Example Sheet 2
Example Class: Thursday, 27 Feb 2020, 3:30pm, MR13

Part III Astrostatistics

1 Warm-Ups

1.1 Product of Gaussian densities

Read the "multivariate_gaussian_notes.pdf" posted on the course website. Prove that the product of m multivariate Gaussian densities in random d-dimensional vector $\boldsymbol{x}$:

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

is proportional to a single Gaussian density in $\boldsymbol{x}$. Here the $\{\boldsymbol{\mu}_i, \boldsymbol{C}_i\}$ 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. Simplify for the case of $d = 1$ and $m = 2$.

1.2 Sum of Gaussian random variables

Suppose x, y, are independent univariate Gaussian random variables. The marginal distributions are given by:

$$x \sim N(\mu_x, \sigma_x^2) \tag{2}$$

$$y \sim N(\mu_y, \sigma_y^2) \tag{3}$$

Derive the probability density of $z = x + y$. (Hint: look up *characteristic function*).

1.3 Bayesian Inference for Gaussian data with unknown mean and variance

Data $\{y_i\}$ are iid from Gaussian distribution with unknown population mean μ and variance σ^2:

$$y_i \overset{iid}{\sim} N(\mu, \sigma^2) \tag{4}$$

for $i = 1, \dots, N$.

1. Derive the likelihood function $P(\boldsymbol{y}|\,\mu, \sigma^2)$, expressed in terms of the sufficient statistics: the sample mean,

$$\bar{y} = \frac{1}{N} \sum_{i=1}^{N} y_i \tag{5}$$

and sample variance:

$$s^2 = \frac{1}{N-1} \sum_{i=1}^{N} (y_i - \bar{y})^2. \tag{6}$$

2. Adopt a "non-informative" improper prior density $P(\mu, \sigma^2) \propto \sigma^{-2}$ for $\sigma^2 > 0$. Derive the posterior density $P(\mu, \sigma^2 | \boldsymbol{y})$. Show that:

$$P(\mu | \sigma^2, \boldsymbol{y}) = N(\mu | \bar{y}, \sigma^2/n) \tag{7}$$

and

$$P(\sigma^2 | \boldsymbol{y}) = \text{Inv-}\chi^2(\sigma^2 | n-1, s^2) \tag{8}$$

where the scaled inverse χ^2 distribution has an unnormalised density:

$$\text{Inv-}\chi^2(\theta \,|\, n-1, s^2) \propto \theta^{(-\nu/2+1)} \exp(-\nu s^2/(2\theta)). \tag{9}$$

3. Show that the marginal $P(\mu | \boldsymbol{y})$ is a t-distribution and derive its parameters. A t-random variable has unnormalised density:

$$t_\nu(\theta | \mu, \sigma^2) \propto \left[1 + \frac{1}{\nu} \left(\frac{\theta - \mu}{\sigma} \right)^2 \right]^{-(\nu+1)/2}. \tag{10}$$

4. Suppose the sufficient statistics of the data are $\bar{y} = 0$ and $s^2 = 1$. Plot the marginal posterior density $P(\mu | \boldsymbol{y})$ for $N = 2, 5, 10, 30$, and compare against a Gaussian density with mean $\bar{y}$ and variance s^2/N.

2 Linear Regression with heteroskedastic $(x, y)-$measurement error and intrinsic dispersion: Quasar X-ray Spectral Slopes vs. Eddington Ratios

In class we examined 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:

$$\xi_i \sim N(\mu, \tau^2) \tag{11}$$

$$\eta_i | \xi_i \sim N(\alpha + \beta \xi_i, \sigma^2) \tag{12}$$

$$x_i | \xi_i \sim N(\xi_i, \sigma^2_{x,i}) \tag{13}$$

$$y_i | \eta_i \sim N(\eta_i, \sigma^2_{y,i}) \tag{14}$$

The astronomer measures values $\mathcal{D} = \{x_i, y_i\}$ with known measurement error variances $\{\sigma^2_{x,i}, \sigma^2_{y,i}\}$, for $i = 1, \dots, N$ 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. Derive the observed data likelihood function for all the quasars:

$$L(\alpha, \beta, \sigma^2, \mu, \tau^2) = \prod_{i=1}^{N} P(x_i, y_i | \alpha, \beta, \sigma^2, \mu, \tau^2). \tag{15}$$

Show all steps and maximally simplify.

