UNIVERSITY OF CAMBRIDGE

MATHEMATICAL TRIPOS Part III

Monday, 13 June, 2022 9:00 am to 12:00 pm

PAPER 219

ASTROSTATISTICS

Before you begin please read these instructions carefully

Candidates have THREE HOURS to complete the written examination.

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 are bright stellar explosions used to measure astronomical distances and the expansion rate of the Universe. Siblings are two or more supernovae that occur in the same galaxy, and thus have the same true distance from Earth. Because they have a common astrophysical origin, we might expect supernova siblings within the same galaxy to be more similar to each other than to supernovae occurring in other galaxies. Suppose the peak absolute magnitude M_s of a supernova s in galaxy $g(s)$ can be modelled as,

$$M_s = M_0 + \Delta M_{g(s)} + \delta M_s,$$

where M_0 is the mean absolute magnitude, $\Delta M_{g(s)} \sim N(0, \sigma_{\mathrm{G}}^2)$ is a random magnitude fluctuation that has the same value for all supernovae s in the same galaxy $g(s)$, and $\delta M_s \sim N(0, \sigma_{\mathrm{I}}^2)$ are random magnitude fluctuations that are independent between supernovae in the same galaxy. Both $\Delta M_{g(s)}$ and δM_s are independent between supernovae in different galaxies. Suppose the variances σ_{G}^2 and σ_{I}^2 are known.

The true absolute magnitude M_s 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 $\mu_s = 25 + 5\log_{10}[d_s \text{ Mpc}^{-1}]$, where Mpc is a megaparsec. Assume astronomers observe the apparent magnitude m_s of each supernova s with negligible measurement error.

Astronomers observe $K \geqslant 2$ sibling supernovae, labelled $k = 1, \ldots, K$, in a single nearby galaxy, in which they also observe Cepheid stars to use as independent distance indicators. The analysis of the Cepheid stars yields an unbiased measurement $\hat{\mu}$ of this galaxy's true distance modulus μ with Gaussian measurement error with variance σ_μ^2.

They also observe a much larger ("Hubble Flow") set of N supernovae, labelled $i = 1, \ldots, N$, 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. Each supernova i in this set follows the Hubble law, the linear relation between their recession velocities $v_i = c\, z_i$ and their distances d_i: $d_i = c\, z_i / H_0$, where c is the speed of light and z_i is the redshift. Assume the redshift is measured exactly for each supernova in this set. In this set, only one supernova is observed in each galaxy, and every supernova is independent. The units of the Hubble constant H_0 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 $\alpha = 5/\ln 10$. In each part below, show all steps.

(a) Derive the covariance and correlation between the absolute magnitudes M_s of any two siblings (s, s') in the calibrator galaxy. What is the total variance $\sigma_{\mathrm{tot}}^2 = \mathrm{Var}[M_s]$ for any supernova s?

(b) Derive the likelihood function $L(M_0, \theta)$ using all the data of the calibrator set $\{m_k\}, \hat{\mu}$ and the Hubble Flow set $\{m_i, z_i\}$.

(c) Derive the maximum likelihood estimators $\hat{M}_0, \hat{\theta}$. Derive the bias and variance of each, and compare $\mathrm{Var}[\hat{\theta}]$ to the Cramér-Rao bound.

(d) What is the maximum likelihood estimator $\hat{h}$ for h? Approximate the fractional variance $\mathrm{Var}[\hat{h}/h]$ to lowest order in $\sigma_\theta^2 \equiv \mathrm{Var}[\hat{\theta}]$. In which of the following cases is the fractional variance smaller: (i) $\sigma_{\mathrm{G}} = \sigma_{\mathrm{tot}}$, $\sigma_{\mathrm{I}} = 0$, or (ii) $\sigma_{\mathrm{G}} = 0$, $\sigma_{\mathrm{I}} = \sigma_{\mathrm{tot}}$?

