UNIVERSITY OF CAMBRIDGE

# MATHEMATICAL TRIPOS    Part III

Monday, 11 June, 2018    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*

***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** Suppose $N$ Type Ia supernovae are observed, all at the same observed redshift $z$, and analysis of their brightnesses yields unbiased, independent estimates of their distance moduli, $\hat{\mu}_i$, $i = 1 \ldots N$. The distance modulus is a logarithmic measure of distance $d$:

$$\mu = 25 + 5 \log_{10}[d \text{ Mpc}^{-1}],$$

where Mpc is a mega-parsec, a unit of distance. In a smooth, homogeneous, isotropic, expanding universe, the supernovae must all be at the same true distance modulus $\mu$, since they all have the same redshift $z$. However, the estimates $\hat{\mu}_i$ have different sampling variances $\sigma_i^2$ around the true $\mu$, because of observational heteroskedastic measurement error. We wish to combine the $N$ independent estimates from the $N$ individual supernovae to determine the "best" single estimate of the distance modulus $\mu$ to redshift $z$.

(i) Consider $N = 2$ supernovae. Consider all estimators that are linear combinations of the data $\hat{\mu}_1, \hat{\mu}_2$: $\hat{\mu} = \alpha_1 \hat{\mu}_1 + \alpha_2 \hat{\mu}_2$. What restriction is required of all *unbiased* linear estimators of $\mu$?

(ii) For $N = 2$, what is the sampling variance $\mathrm{Var}[\hat{\mu}]$ of the unbiased linear estimators in part 1? Find the *minimum variance* unbiased linear estimator by solving for the appropriate coefficients. Show that they can be expressed as $\alpha_i = K \sigma_i^\gamma$, and determine $K$ and $\gamma$. What is the variance of the minimum variance unbiased linear estimator?

(iii) Now generalise to $N > 2$. Consider all linear estimators of the form $\hat{\mu} = \sum_{i=1}^{N} \alpha_i \, \hat{\mu}_i$. What are the coefficients of the minimum variance unbiased linear estimator? Verify that they satisfy the first- and second-derivative conditions for a local minimum.

(iv) For $N > 2$, suppose all the uncertainties of the individual estimates are the same, $\sigma_i = \sigma$ for $i = 1, \ldots, N$. What is the variance of the minimum variance unbiased linear estimator, and how does it scale with the number of supernovae $N$?

(v) Now suppose, because of systematic uncertainties, the distance errors for $N > 2$ supernovae are jointly Gaussian and correlated between supernovae, with known pairwise covariances $\mathrm{Cov}[\hat{\mu}_i, \hat{\mu}_j] \equiv C_{ij} = \sigma_i \, \sigma_j \, \rho_{ij}$, and correlation coefficients $|\rho_{ij}| < 1$. What is required for the matrix $\boldsymbol{C}$ to be a valid covariance matrix? Assuming $\boldsymbol{C}$ is a valid covariance matrix, derive the maximum likelihood estimator $\hat{\mu}_{\text{MLE}}$. Compute the bias and variance of the MLE. Compare the variance to the Cramér-Rao bound. You may leave your answers in terms of elements $\Lambda_{ij}$ of the inverse of the covariance matrix, $\boldsymbol{\Lambda} = \boldsymbol{C}^{-1}$.

Question 1

(i) Consider $N = 2$ supernovae. Consider all estimators that are linear combinations of the data $\hat{\mu}_1, \hat{\mu}_2$: $\hat{\mu} = \alpha_1\hat{\mu}_1 + \alpha_2\hat{\mu}_2$. What restriction is required of all *unbiased* linear estimators of $\mu$?

Idea: Unbiased $\Leftrightarrow$ $\mathbb{E}[\hat{\mu}] = \mu$

$\mathbb{E}[\hat{\mu}] = \alpha_1\mathbb{E}[\hat{\mu}_1] + \alpha_2\mathbb{E}[\hat{\mu}_2] = (\alpha_1 + \alpha_2)\mu$ so require $\alpha_1 + \alpha_2 = 1$

(ii) For $N = 2$, what is the sampling variance $\text{Var}[\hat{\mu}]$ of the unbiased linear estimators in part 1? Find the *minimum variance* unbiased linear estimator by solving for the appropriate coefficients. Show that they can be expressed as $\alpha_i = K\sigma_i^\gamma$, and determine $K$ and $\gamma$. What is the variance of the minimum variance unbiased linear estimator?