3. Write a code to find the maximum likelihood estimate, if given $\{x_i, y_i\}$ and their known measurement variances for $i = 1 \dots N$ quasars. Find an approximate 68% confidence interval for each parameter using the observed Fisher information. (Use a generic optimisation library or toolbox to numerically minimise a given function, e.g. scipy.optimize in Python, fmincon in Matlab, or optim in R, or equivalent).

4. Using the dataset provided online ("quasar_data.txt"), find the maximum likelihood estimates (MLEs) of the parameters $\alpha, \beta, \sigma^2, \mu, \tau^2$, and their uncertainties.

5. Suppose the distribution of the latent (true) independent variables $\{\xi_i\}$ is much wider than their individual uncertainties $\sigma_{x,i}$. If $\tau \gg \max(\sigma_{x,i})$, show that Eq. 15 factors

$$L(\alpha, \beta, \sigma^2, \mu, \tau^2) \approx L_1(\alpha, \beta, \sigma^2) \times L_2(\mu, \tau^2) \tag{16}$$

so that the estimation of the regression parameters $(\alpha, \beta, \sigma^2)$ decouples from the estimation of the latent distribution of the independent variables. Find $L_1(\alpha, \beta, \sigma^2)$ and $L_2(\mu, \tau^2)$. What are the maximum likelihood estimators for μ, τ^2 ?

6. Compare your MLE for β using Eq 15 against what you get using ordinary least squares (OLS), minimum χ^2, FITEXY modified χ^2 methods, and the MLE in the non-informative $\tau \gg \max(\sigma_{x,i})$ limit.

   (a) Ordinary Least Squares minimises the residual sum of squares (RSS) with respect to the parameters:

   $$RSS = \sum_{i=1}^{n} (y_i - \alpha - \beta\, x_i)^2 . \tag{17}$$

   Estimate the uncertainty (variance) of this $\hat{\beta}_{\rm OLS}$.

   (b) Minimum χ^2 or Weighted Least Squares minimises the following with respect to the parameters:

   $$\chi^2 = \sum_{i=1}^{N} \frac{(y_i - \alpha - \beta\, x_i)^2}{\sigma_{y,i}^2}. \tag{18}$$

   Estimate the uncertainty (variance) of this $\hat{\beta}$.

   (c) The FITEXY methods (Press et al. *Numerical Recipes in C*) minimise an "effective" χ^2 statistic that takes in account x-measurement errors

   $$\chi^2_{EXY} = \sum_{i=1}^{N} \frac{(y_i - \alpha - \beta\, x_i)^2}{\sigma_{y,i}^2 + \beta^2 \sigma_{x,i}^2}. \tag{19}$$

   (d) The maximum likelihood solution assuming a non-informative distribution on the $\{\xi_i\}$ is obtained by minimising $-\log L_1(\alpha, \beta, \sigma^2)$, which you derived in part (5) above. Estimate the uncertainty (variance) of this $\hat{\beta}$.

7. State and employ appropriate non-informative priors on the parameters $\alpha, \beta, \sigma^2, \mu, \tau^2$ defined in Eqs. 11 - 15. Construct and implement a MCMC algorithm to sample from the posterior probability density:

$$P(\alpha, \beta, \sigma^2, \mu, \tau^2 | \mathcal{D}) \propto L(\alpha, \beta, \sigma^2, \mu, \tau^2) \times P(\alpha, \beta, \sigma^2, \mu, \tau^2) \tag{20}$$

Run 4 independent chains, initialise appropriately at different starting points, to diagnose convergence using the Gelman-Rubin ratio. Remove "burn-in" and use the combined chains to compute the marginal distributions of the parameters, and compare against the point estimates you obtained with the other methods.

