Part III Astrostatistics: Example Sheet 4
Example Class: Thursday, 26 Apr 2018, 1:00pm, MR5

1 Combining Uncertain Estimates

Suppose two Type Ia supernovae are observed in a single external galaxy. Since the distance from Earth to the galaxy is so much greater than the size of the galaxy, we can assume that the two supernovae, and the galaxy, are all at nearly the same distance from Earth. Analysis of the supernovae yield unbiased distance estimates $\hat{d}_1$ and $\hat{d}_2$, which have Gaussian error with known standard deviations σ_1 and σ_2.

1. Consider all estimators that are linear combinations of the two distances: $\hat{d} = \alpha_1 \hat{d}_1 + \alpha_2 \hat{d}_2$. What relation between the coefficients is required of the subset of unbiased estimators?
2. Find the unbiased linear estimator with minimum variance. What is the variance of this estimator?
3. Suppose the standard deviations are the same $\sigma_1 = \sigma_2 = \sigma$. What is the minimum variance of the combined estimate?
4. Suppose because of observational systematic errors, the distance errors are correlated with correlation coefficient ρ, such that $|\rho| < 1$. What is the minimum variance unbiased linear estimator in this case?

2 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,

$$\phi_i(x) = \exp\left[-\frac{(x - c_i)^2}{l^2} \right] \tag{1}$$

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:

$$f(x) = \sum_{i=1}^{J} w_i \, \phi_i(x). \tag{2}$$

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

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?

3 Periodic Gaussian Processes

Many astronomical time-domain phenomena exhibit periodic signals (e.g. variable stars or exoplanet transits). Consider the zero-mean Gaussian process on the plane $\boldsymbol{x} \in \mathbb{R}^2$ with the squared exponential kernel: $f(\boldsymbol{x}) \sim \mathcal{GP}(\boldsymbol{0}, k(\boldsymbol{x}, \boldsymbol{x}'))$:

$$k(\boldsymbol{x}, \boldsymbol{x}') = A^2 \exp\left(-\frac{|\boldsymbol{x} - \boldsymbol{x}'|^2}{2l^2}\right). \tag{3}$$

Now consider the process $g(t) = f(\boldsymbol{u}(t))$ restricted to the circle:

$$\boldsymbol{u}(t) = \left(r \sin \frac{2\pi t}{T}, r \cos \frac{2\pi t}{T}\right). \tag{4}$$

1. Derive the covariance $k(t, t')$ between $g(t)$ and $g(t')$. Show that the Gaussian process on the circle is stationary.

2. What is the period of functions drawn from this GP? Justify.

3. (*non-exam*). Draw random functions $g(t) \sim \mathcal{GP}(\boldsymbol{0}, k(t, t'))$ to verify their period and explore their behaviour as you vary $\tilde{l} = l/r$ from small to large.

4 Probabilistic Graphical Model and Gibbs Sampling for the Normal-Normal Hierarchical Model

Consider our Normal-Normal hierarchical Bayesian model for supernova magnitudes. The true, latent absolute magnitudes are drawn from a Gaussian distribution with the population mean μ and variance τ^2 as hyperparameters:

$$M_s | \, \mu, \tau^2 \sim N(\mu, \tau^2). \tag{5}$$

Each absolute magnitude is observed to yield the data D_s with measurement error with known variance σ_s^2.

$$D_s | \, M_s \sim N(M_s, \sigma_s^2) \tag{6}$$

Suppose we observe $s = 1, \dots, N_{\mathrm{SN}}$ independent supernovae.

1. For a single supernova s, write down the joint probability density of the datum D_s and the latent variable M_s, conditional on the hyperparameters μ, τ^2.

2. For all the N_{SN} supernovae, write down the joint probability density of all the data $\mathcal{D} = \{D_s\}$ and the latent variables $\{M_s\}$ given the hyperparameters.

3. Adopt a "non-informative" hyperpriors on the hyperparameters $P(\mu, \tau^2) \propto 1$, $\tau^2 > 0$. Write down the joint probability density of all data $\mathcal{D}$, latent variables $\{M_s\}$, and hyperparameters μ, τ^2.

4. Draw a probabilistic graphical model or directed acyclic graph representing this joint probability density.

5. Construct a Gibbs sampling algorithm to sample from the full posterior $P(\{M_s\}, \mu, \tau^2 | \mathcal{D})$ by deriving a complete set of tractable conditional posterior densities. You may choose to draw "blocks" or parameter subsets from their conditionals in a single step, rather than drawing a single parameter in each step. You may assume that you have access to algorithms that generate random draws from Gaussian distributions, and inverse-gamma distributions with shape parameter a and scale parameter b:

$$\text{Inv-Gamma}(x | a, b) \propto x^{-(a+1)} \exp(-b/x), x > 0. \tag{7}$$

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

7. (*non-exam*) Implement your Gibbs sampler in code and apply it to analyse the data from Example Sheet 1, Problem 1.

5 PGM and Gibbs Sampling for the Hierarchical Linear Regression Model

Consider 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:

$$\xi_i | \mu, \tau^2 \sim N(\mu, \tau^2) \tag{8}$$

$$\eta_i | \xi_i; \alpha, \beta, \sigma^2 \sim N(\alpha + \beta\xi_i, \sigma^2) \tag{9}$$

$$x_i | \xi_i \sim N(\xi_i, \sigma^2_{x,i}) \tag{10}$$

$$y_i | \eta_i \sim N(\eta_i, \sigma^2_{y,i}) \tag{11}$$

The astronomer observes values $\mathcal{D} = \{x_i, y_i\}$ with heteroskedastic measurement error of known variances $\{\sigma^2_{x,i}, \sigma^2_{y,i}\}$, 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.

2. Adopt "non-informative" hyperpriors on the hyperparameters $P(\alpha, \beta, \sigma^2) \propto 1, \sigma^2 > 0$ and $P(\mu, \tau^2) \propto 1, \tau^2 > 0$. 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 / directed acyclic graph 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 complete set of conditional posterior densities. You may choose to draw "blocks" or parameter subsets from their conditionals in a single step, rather than drawing a single parameter in each step. You may assume that you have access to algorithms that generate random draws from Gaussian distributions, and inverse-gamma distributions with shape parameter a and scale parameter b:

$$\text{Inv-Gamma}(x | a, b) \propto x^{-(a+1)} \exp(-b/x), x > 0. \tag{12}$$

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

6. (*non-exam*) Implement your Gibbs sampler in code and apply it to analyse the data from Example Sheet 2, Problem 1.

6 Markov Chain Monte Carlo

Consider a general Bayesian inference of unknown parameters $\boldsymbol{\theta}$ from data $\boldsymbol{D}$. You have a likelihood function $L(\boldsymbol{\theta}) = P(\boldsymbol{D}|\,\boldsymbol{\theta})$ and a proper prior $P(\boldsymbol{\theta})$.

1. Construct an MCMC algorithm that samples from the posterior density $P(\boldsymbol{\theta}|\,\boldsymbol{D}) \propto L(\boldsymbol{\theta})P(\boldsymbol{\theta})$ in the long-run.

2. Show that the posterior $P(\boldsymbol{\theta}|\,\boldsymbol{D})$ is the stationary distribution of the chain. That is, if, at step $i-1$, $\boldsymbol{\theta}_{i-1} \sim P(\boldsymbol{\theta}|\,\boldsymbol{D})$ is drawn from the posterior density, then so will be $\boldsymbol{\theta}_i$ after a complete iteration of the chain.

3. Describe how you would implement this algorithm, diagnose convergence, and prepare the output for analysis.