

# Part III Astrostatistics: Example Sheet 4

## Example Class: Tuesday, 30 April 2019 (To Be Confirmed)

### 1 Warm-Up

Prove that the product of  $m$  multivariate Gaussian densities in random  $d$ -dimensional vector  $\mathbf{x}$ :

$$\prod_{i=1}^m N(\mathbf{x} | \boldsymbol{\mu}_i, \mathbf{C}_i) \quad (1)$$

is proportional to a single Gaussian density in  $\mathbf{x}$ . Here the  $\{\boldsymbol{\mu}_i\}$  and  $\{\mathbf{C}_i\}$  are  $m$  pairs of constant mean  $d$ -vectors and  $d \times d$  covariance matrices. Find the mean and covariance matrix of the single resulting Gaussian.

### 2 A Hierarchical Bayesian Model for Supernovae and Dust

Consider the following hierarchical Bayesian generative model for supernova colours. The latent intrinsic colour of a supernova  $s$  is  $C_s$  and is drawn from a Gaussian distribution with mean colour  $\mu_C$  and variance  $\sigma_{\text{int}}^2$ :  $C_s \sim N(\mu_C, \sigma_{\text{int}}^2)$ . The latent reddening due to interstellar dust in the supernova's galaxy is  $E_s$ , and is drawn from an exponential distribution with mean  $\tau$ :  $E_s \sim \text{Exponen}(\tau)$ , i.e.,

$$P(E_s | \tau) = \tau^{-1} \exp(-E_s/\tau) \times H(E_s), \quad (2)$$

where  $H(x)$  is the Heaviside step function:

$$H(x) = \begin{cases} 1, & x > 0, \\ 0, & x \leq 0. \end{cases} \quad (3)$$

Because the presence of interstellar dust can only redden the colours, the probability density is only positive for  $E_s > 0$ . The measured, observed colour  $\hat{O}_s$  results from the sum of the intrinsic colour, reddening, and measurement error with known variance  $\sigma_{O,s}^2$ :  $\hat{O}_s | E_s, C_s \sim N(C_s + E_s, \sigma_{O,s}^2)$ . There are  $s = 1, \dots, N$  independent supernovae in our sample. For hyperpriors, you may use the improper, noninformative  $P(\mu_C) \propto 1$ ,  $P(\tau) \propto H(\tau)$ , and  $P(\sigma_{\text{int}}^2) \propto H(\sigma_{\text{int}}^2)$ .

1. Write down the joint probability distribution of the observed data  $\{\hat{O}_s\}$ , latent variables  $\{C_s, E_s\}$ , and hyperparameters  $\mu_C, \sigma_{\text{int}}^2, \tau$  for the sample of  $N$  supernovae.
2. Draw a probabilistic graphical model or directed acyclic graph representing this joint distribution.

3. Construct a Gibbs sampler that generates an MCMC to sample the joint posterior probability density of the unknown latent variables and hyperparameters given the observed colours,  $P(\{C_s, E_s\}, \mu_C, \sigma_{\text{int}}^2, \tau | \{\hat{O}_s\})$ , by deriving the  $2N + 3$  conditional posterior densities that one can directly sample from. You may assume that you have access to algorithms that allow you to directly sample random variates from the following probability densities:
  - (a) Gaussian  $N(x | \mu, \sigma^2)$ .
  - (b) truncated Gaussian  $\propto H(x) \times N(x | \mu, \sigma^2)$ .
  - (c) Inverse gamma: Inv-Gamma( $x | a, b$ )  $\propto x^{-(a+1)} \exp(-b/x)$ ,  $x > 0$ .

Briefly describe how you would implement the sampler, and analyse and assess the convergence of the MCMC.

4. Apply your sampler to analyse the data from the Table 3 dataset from Jha, Riess & Kirshner. (2007), “Improved Distances to Type Ia Supernovae with Multicolor Light-Curve Shapes: MLCS2k2.” Compute posterior summaries of the hyperparameters  $\mu_C, \sigma_{\text{int}}^2, \tau$  and the latent variables  $\{C_s, E_s\}$ . The data is provided online.

### 3 Harmonic Mean Estimator for Bayesian Evidence

Consider a general Bayesian inference problem with data  $y$ , parameter  $\theta$ , likelihood function  $P(y | \theta)$  and a proper prior  $P(\theta)$ . We wish to compute the evidence or marginal likelihood  $Z \equiv P(y) = \int P(y | \theta) P(\theta) d\theta$ .

1. Suppose you have  $i = 1, \dots, m$  independent, random samples from the posterior distribution,  $\theta_i \sim P(\theta | y)$ . Consider the estimator

$$\hat{I} \equiv \frac{1}{m} \sum_{i=1}^m P(y | \theta)^{-1}. \quad (4)$$

Show that  $\mathbb{E}_{\theta|y}[\hat{I}] = Z^{-1}$ , where the expectation is taken with respect to the posterior density  $P(\theta | y)$ . Thus, by LLN, as  $m \rightarrow \infty$ , this converges to the inverse of the evidence.

2. Suppose the sampling distribution of the data is  $y \sim N(\theta, \sigma^2)$  and the proper prior is  $\theta \sim N(0, \tau^2)$ . The measurement variance  $\sigma^2$  and the prior variance  $\tau^2$  are known, and  $\tau \gg \sigma$ .
  - (a) What is the posterior density  $P(\theta | y)$ ?
  - (b) What is the exact evidence  $Z$ ?
  - (c) What is the expectation of the estimator  $\mathbb{E}_{\theta|y}[\hat{I}]$ ?
  - (d) What is the variance of the estimator  $\text{Var}_{\theta|y}[\hat{I}]$ ?

### 4 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  $m \in (-1, 1)$ :

$$P(x | m, H_1) = \frac{1}{2}(1 + mx), \quad (5)$$

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  $H_1$ ?

2. Datapoints  $\{(x_i, t_i)\}$  are believed to come from a straight line. The experimenter chooses  $x_i$ , and  $t_i$  is Gaussian-distributed about  $y_i = w_0 + w_1 x_i$  with variance  $\sigma^2$ . According to model  $H_1$ , the straight line is horizontal, so  $w_1 = 0$ . According to model  $H_2$ ,  $w_1$  is a parameter with prior distribution  $\text{Normal}(0, 1)$ . Both models assign a prior distribution  $\text{Normal}(0, 1)$  to  $w_0$ . Given the data set  $D = \{(-8, 8), (-2, 10), (6, 11)\}$ , and assuming the noise level is  $\sigma = 1$ , what is the evidence for each model?

## 5 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  $c_i$ ,

$$\phi_i(x) = \exp \left[ -\frac{(x - c_i)^2}{l^2} \right] \quad (6)$$

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:

$$f(x) = \sum_{i=1}^J w_i \phi_i(x). \quad (7)$$

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:  $J \rightarrow \infty$  and  $c_1 \rightarrow -\infty$ ,  $h \rightarrow \infty$ .
3. What is the variance of the resulting Gaussian process at any  $x$ ?