3 Importance Sampling for Bayesian Estimates of the Milky Way Mass using Angular Momentum Measurements

Look up the paper Patel et al. 2017, "Orbits of massive satellite galaxies – II. Bayesian estimates of the Milky Way and Andromeda masses using high-precision astrometry and cosmological simulations." *Monthly Notices of the Royal Astronomical Society*, 468, 3428. Use the measurements in Table 1 and the online data from the Illustris simulation to estimate the Milky Way mass using angular momentum j and the rotational velocity $v_{\max}$ of the Large Magellanic Cloud (LMC). In this context, the Milky Way is the central (host) galaxy of the system, and the LMC is a "satellite" galaxy. We wish to infer the log of the Milky Way mass, $m = \log_{10} M$.

1. Let $\boldsymbol{x} = (v_{\max}, j)$ be the latent parameters, and let $\boldsymbol{d} = (v_{\max}^{\text{obs}}, j^{\text{obs}})$ be their measured values, with uncertainties shown in Table 1. Write down the likelihood function $P(\boldsymbol{d}|\,\boldsymbol{x})$, assuming Gaussian measurement errors.

2. The Illustris simulation implicitly encodes a joint distribution between these latent dynamical parameters of satellites and the $\log_{10}$ masses of central (or host) galaxies, $P(\boldsymbol{x}, \log_{10} M)$. Assuming this exists, write down an expression for the normalised posterior probability density of the Milky Way $\log_{10}$ mass.

3. Write down an expression for the posterior mean estimate of the $\log_{10}$ MW mass in terms of integrals involving the likelihood and prior.

4. Using an arbitrary importance sampling distribution $Q(\boldsymbol{x}, \log_{10} M)$ from which we can easily draw samples, rewrite this expression in terms of expectations with respect to Q.

5. Rewrite this expression now assuming now that the importance sampling distribution is the same as the prior $Q(\boldsymbol{x}, \log_{10} M) = P(\boldsymbol{x}, \log_{10} M)$. Approximate this expression with weighted sums over the prior samples, suitable for the Monte Carlo method, and derive the importance weights.

6. Use the Illustris host-satellite data in the online file "Patel17b_Illustris_Data_KM.txt" as samples from the prior. Use the columns labelled "MVIR", "SATVMAX" and "SATJMAG". Compute the importance weights, and estimate the posterior mean and standard deviation of the $\log_{10}$ MW mass, given the LMC data $\boldsymbol{d}$. Also compute an effective sample size using Eq. B2 in the paper, and compare against the number of samples from the prior.

7. Using the bandwidth Eq. B1, create a weighted KDE representation of the posterior distribution $P(\log_{10} M|\boldsymbol{d})$. Plot it over a KDE representation of the marginal prior $P(\log_{10} M)$.

8. Prove that the optimal importance function for approximating the posterior mean of $m = \log_{10} M$, in the sense of minimum variance, is

$$Q^*(m) = \frac{|m|\, P(m|\,\boldsymbol{d})}{\int |m|\, P(m|\,\boldsymbol{d})\, dm} \tag{21}$$

   (You may use Jensen's Inequality: $\mathbb{E}[g(X)] \geq g(\mathbb{E}[X])$ where g is a convex function and X is a generic random variable.) Compare $Q^*(m)$ to your marginal importance function you have plotted, the marginal prior $P(m)$.

4 Bayesian Inference for the Hubble Constant

Consider the Hubble constant estimation problem from Example Sheet 1. Assume the measurement errors are heteroskedastic (of both the measured magnitudes and the Cepheid distance estimates), and all errors are independent, unless stated otherwise.

1. Write down the likelihood function for the parameters M_0, σ_{int}, and $\theta = 5\log_{10} h$, where $h = H_0/(100 \text{ km s}^{-1} \text{ Mpc}^{-1})$.