2 Consider a quasar whose stochastic brightness over time $y(t)$ can be modelled as a realisation of a Gaussian process,

$$y(t) \sim \mathcal{GP}(\mu, k(t, t')),$$

with prior mean level μ and a symmetric, stationary covariance kernel $k(t, t')$. An astronomer has measured the brightness of the quasar at times t_1 and $t_2 > t_1$, with negligible measurement error, yielding $y_1 \equiv y(t_1)$ and $y_2 \equiv y(t_2)$. Denote the kernel values for indexed times as $R_{ij} \equiv k(t_i, t_j)$ and let $R \equiv k(0, 0)$. In all parts below, show all steps.

(a) We wish to predict the quasar's brightness $y_3 = y(t_3)$ at a future third time $t_3 > t_2$. Under what condition(s) is y_3 conditionally independent of y_1 given y_2? Under this condition(s), derive and fully simplify the posterior predictive distribution $P(y_3 | \, y_2, y_1)$, and posterior predictive mean and variance. Explicitly demonstrate that they are independent of y_1 and t_1.

(b) Henceforth, consider covariance kernels of the form,

$$k_\tau(t, t') = \exp(-|t - t'|^\eta / \tau^\eta),$$

where $\tau > 0$ is a characteristic timescale. For which value of η does the conditional independence in part (a) hold? Justify your answer. For this value of η and covariance kernel, derive and fully simplify explicit expressions for the posterior predictive mean and variance of y_3 given the observed data. Derive their limiting values as $t_3 \to \infty$.

(c) An astronomer now additionally observes $y_3 = y(t_3)$ without measurement error. For the value of η found in part (b), derive and fully simplify an expression for the likelihood function $P(\boldsymbol{y} | \, \boldsymbol{t}, \mu, \tau)$, where $\boldsymbol{y} = (y_1, y_2, y_3)^T$ and $\boldsymbol{t} = (t_1, t_2, t_3)^T$, in the form of a product of three univariate probability densities. Assuming τ is known, derive and fully simplify the maximum likelihood estimator $\hat{\mu}$ for μ. Is $\hat{\mu}$ unbiased? Justify your answer.

3 Suppose stars of a particular type have intrinsic absolute magnitudes and colours that are drawn from a Gaussian population distribution. For a random star s, its intrinsic absolute magnitude M_s and intrinsic colour c_s are drawn from a correlated population distribution $P(M_s, c_s)$, which can be expressed conditionally as:

$$c_s \sim N(c_0, \sigma_c^2), \qquad M_s|\, c_s \sim N(M_0 + \beta\, c_s, \sigma_M^2),$$

where $M_0, c_0, \beta, \sigma_M^2$, and σ_c^2 are hyperparameters. However, the stellar light is modified by interstellar dust along the line-of-sight from the star to the Earth before it is observed. The effect $E_s \geqslant 0$ of interstellar dust is to dim (increase) the magnitude, $\tilde{M}_s = M_s + R\, E_s$ and redden (increase) the colour $\tilde{c}_s = c_s + E_s$. The hyperparameter R is a property of the dust. The reddening E_s is drawn from an independent exponential distribution,

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

for $E_s \geqslant 0$, and zero otherwise, with scale hyperparameter τ. In each part below, show all steps.

(a) Suppose all hyperparameters are known. Derive and fully simplify the conditional probability density $P(E_s|\, \tilde{c}_s)$ of the reddening E_s given a star's dusty apparent colour $\tilde{c}_s$. Derive the colour-magnitude relation, i.e. the conditional expectation $\mathbb{E}[\tilde{M}_s|\, \tilde{c}_s]$ of the dusty absolute magnitude $\tilde{M}_s$ as a function of a given dusty apparent colour $\tilde{c}_s$. What are the asymptotic slopes of this relation as $\tilde{c}_s \to -\infty$ and as $\tilde{c}_s \to +\infty$?

[*Note that a random variable x with a truncated Gaussian density with location and scale parameters μ, σ and a lower truncation at $x = 0$ has a density function:*

$$TN(x|\, \mu, \sigma^2) = \frac{\phi((x-\mu)/\sigma)}{\sigma\, \Phi(\mu/\sigma)}$$

for $x \geqslant 0$ and zero otherwise. It has expectation value $\mathbb{E}[x] = \mu + \sigma\phi(\mu/\sigma)/\Phi(\mu/\sigma)$, where $\phi(y) \equiv N(y|\, 0, 1)$ and $\Phi(y) \equiv \int_{-\infty}^{y} \phi(t)\, dt$. For $y \ll 0$, $\phi(y)/\Phi(y) \approx -y$].