$\text{Var}(\hat{\mu}) = \alpha_1^2\sigma_1^2 + \alpha_2^2\sigma_2^2$

Consider $\gamma_i = \frac{\alpha_i}{\sigma_i}$ then $\text{Var}(\hat{\mu}) = \gamma_1^2 + \sigma_2^2$ which is minimised at

$\gamma_1 = \gamma_2 = 1/2$

$\Rightarrow$ $\alpha_i = \frac{1}{2}\sigma_i$ $\Rightarrow$ $\gamma = 1$ $k = 1/2$

**2** Suppose Type Ia supernovae (SN) are standard candles: the true absolute magnitude $M_s$ (proportional to the logarithm of the luminosity) of each individual supernova $s$ is an independent draw from a narrow Gaussian population distribution

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

with unknown mean $M_0$ and unknown intrinsic "dispersion" or variance $\sigma^2_{\text{int}}$. The dimming effect of distance relates the true absolute magnitude $M_s$ to the true apparent magnitude $m_s$ for each SN $s$:

$$m_s = M_s + \mu(z_s; H_0, w, \Omega_M),$$

where the true distance modulus at the observed redshift $z_s$ is

$$\mu(z_s; H_0, w, \Omega_M) = 25 + 5\log_{10}\left[\frac{c}{H_0}\tilde{d}(z_s; w, \Omega_M)\ \text{Mpc}^{-1}\right],$$

where Mpc is a mega-parsec (a unit of distance), $c$ is the speed of light, $H_0$ is the Hubble constant, and $(w, \Omega_M)$ are other cosmological parameters, and

$$\tilde{d}(z; w, \Omega_M) = (1+z)\int_0^z \frac{dz'}{\sqrt{\Omega_M(1+z')^3 + (1-\Omega_M)(1+z')^{3(1+w)}}}$$

is a dimensionless deterministic function. However, due to heteroskedastic measurement error, the measured apparent magnitude (data) is $\hat{m}_s$. Assume that these are unbiased estimates of $m_s$ with zero-mean Gaussian error of *known* standard deviation $\sigma_{m,s}$: $\hat{m}_s|\, m_s \sim N(m_s, \sigma^2_{m,s})$. The redshift $z_s$ for each SN $s$ is known perfectly. We have independent measurements of $N$ supernovae.

(i) What is the likelihood function for one supernova $s$: $P(\hat{m}_s|\, z_s, M_0, \sigma^2_{\text{int}}, H_0, w, \Omega_M)$? Write down the likelihood function for $N$ supernovae:

$$L(M_0, \sigma^2_{\text{int}}, H_0, w, \Omega_M) = P(\{\hat{m}_s\}|\, \{z_s\}, M_0, \sigma^2_{\text{int}}, H_0, w, \Omega_M)$$

(ii) Suppose the prior is of the form $P(M_0)P(\sigma^2_{\text{int}})P(H_0)P(w, \Omega_M)$. Write down the unnormalised posterior density of $(M_0, \sigma^2_{\text{int}}, H_0, w, \Omega_M)$ given data $\{\hat{m}_s\}, \{z_s\}$.

(iii) Assume a flat improper prior on the mean absolute magnitude $M_0 \sim U(-\infty, \infty)$. Show that the integral

$$I = \int_{-\infty}^{+\infty} L(M_0, \sigma^2_{\text{int}}, H_0, w, \Omega_M)P(M_0)\, dM_0$$

is independent of $H_0$. Suppose we use a prior on the Hubble constant $P(H_0) = N(H_0|\, a, b^2)$, where $a = 73.24$ km s$^{-1}$ Mpc$^{-1}$ and $b = 1.74$ km s$^{-1}$ Mpc$^{-1}$. What is the marginal posterior of $P(H_0|\{\hat{m}_s\}, \{z_s\})$ conditional on the supernova data? Justify your answer.

(iv) We ultimately want to obtain the marginal posterior density of the dark energy equation-of-state parameter $w$: $P(w|\{\hat{m}_s\}, \{z_s\})$. Assume we have constraints from external data (e.g. baryonic acoustic oscillations or cosmic microwave background) in the form of a proper prior density $P(w, \Omega_M)$. Specify an appropriate non-informative prior on $\sigma^2_{\text{int}}$. Describe how you would implement an MCMC algorithm to generate samples from the joint parameter space. What quantities are included in the parameter vector $\boldsymbol{\theta}$ in your chain, and how would you diagnose convergence and determine an appropriate thinning factor? Describe how you would use the resulting posterior samples to compute the marginal posterior mean and standard deviation of $P(w|\{\hat{m}_s\}, \{z_s\})$.

(v) Show that the MCMC algorithm you constructed satisfies detailed balance.

**3**

(i) Consider the following plots (P1, P2, P3), which show random functions of time drawn from Gaussian process priors with different kernels. Match each of the following kernels with the plot it most likely generated:

(a) $k(t, t') = A^2 \exp(-|t - t'|/\tau)$;

(b) $k(t, t') = A^2 \exp(-|t - t'|^2/\tau^2)$;

(c) $k(t, t') = A^2 \exp\left[-\frac{2}{l^2} \sin^2\left(2\pi \frac{|t-t'|}{T}\right)\right]$.

