UNIVERSITY OF CAMBRIDGE

MAT3, MAMA, NST3AS

MATHEMATICAL TRIPOS Part III

Monday, 10 June, 2019 1:30 pm to 4:30 pm

PAPER 219

ASTROSTATISTICS

*Attempt no more than **THREE** questions.*

*There are **FOUR** questions in total.*

The questions carry equal weight.

STATIONERY REQUIREMENTS
Cover sheet
Treasury Tag
Script paper
Rough paper

SPECIAL REQUIREMENTS
None

You may not start to read the questions printed on the subsequent pages until instructed to do so by the Invigilator.

1 Type Ia supernovae can be observed at great distances, and are used to estimate the current expansion rate of the Universe, the Hubble constant H_0. Suppose the peak absolute magnitudes of Type Ia supernovae are independent draws from an intrinsic Gaussian distribution with population mean M_0 and variance $\sigma^2_{\rm int}$:

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

for every supernova s. The true absolute magnitude is related to the true apparent magnitude m_s via the true distance modulus μ_s, which is a logarithmic measure of the true distance d_s:

$$m_s = M_s + \mu_s.$$

The definition of the distance modulus is $\mu = 25 + 5\log_{10}[d \text{ Mpc}^{-1}]$, where Mpc is a megaparsec. For every supernova s, we obtain an estimate of its peak apparent magnitude $\hat{m}_s$ with Gaussian error of known variance $\sigma^2_{m,s}$, so that

$$\hat{m}_s | \, m_s \sim N(m_s, \sigma^2_{m,s}).$$

Suppose we have a calibrator set of $k = 1, \ldots, K$ supernovae located in nearby galaxies in which we can observe Cepheid variable stars. For each calibrator supernova, we have an unbiased distance modulus estimate $\hat{\mu}_{C,k}$ with a Gaussian error with variance $\sigma^2_{C,k}$, obtained from analyses of the Cepheid stars as distance indicators in the same galaxy:

$$\hat{\mu}_{C,k} | \, \mu_k \sim N(\mu_k, \sigma^2_{C,k}).$$

We also have a much larger ("Hubble Flow") set of $i = 1, \ldots, N$ supernovae which are much further away, so the Cepheids stars cannot be observed in their galaxies. However, they are far enough away that they participate in the smooth, overall expansion of the Universe. Thus, they follow the Hubble law, the linear relation between their recession velocities $v_i = c\, z_i$ and their distances d_i: $d_i = cz_i/H_0$, where c is the speed of light and z_i is the redshift. Assume the redshift is measured exactly for each supernova i. The units of the Hubble constant are km s^{-1} Mpc^{-1}. Define $h \equiv H_0/(100 \text{ km s}^{-1} \text{ Mpc}^{-1})$, $\theta \equiv 5\log_{10} h$, and $\alpha \equiv 5/\ln 10 \approx 2$. Assume all errors are independent.

(a) Write down the likelihood function of (M_0, θ) in terms of the data of the calibrator set $\{\hat{m}_k, \hat{\mu}_{C,k}\}$ and the Hubble flow sample $\{\hat{m}_i, z_i\}$, and the relevant variances.

(b) Assume all error variances and the intrinsic dispersion $\sigma^2_{\rm int}$ are known. Derive the maximum likelihood estimators for (M_0, θ), checking 1st- and 2nd-order conditions.

(c) Simplify for the case where each source of measurement error is homoskedastic, i.e. $\sigma_{C,k} = \sigma_C$ for all calibrators, and $\sigma_{m,s} = \sigma_m$ for all supernovae. Evaluate the bias and variance of your estimators $(\hat{M}_0, \hat{\theta})$, and compare against the Cramér-Rao bound.

(d) As in part (c), assume that $\sigma_{C,k} = \sigma_C$ for all calibrators, and $\sigma_{m,s} = \sigma_m$ for all supernovae. What is the maximum likelihood estimator $\hat{h}$ for h? What is the sampling distribution of $\hat{h}$? Derive approximations for the fractional bias $(\mathbb{E}[\hat{h}] - h)/h$ and the fractional variance $\text{Var}[\hat{h}]/h^2$ to leading order in $\sigma^2_{\hat{\theta}} \equiv \text{Var}[\hat{\theta}]$.