2. Adopting flat, improper priors on M_0 and θ, and a positive flat improper prior on $\sigma_{\text{int}} > 0$, write down the posterior density

$$P(M_0, \theta, \sigma_{\text{int}} | \mathcal{D}_K, \mathcal{D}_N) \tag{22}$$

   where $\mathcal{D}_K = \{\hat{m}_k, \hat{\mu}_{C,k}\}$ is the data of the calibrators and $\mathcal{D}_N = \{\hat{m}_i, \hat{z}_i\}$ is the data of the Hubble flow sample.

3. From the joint posterior, derive useful forms for the following:

$$P(M_0 | \theta, \sigma_{\text{int}}, \mathcal{D}_K, \mathcal{D}_N) \tag{23}$$

$$P(\theta | M_0, \sigma_{\text{int}}, \mathcal{D}_K, \mathcal{D}_N) \tag{24}$$

$$P(\theta | \sigma_{\text{int}}, \mathcal{D}_K, \mathcal{D}_N) \tag{25}$$

4. Assuming $\sigma_{\text{int}} = 0.12$, construct and implement an algorithm to sample the posterior

$$P(h | \sigma_{\text{int}} = 0.12, \mathcal{D}_K, \mathcal{D}_N). \tag{26}$$

   Apply this to the data from Dhawan et al. 2018, available in the machine-readable tables provided online. Plot a histogram of the posterior samples of h, and estimate the posterior mean and standard deviation of h. Express the uncertainty as a percentage fractional standard deviation.

5. Now with σ_{int} unknown, construct and implement an MCMC algorithm to generate samples from the marginal posterior:

$$P(h, \sigma_{\text{int}} | \mathcal{D}_K, \mathcal{D}_N) \tag{27}$$

   Describe how you initialise your chains, tune any proposals necessary, and assess convergence. Make a scatter plot of the samples from the joint distribution, plot marginal posterior histograms of the parameters, and compute the posterior mean and fractional standard deviation of h.

6. Suppose the astronomer now realises that there is a systematic error in measuring the magnitudes of the Hubble Flow sample. For only the Hubble flow supernovae, the measured magnitude is now related to the true magnitude by

$$\hat{m}_i | m_i = m_i + \epsilon_{m,i} + \xi \tag{28}$$

   The photometric errors $\epsilon_{m,i} \sim N(0, \sigma^2_{m,i})$ are still mutually independent. The systematic error is ξ, which affects all the Hubble flow supernovae equally. The true value of the systematic error ξ is unknown, but we have some prior knowledge: $\xi \sim N(0, \sigma^2_\xi)$, where the variance σ^2_ξ is known.

   Modify the likelihood, posterior, and sampler appropriately, and compute the posterior mean and fractional standard deviation of h for $\sigma_\xi = 0.01, 0.02$, and 0.05.

1.1 Product of Gaussian densities

Read the "multivariate_gaussian_notes.pdf" posted on the course website. Prove that the product of m multivariate Gaussian densities in random d-dimensional vector $\boldsymbol{x}$:

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

is proportional to a single Gaussian density in $\boldsymbol{x}$. Here the $\{\boldsymbol{\mu}_i, \boldsymbol{C}_i\}$ 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. Simplify for the case of $d = 1$ and $m = 2$.

Let $X_i \sim N(\mu_i, C_i)$ gaussians.

Then $X = \begin{pmatrix} X_1 \\ \vdots \\ X_m \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \vdots \\ \mu_m \end{pmatrix}, \begin{bmatrix} C_1 & & 0 \\ & \ddots & \\ 0 & & C_m \end{bmatrix} \right)$

and $P(X = \begin{pmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_m^T \end{pmatrix} = \prod_{i=1}^{m} N(x|\mu_i, C_i)$

$d=1,\ m=2$ $\quad X_1 \sim N(\mu_1, \sigma_1^2) \quad X_2 \sim N(\mu_2, \sigma_2^2)$

$\Rightarrow$