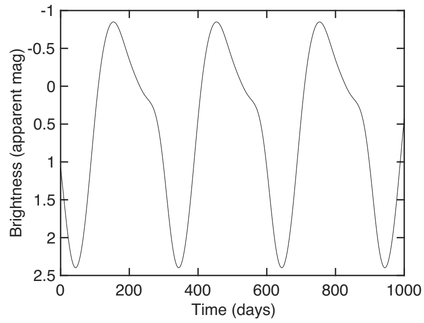


Figure 1: Plot P1: A random function of time drawn from a GP.

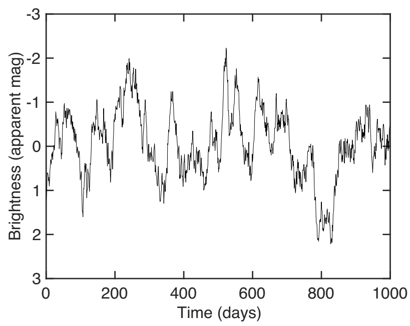


Figure 2: Plot P2: A random function of time drawn from a GP.

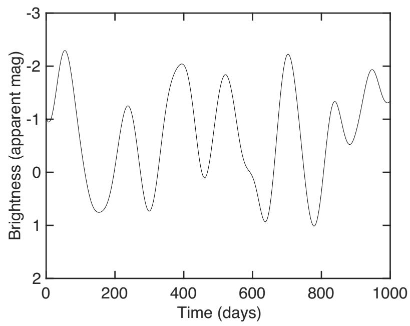


Figure 3: Plot P3: A random function of time drawn from a GP.

(ii) Suppose we have irregularly timed observations of the brightness of a periodic variable star. The true brightness light curve $f(t)$ of the star repeats every $P$ days. The mean brightness has been subtracted, so $f(t)$ may be assumed to have a long-term average of zero. The measurement of the latent brightness $f(t_i)$ at observation time $t_i$ is $y_i$ with zero-mean heteroskedastic Gaussian error with known standard deviation $\sigma_i$, for $i = 1, \dots, N$ data points. Assume a zero-mean Gaussian process prior and an appropriate covariance function, with hyperparameters $\boldsymbol{H}$, for the underlying light curve. Derive a marginal likelihood function $P(\boldsymbol{y}|\, \boldsymbol{t}, \boldsymbol{H})$, and specify the hyperparameters $\boldsymbol{H}$. How would you estimate the period $P$ of the variable star and its $1\sigma$ uncertainty?

(iii) Having now determined estimates $\hat{\boldsymbol{H}}$ of the hyperparameters $\boldsymbol{H}$, we would like to make predictions of the true, latent light curve on a regular grid of future times $t_j^*$, $j = 1 \dots M$, such that $t_1^* > t_N$. Fixing, $\boldsymbol{H} = \hat{\boldsymbol{H}}$, derive an expression for the joint posterior predictive probability of the future light curve $\boldsymbol{f}(\boldsymbol{t}^*)$, which has elements $f(t_j^*)$: $P(\boldsymbol{f}(\boldsymbol{t}^*)|\, \boldsymbol{t}^*, \boldsymbol{y}, \boldsymbol{t}, \hat{\boldsymbol{H}})$.

**4** 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 $\sigma^2_{\text{int}}$: $C_s \sim N(\mu_C, \sigma^2_{\text{int}})$. 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.,

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

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

$$H(x) = \begin{cases} 1, & x \geqslant 0, \\ 0, & x < 0. \end{cases}$$

The measured, observed colour $\hat{O}_s$ results from the sum of the intrinsic colour, reddening, and Gaussian measurement error with mean zero and known variance $\sigma^2_{O,s}$: $\hat{O}_s|E_s, C_s \sim N(C_s + E_s, \sigma^2_{O,s})$. There are $s = 1, \ldots, N$ independent supernovae in our sample. You may assume improper, noninformative and independent priors on $\ln\tau$, $\mu_C$ and $\ln\sigma^2_{\text{int}}$:

$$P(\ln\tau) \propto 1,$$
$$P(\mu_C) \propto 1,$$
$$P(\ln\sigma^2_{\text{int}}) \propto 1.$$

(i) Write down the joint probability distribution of the observed data $\{\hat{O}_s\}$, latent variables $\{C_s, E_s\}$ , and hyperparameters $\mu_C, \sigma^2_{\text{int}}, \tau$ for the sample of $N$ supernovae.

(ii) Draw a probabilistic graphical model or directed acyclic graph representing this joint distribution.

(iii) 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, $P(\{C_s, E_s\}, \mu_C, \sigma^2_{\text{int}}, \tau|\,\{\hat{O}_s\})$, 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 $TN(x|\,\mu, \sigma^2) \propto H(x) \times N(x|\,\mu, \sigma^2)$.

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

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

## END OF PAPER