2 The physics of galaxy formation often produces complex galactic systems, each composed of a massive, central host galaxy surrounded by several dwarf galaxies with different dynamical properties, such as the true angular momentum. We want to infer the mass of our Milky Way (MW) galaxy using measurements of the dynamical properties of its satellite dwarf galaxies. For each observed satellite s, we have a measurement d_s of (the magnitude of) its angular momentum x_s. We also have a catalog of K satellite dwarf galaxies found around massive host galaxies in a cosmological simulation. Each row i in this catalog gives the true angular momentum x_i of a satellite i, and the log mass, $m_i = \log_{10} M_i$, of its associated massive host galaxy. Each entry can be thought of as a random draw (x_i, m_i) from a prior distribution $P(x, m)$ that encodes the correlation between the true angular momentum and host galaxy log mass induced by the physics of galaxy formation encoded in the simulation. Assume the MW galaxy and each of its satellites is a representative draw from this distribution.

(a) First, consider a single satellite of the MW, the Large Magellanic Cloud. The measurement d is an unbiased estimate of x, but its error is Gaussian with known variance σ^2. Write down the likelihood function $P(d\,|\,x)$. Write down an expression for the normalised posterior probability density of the MW log mass, $P(m|\,d)$, and the posterior mean $\bar{m}$, conditional on this single satellite.

(b) Derive an estimate of $\bar{m} \approx \sum_{i=1}^{K} m_i\, w_i$ and specify the importance weights w_i.

(c) Because the importance weights are unequal, not all prior samples contribute equally to the posterior estimate. A measure of *effective sample size* is $\mathrm{ESS} = K/[1+\mathrm{CV}^2(w)]$, where the squared coefficient of variation is defined as $\mathrm{CV}^2(w) = \mathrm{Var}[w]/(\mathbb{E}[w])^2$, and the weight w is regarded as a random variable. Show that the estimate $\widehat{\mathrm{ESS}}$ is only a function of $\sum_{i=1}^{K} w_i^2$, and derive the function.

(d) The MW galaxy has $N_{\mathrm{sat}} = 9$ classical dwarf satellite galaxies. Derive an expression for estimating the posterior mean MW log mass, simultaneously conditioning on the independent measurements $\boldsymbol{d} = \{d_s\}$, $s = 1, \ldots, N_{\mathrm{sat}}$, of all of these satellites (each with known measurement variance σ_s^2). You may assume that the true angular momenta of the satellites are conditionally independent from each other, given the host galaxy's log mass $P(\{x_s\}|\,m) = \prod_{s=1}^{N_{\mathrm{sat}}} P(x_s|\,m)$. You may use the catalog simulation samples to construct kernel density estimates of the marginal prior density $P(m)$, and the posterior densities $P(m|\,d_s)$, conditioning on each satellite s individually, without specifying the optimal bandwidths.

(e) Suppose that we can compute $P(m|\,\boldsymbol{d})$ (with proper normalisation) but cannot directly sample from it, and we want to compute the posterior mean using importance sampling. Prove that the optimal importance function to sample for approximating the posterior mean of m, in the sense of minimum variance, is

$$Q^*(m) = \frac{|m|\, P(m|\,\boldsymbol{d})}{\int |m|\, P(m|\,\boldsymbol{d})\, dm}.$$

(You may use Jensen's Inequality: $\mathbb{E}[g(X)] \geqslant g(\mathbb{E}[X])$ where g is a convex function and X is a generic random variable.) Explain why this is not likely to be a useful importance function.

