

$$\propto \prod_{i=1}^{N} N(\beta | (\eta_i - \alpha)/\xi_i, \delta_i^2 \equiv \sigma^2/\xi_i^2)$$

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

$$\bar{\beta} = \frac{\sum_{i=1}^{N} \delta_i^{-2} (\eta_i - \alpha)/\xi_i}{\sum_{i=1}^{N} \tau_i^{-2}} \tag{20}$$

$$V_\beta = \left(\sum_{i=1}^{N} \delta_i^{-2}\right)^{-1} \tag{21}$$

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

$$P(\{\xi_i, \eta_i\}, \alpha, \beta, \sigma^2, \mu, \tau^2 | \mathcal{D}) \propto$$

$$\Big[\prod_{i=1}^{N} N(x_i | \xi_i, \sigma_{x,i}^2) N(y_i, | \eta_i, \sigma_{y,i}^2) N(\eta_i | \alpha + \beta\xi_i, \sigma^2) N(\xi_i | \mu, \tau^2)\Big]$$

$$\times H(\sigma^2)\, H(\tau^2)$$

$$\text{(e) } \mu, \tau^2 \sim P(\mu, \tau^2 | \ \ldots; \mathcal{D}) = P(\mu | \tau^2 \{\xi_i\})\, P(\tau^2 | \{\xi_i\})$$

(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 1.3 of Example Sheet 2.

$$P(\mu, \tau^2 | \{\xi_i\}) \propto \prod_{i=1}^{N} N(\xi_i | \mu, \tau^2) \tag{22}$$

Using the same derivation, we can factor

$$\begin{aligned} P(\mu, \tau^2 | \{\xi_i\} &= P(\mu | \tau^2, \{\xi_i\}) \times P(\tau^2 | \{\xi_i\}) \\ &= N(\mu | \bar{\xi}, \tau^2/N) \times \textbf{Inv-}\chi^2 \left(\tau^2 \Big| (N-3), \frac{(N-1)}{(N-3)} S_\xi^2\right) \end{aligned} \tag{23}$$

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

$$S_\xi^2 = \frac{1}{N-1} \sum_{i=1}^{N} (\xi_i - \bar{\xi})^2. \tag{24}$$

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

$$\text{(e1) } \tau^2 \sim \text{Inv-}\chi^2 \left(\tau^2 \Big| (N-3), \frac{(N-1)}{(N-3)} S_\xi^2\right) \quad \text{(e2) } \mu | \tau^2 \sim N(\mu | \bar{\xi}, \tau^2/N)$$

Summary

Update Sequentially:

For every quasar i: $(\mathrm{a})_i\ (\xi_i, \eta_i) \sim P(\xi_i, \eta_i | \alpha, \beta, \sigma^2, \mu, \tau^2; \mathcal{D})$

Then update hyperparameters

$$(\mathrm{b})\ \sigma^2 \sim P(\sigma^2 | \ldots; \mathcal{D}) = P(\sigma^2 | \alpha, \beta, \{\eta_i, \xi_i\})$$

$$(\mathrm{c})\ \alpha \sim P(\alpha | \ldots; \mathcal{D}) = P(\alpha | \beta, \sigma^2, \{\eta_i, \xi_i\})$$

$$(\mathrm{d})\ \beta \sim P(\beta | \ldots; \mathcal{D}) = P(\beta | \alpha, \sigma^2, \{\eta_i, \xi_i\})$$

$$(\mathrm{e})\ \mu, \tau^2 \sim P(\mu, \tau^2 | \ldots; \mathcal{D}) = P(\mu | \tau^2 \{\xi_i\})\, P(\tau^2 | \{\xi_i\})$$

$$(\mathrm{e1})\ \tau^2 \sim P(\tau^2 | \{\xi_i\})$$

$$(\mathrm{e2})\ \mu | \tau^2 \sim P(\mu | \tau^2, \{\xi_i\})$$

Cycle through steps a-e many, many times through reaching convergence criteria and until reaching sufficient effective sample size

Implementation

- Initialisation
- Tuning?
- Assessing Convergence/Mixing - how long to run MCMC?
- Postprocessing
- Computing posterior inferences from MCMC samples