(b) Henceforth, suppose all hyperparameters are unknown, and the dusty absolute magnitude $\tilde{M}_s$ and apparent colour $\tilde{c}_s$ are measured without error for a sample of N stars, labelled $s = 1, \ldots, N$. Adopt independent, flat improper hyperpriors on each of M_0, c_0, β and R, and independent, flat positive improper hyperpriors on each of σ_M^2, σ_c^2, and τ. For the full sample of N stars, write down the unnormalised posterior probability density $P(\{E_s\}, M_0, c_0, \beta, \sigma_M^2, \sigma_c^2, R, \tau|\{\boldsymbol{d}_s\})$, where the data are $\boldsymbol{d}_s = \{\tilde{M}_s, \tilde{c}_s\}$. Draw the corresponding directed acyclic graph.

(c) Construct an MCMC algorithm to sample this joint posterior density by deriving a sequence of proposed moves that are always accepted. Specify the order of your sequence. You may assume you have access to functions that generate random draws from these probability densities:

(1) Gaussian: $N(x|\, \mu, \sigma^2)$,

(2) truncated Gaussian: $TN(x|\, \mu, \sigma^2)$,

(3) scaled inverse χ^2: Inv-$\chi^2(x|\, \nu, b^2) \propto x^{-(\nu/2+1)} e^{-\nu b^2/(2x)}$, for $x \geqslant 0$ and zero otherwise, for scale parameter b and degrees of freedom ν.

4

(a) Consider a general Bayesian inference problem with observed data y, parameter θ, likelihood function $P(y|\,\theta)$ and a proper prior $P(\theta)$. We wish to compute the evidence or marginal likelihood $Z \equiv P(y) = \int P(y|\,\theta)\, P(\theta)\, d\theta$. Assume $P(y|\,\theta) > 0$ for all possible values of y and θ. In each part below, show all steps.

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

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

Compute $\mathbb{E}_{\theta|y}[\hat{I}]$, where the expectation is taken with respect to the posterior density $P(\theta|\,y)$. How might you use $\hat{I}$ to estimate Z?

(ii) Henceforth, 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 σ^2 and the prior variance τ^2 are known. Derive and fully simplify the posterior density $P(\theta|\,y)$. What are the posterior mean and variance?

(iii) Derive the expectation of the estimator, $\mathbb{E}_{\theta|y}[\hat{I}]$, with respect to random draws from the posterior for fixed data y. Fully simplify in terms of y, σ, and τ.

(iv) Derive the variance of the estimator, $\mathrm{Var}_{\theta|y}[\hat{I}]$, for fixed data y and $\tau < \sigma$. What about for $\tau \geqslant \sigma$? Comment on the suitability of this estimator in the typical case where the prior is more diffuse than the likelihood.

(b) We wish to compare two probabilistic models for data $\mathcal{D}$: $\mathcal{M}_0$ with parameter ϕ, and $\mathcal{M}_1$ with parameters ϕ and ψ. The more complex model $\mathcal{M}_1$ reduces to the simpler model $\mathcal{M}_0$ in the special case of $\psi = 0$. Suppose further that the proper prior under $\mathcal{M}_1$ is separable, $P(\phi, \psi|\,\mathcal{M}_1) = P(\phi|\,\mathcal{M}_1)P(\psi|\,\mathcal{M}_1)$, the proper prior for ϕ is the same under each model, and all priors are nonzero for every parameter value. Show that the Bayes factor between the two models reduces to:

$$B_{01} = \frac{P(\mathcal{D}|\,M_0)}{P(\mathcal{D}|\,M_1)} = \left.\frac{P(\psi|\,\mathcal{D}, M_1)}{P(\psi|\,M_1)}\right|_{\psi=0}.$$

END OF PAPER