3 Quasar light curves (brightness time series) $f(t)$ (in magnitudes) are often modelled as realisations of an Orstein-Uhlenbeck (O-U) process, also called a damped random walk. The O-U process is a Gaussian process (GP):

$$f(t) \sim \mathcal{GP}(m(t), k_f(t,t')),$$

with prior mean function $m(t) = c$ and covariance function or kernel

$$\mathrm{Cov}[f(t), f(t')] = k_f(t,t') = A_f^2 \exp(-|t-t'|/\tau_f)$$

with characteristic amplitude $A_f^2 = \tau_f \sigma^2/2$. The short-term variability of the process is controlled by σ^2, and the characteristic timescale for the quasar brightness to revert to the mean c is τ_f. In a doubly-lensed quasar system, a galaxy, along the line-of-sight between the quasar and Earth, acts as a strong gravitational lens and produces two observed images of the same quasar. However, the brightness time series of image 2 will have a constant time delay (horizontal shift) Δt and constant magnification (vertical shift) Δm relative to image 1 due to strong lensing effects. Each image's light curve is further smoothly modulated by time-dependent *microlensing* magnification functions $g_i(t)$, caused by the movement of stars within the lens, which we can model as two ($\{i = 1, 2\}$) independent realisations of a squared-exponential GP:

$$g_i(t) \sim \mathcal{GP}(0, k_g(t,t'))$$

with prior mean zero and covariance function:

$$\mathrm{Cov}[g_i(t), g_i(t')] = k_g(t,t') = A_g^2 \exp(-(t-t')^2/2\tau_g),$$

with known hyperparameters A_g and τ_g. These functions $g_1(t)$ and $g_2(t)$ are independent from each other and $f(t)$. An astronomer measures the brightness time series $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ of two images of a lensed quasar at times $\boldsymbol{t}$, so that

$$y_{1,j} = f(t_j) + g_1(t_j) + \epsilon_{1,j}$$

$$y_{2,j} = f(t_j - \Delta t) + \Delta m + g_2(t_j) + \epsilon_{2,j}$$

for $j = 1, \ldots, N$ observation times. Assume the heteroskedastic measurement errors $\epsilon_{i,j}$ are independent, zero-mean Gaussian random variables with known variances, $\sigma_{i,j}^2$.

(a) Derive the marginal likelihood function $P(\boldsymbol{y}_1, \boldsymbol{y}_2 | \boldsymbol{\theta})$ of the time delay, magnification, and O-U process parameters $\boldsymbol{\theta} = (\Delta t, \Delta m, c, A_f, \tau_f)$, with the two time series $\boldsymbol{y}_1, \boldsymbol{y}_2$ considered jointly.

(b) Suppose you have optimised this marginal likelihood to find point estimates $\hat{\boldsymbol{\theta}} = (\Delta\hat{t}, \Delta\hat{m}, \hat{c}, \hat{A}_f, \hat{\tau}_f)$. Treating these as perfectly known, now you want to infer the underlying light curve $f(t)$ on a fine grid of times $\boldsymbol{t}^*$, when the quasars were not observed. Derive expressions for computing the posterior mean and variance of $f(t)$ at all times $\boldsymbol{t}^*$, conditional on the observed data $\boldsymbol{y}_1, \boldsymbol{y}_2$ and times $\boldsymbol{t}$.

(c) Using suitable non-informative priors, write down a posterior density $P(\boldsymbol{\theta} | \boldsymbol{y}_1, \boldsymbol{y}_2)$. Describe an MCMC algorithm to sample from this posterior density. Describe how you would initialise, run and evaluate the MCMC.

(d) Prove that your MCMC algorithm respects detailed balance with a stationary distribution equal to the posterior distribution.

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

$$\xi_i |\, \mu, \tau^2 \sim N(\mu, \tau^2)$$

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

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

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

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

(a) 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, conditional on the hyperparameters.

(b) 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 $\alpha, \beta, \sigma^2, \mu, \tau^2$.

(c) Draw a probabilistic graphical model representing this joint distribution.

(d) Construct a Gibbs sampler for the full joint 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 inverse-gamma distributions with shape parameter $a > 0$ and scale parameter $b > 0$. An inverse-gamma probability density is:

$$\text{Inv-Gamma}(x \,|\, a, b) = \frac{b^a}{\Gamma(a)} x^{-(a+1)} \exp(-b/x),$$

for $x > 0$, and zero otherwise, and $\Gamma(a)$ is a gamma function.

END OF PAPER