Bayesian Model Comparison

1. Data points $\{x_i\}$ come independently from a probability distribution $P(x)$. According to model H_0, $P(x)$ is a uniform distribution $P(x|H_0) = \frac{1}{2}$ for $x \in (-1, 1)$. According to model H_1, $P(x)$ is a nonuniform distribution with an unknown parameter $m \in (-1, 1)$:

$$P(x|\, m, H_1) = \frac{1}{2}(1 + m\, x), \tag{25}$$

for $x \in (-1, 1)$. Given the data $\mathcal{D} = \{0.3, 0.5, 0.7, 0.8, 0.9\}$, what is the evidence for H_0 and H_1?

- Strategy is to compute the evidence (marginal likelihood) under each model H:

$$Z_H = P(\boldsymbol{D}|H) = \int P(\boldsymbol{D}|\, \boldsymbol{\theta}, H) P(\boldsymbol{\theta}|H)\, d\boldsymbol{\theta}$$

Bayesian Model Comparison

1. Data points $\{x_i\}$ come independently from a probability distribution $P(x)$. According to model H_0, $P(x)$ is a uniform distribution $P(x|H_0) = \frac{1}{2}$ for $x \in (-1, 1)$. According to model H_1, $P(x)$ is a nonuniform distribution with an unknown parameter $m \in (-1, 1)$:

$$P(x|\, m, H_1) = \frac{1}{2}(1 + m\, x), \tag{25}$$

for $x \in (-1, 1)$. Given the data $\mathcal{D} = \{0.3, 0.5, 0.7, 0.8, 0.9\}$, what is the evidence for H_0 and H_1?

Solution: We note that all the observed data points are within the range of the sampling distribution.

Under model H_0, we have

$$P(\boldsymbol{x}|\, H_0) = \prod_{i=1}^{N} P(x_i|\, H_0) = \left(\frac{1}{2}\right)^N$$

(no parameters to integrate over)

Bayesian Model Comparison

1. Data points $\{x_i\}$ come independently from a probability distribution $P(x)$. According to model H_0, $P(x)$ is a uniform distribution $P(x|H_0) = \frac{1}{2}$ for $x \in (-1, 1)$. According to model H_1, $P(x)$ is a nonuniform distribution with an unknown parameter $m \in (-1, 1)$:

$$P(x|\, m, H_1) = \frac{1}{2}(1 + m\, x), \tag{25}$$

for $x \in (-1, 1)$. Given the data $\mathcal{D} = \{0.3, 0.5, 0.7, 0.8, 0.9\}$, what is the evidence for H_0 and H_1? **Under model H_1,**

$$\begin{aligned} P(\boldsymbol{x}|\, H_1) &= \int dm\, P(\boldsymbol{x}|\, m, H_1)\, P(m|\, H_1) \\ &= \int \left[\prod_{i=1}^{N} P(x_i|\, m, H_1) P(m|\, H_1) \right] dm \\ &= \left(\frac{1}{2}\right)^N \int_{-1}^{1} \prod_{i=1}^{N} (1 + m\, x_i)\, dm \\ &= \left(\frac{1}{2}\right)^4 \int_{-1}^{1} (1 + m\, x_1)(1 + m\, x_2)(1 + m\, x_3)(1 + m\, x_4)\, dm \\ &= \left(\frac{1}{2}\right)^4 \int_{-1}^{1} (1 + S_1 m + S_2 m^2 + S_3 m^3 + P m^4)\, dm \end{aligned}$$

$$S_2 \equiv x_1x_2 + x_1x_3 + x_1x_4 + x_2x_3 + x_2x_4 + x_3x_4 = 1.91$$

$$P \equiv x_1x_2x_3x_4 = 0.0756$$

(S_1, S_2 don't matter due to symmetry)

Bayesian Model Comparison

1. Data points $\{x_i\}$ come independently from a probability distribution $P(x)$. According to model H_0, $P(x)$ is a uniform distribution $P(x|H_0) = \frac{1}{2}$ for $x \in (-1, 1)$. According to model H_1, $P(x)$ is a nonuniform distribution with an unknown parameter $m \in (-1, 1)$:

$$P(x|\, m, H_1) = \frac{1}{2}(1 + m\, x), \tag{25}$$

for $x \in (-1, 1)$. Given the data $\mathcal{D} = \{0.3, 0.5, 0.7, 0.8, 0.9\}$, what is the evidence for H_0 and H_1?

