Part III Astrostatistics 2020: 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:

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

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

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

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

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 $i = 1, \dots, N$ 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.

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 $\alpha, \beta, \sigma^2, \mu, \tau^2$.

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

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. A scaled inverse-χ^2 distribution with scale parameters s and degrees of freedom ν has a density function:

$$\text{Inv-}\chi^2(x|\, \nu, s^2) = \frac{(\nu/2)^{\nu/2}}{\Gamma(\nu/2)} s^\nu\, x^{-(\nu/2+1)} e^{-\nu s^2/(2x)}, \tag{5}$$

   for $x > 0$, and is zero otherwise.

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

6. Implement your Gibbs sampler in code and apply it to analyse the data from Example Sheet 2, Problem 2.

2 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 $m \in (-1, 1)$:

$$P(x|\, m, H_1) = \frac{1}{2}(1 + m\, x), \tag{6}$$

   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? Compute the relative Bayes factor of H_1 relative to H_0.

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).

   (c) Consider three datasets. In each dataset, $\boldsymbol{x} = (-8, -2, 6)$.

      i. In dataset A, $\boldsymbol{y} = (-6.1, -1.0, 5.5)$.

      ii. In dataset B, $\boldsymbol{y} = (2.7, 1.8, 1.51$.

      iii. In dataset C, $\boldsymbol{y} = (7.9, 10.1, 11.0)$

      For each dataset, compute the posterior mean and standard deviation of the parameters $\boldsymbol{w}$, and compute the log Bayes factor of H_1 relative to H_0, for $\sigma = 1$ and $\tau = 1$. Plot the dependence of the posterior estimates and the log Bayes factor on $\log_{10} \tau$ ranging from -2 to 2. Interpret what you find.

3. Consider Bayesian model selection over p-order polynomials. Under hypothesis H_p, the model for the data is:

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

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

   (b) Analyse the following dataset:

$$\boldsymbol{x} = (-2.0, -1.5, 1.0, -0.5, 0, 0.5, 1.0, 1.5, 2.0) \tag{8}$$

$$\boldsymbol{y} = (-5.76, -1.54, -0.55, 0.29, -0.90, -1.21, -0.95, 2.80, 6.43) \tag{9}$$

   (c) 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.

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 $k(x, x') = \mathrm{Cov}[f(x), f(x')]$?
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 $c_1 \to -\infty$, $h \to \infty$.
3. What is the variance of the resulting Gaussian process at any x?