Exploiting symmetries of the integrand, we find:

$$P(\boldsymbol{x}|\, H_1) = \left(\frac{1}{2}\right)^5 \int_{-1}^{1} (1 + S_1 m + S_2 m^2 + S_3 m^3 + P m^4)\, dm$$
$$= \frac{1}{16}\,[1 + S_2/3 + P/5]$$

where $S_2 \equiv x_1x_2 + x_1x_3 + x_1x_4 + x_2x_3 + x_2x_4 + x_3x_4 = 1.91$ **and** $P \equiv x_1x_2x_3x_4 = 0.0756$**. Thus** $P(\boldsymbol{x}|\, H_1) = 0.0883$ **and** $P(\boldsymbol{x}|\, H_0) = 0.0625$**, so the Bayes Factor** $B_{10} = P(\boldsymbol{x}|\, H_1)/P(\boldsymbol{x}|\, H_0) = 1.41$**, or** $\log B_{10} = 0.34$**. According to the Jeffrey's scale this is not significantly more evidence for model** H_1 **vs. model** H_0**.**

Bayesian Model Comparison 2

Datapoints $\mathcal{D} = \{x_i, y_i\}$ are believed to arisen from a straight line. The experimenter chooses x_i, and y_i is Gaussian-distributed around $w_0 + w_1 x_i$ with variance σ^2. According to model H_0, the straight line is horizontal, so $w_1 = 0$. According to model H_1, w_1 is a parameter with prior distribution $w_1 \sim N(0, \tau^2)$. Both models assign a prior distribution $w_0 \sim N(0, \tau^2)$.

(a) For each model, derive an expression for the posterior distribution of the regression parameter(s), $P(w_0|\,\mathcal{D}, H_0)$ and $P(w_0, w_1|\,\mathcal{D}, H_1)$.

(b) For each model, derive an expression for the evidence (or marginal likelihood).

Solution: For H_0, defining $\bar{y} = (\sum_{i=1}^{N} y_i)/N$, we have the posterior:

$$\begin{aligned} P(w_0|\mathcal{D}) &\propto \left[\prod_{i=1}^{N} N(y_i|w_0, \sigma^2)\right] N(w_0|\,0, \tau^2) \\ &\propto N(w_0|\,\bar{y}, \sigma^2/N) N(w_0, \tau^2) \\ &= N\left[w_0\Big|\, \frac{N\sigma^{-2}\,\bar{y}}{N\sigma^{-2} + \tau^{-2}}, \left(N\sigma^{-2} + \tau^{-2}\right)^{-1}\right]. \end{aligned}$$

Bayesian Model Comparison 2

Datapoints $\mathcal{D} = \{x_i, y_i\}$ are believed to arisen from a straight line. The experimenter chooses x_i, and y_i is Gaussian-distributed around $w_0 + w_1 x_i$ with variance σ^2. According to model H_0, the straight line is horizontal, so $w_1 = 0$. According to model H_1, w_1 is a parameter with prior distribution $w_1 \sim N(0, \tau^2)$. Both models assign a prior distribution $w_0 \sim N(0, \tau^2)$.

(a) For each model, derive an expression for the posterior distribution of the regression parameter(s), $P(w_0|\,\mathcal{D}, H_0)$ and $P(w_0, w_1|\,\mathcal{D}, H_1)$.

(b) For each model, derive an expression for the evidence (or marginal likelihood).

The evidence Z_0 can be computed by noting that the product of N independent 1D Gaussian sampling distributions can be re-expressed as a single N-dimensional multivariate Gaussian distribution with a diagonal covariance matrix.

Use MV Gaussian marginalisation

$$
\begin{aligned}
Z_0 &= \int \left[\prod_{i=1}^{N} N(y_i|w_0, \sigma^2) \right] N(w_0|\,0, \tau^2)\, dw_0 \\
&= \int N(\boldsymbol{y}|\mathbf{1}_N\, w_0, \sigma^2 \boldsymbol{I}_N) N(w_0|\,0, \tau^2)\, dw_0 \\
&= N(\boldsymbol{y}|\,\mathbf{0}, \boldsymbol{C}_0 = \sigma^2 \boldsymbol{I}_N + \mathbf{1}_N \mathbf{1}_N^T \tau^2),
\end{aligned}
$$

where $\mathbf{1}_N$ is an column vector of N ones, and $\boldsymbol{I}_N$ is the $N \times N$ dimensional identity matrix. A $N \times N$ square matrix filled with ones results from $\mathbf{1}_N \mathbf{1}_N^T$.

Bayesian Model Comparison 2

Datapoints $\mathcal{D} = \{x_i, y_i\}$ are believed to arisen from a straight line. The experimenter chooses x_i, and y_i is Gaussian-distributed around $w_0 + w_1 x_i$ with variance σ^2. According to model H_0, the straight line is horizontal, so $w_1 = 0$. According to model H_1, w_1 is a parameter with prior distribution $w_1 \sim N(0, \tau^2)$. Both models assign a prior distribution $w_0 \sim N(0, \tau^2)$.

(a) For each model, derive an expression for the posterior distribution of the regression parameter(s), $P(w_0|\,\mathcal{D}, H_0)$ and $P(w_0, w_1|\,\mathcal{D}, H_1)$.

(b) For each model, derive an expression for the evidence (or marginal likelihood).

For H_1, let $\boldsymbol{w} = (w_0, w_1)^T$ and $\boldsymbol{D} = (\mathbf{1}_N, \boldsymbol{x})$

Posterior:

$$P(\boldsymbol{w}|\mathcal{D}) \propto \left[\prod_{i=1}^{N} N(y_i|w_0 + w_1 x_i, \sigma^2)\right] N(w_0|\,0, \tau^2)$$

$$\propto N(\boldsymbol{y}|\,\boldsymbol{D}\,\boldsymbol{w}, \sigma^2 \boldsymbol{I}_N) N(\boldsymbol{w}|\,\mathbf{0}, \tau^2 \boldsymbol{I}_2)$$

Use result from OLS

$$\propto N(\boldsymbol{w}|\,(\boldsymbol{D}^T\boldsymbol{D})^{-1}\boldsymbol{D}^T\boldsymbol{y}, \sigma^2(\boldsymbol{D}^T\boldsymbol{D})^{-1}) \times N(\boldsymbol{w}|\,\mathbf{0}, \tau^2\boldsymbol{I}_2)$$

Product of Gaussian

$$\propto N(\boldsymbol{w}|\,\ldots, \ldots)$$

Bayesian Model Comparison 2

Datapoints $\mathcal{D} = \{x_i, y_i\}$ are believed to arisen from a straight line. The experimenter chooses x_i, and y_i is Gaussian-distributed around $w_0 + w_1 x_i$ with variance σ^2. According to model H_0, the straight line is horizontal, so $w_1 = 0$. According to model H_1, w_1 is a parameter with prior distribution $w_1 \sim N(0, \tau^2)$. Both models assign a prior distribution $w_0 \sim N(0, \tau^2)$.

(a) For each model, derive an expression for the posterior distribution of the regression parameter(s), $P(w_0|\,\mathcal{D}, H_0)$ and $P(w_0, w_1|\,\mathcal{D}, H_1)$.

(b) For each model, derive an expression for the evidence (or marginal likelihood).

The evidence Z_1 is

Using MV Gaussian marginalisation

$$Z_1 = \int N(\boldsymbol{y}|\,\boldsymbol{D}\,\boldsymbol{w}, \sigma^2 \boldsymbol{I}_N) N(\boldsymbol{w}|\,\boldsymbol{0}, \tau^2 \boldsymbol{I}_2)\, d\boldsymbol{w}$$

$$= N(\boldsymbol{y}|\,\boldsymbol{0}, \boldsymbol{C}_1 = \sigma^2 \boldsymbol{I}_N + \boldsymbol{D}\boldsymbol{D}^T)$$