Bayesian statistics, home exam, spring 2011

Instructions: two exercises from (1)-(3) are enough to pass the exam with a good grade. You can use the literature, and ask for advice from the teacher. To be returned before summer.

1. You end up abroad in a big city, whose size is unknown to you. The first thing you see is a couple of busses with line numbers 87 and 144. Assume that bus lines are numbered by integers starting from 1, and that every bus line has the same probability to be observed.

   Using this observation, what you can deduce about the number of bus lines in the city ?

   Assign a prior to the number of bus lines N . Compute the posterior distribution, the posterior expectation and the maximum of the posterior of N.

   Do this under two different prior, one uninformative and one informative bases on your knowledge.

   If needed you can approximate sums by integrals.

   $$\sum_{n=1}^{\infty} f(n) \sim \int_0^{\infty} f(x)dx$$

2. We compare linear and quadratic regression models:

   $$\mathcal{M}_1 : \quad Y_i = a + bX_i + \sigma\varepsilon_i$$
   $$\mathcal{M}_2 : \quad Y_i = \alpha + \beta X_i + \gamma X_i^2 + \eta\varepsilon_i$$

   jossa $\varepsilon_i \sim \mathcal{N}(0,1)$ are i.i.d. standard -gaussian , $E(\varepsilon_i) = 0,\ E(\varepsilon_i^2) = 0$.

   The Data is given by $(X_i, Y_i, i = 1, \ldots, N)$ where X_i (meters/second) is the speed of a car at the moment of braking, and Y_i (meters) is the braking distance, how many meters away the car stops completely after braking.

   - Define informative priors for the parameters (a, b, σ), in model $\mathcal{M}_1$ and for the parameters $(\alpha, \beta\gamma, \eta)$ in model $\mathcal{M}_2$.

     In case you have been driving or travelled by car, use your own practical experience about braking. If you want you can use also your knowledge from school physics.
     The choice of the prior should not be based on the Y data, so don't look at the Y data before you have chosen the prior.
     Use the gaussian-inverse gamma joint distribution as conjugate prior.
   - Using R look at the linear and quadratic regression curves

     $X = (4, 4, 7, 7, 8, 9, 10, 10, 10, 11, 11, 12, 12, 12, 12, 13, 13, 13, 13, 14, 14, 14, 14, 15, 15, 15,$
     $16, 16, 17, 17, 17, 18, 18, 18, 18, 19, 19, 19, 20, 20, 20, 20, 20, 22, 23, 24, 24, 24, 24, 25)$
     $Y = (2, 10, 4, 22, 16, 10, 18, 26, 34, 17, 28, 14, 20, 24, 28, 26, 34, 34, 56, 26, 36, 60, 80, 20, 26,$
     $54, 32, 40, 32, 40, 50, 42, 56, 76, 84, 36, 46, 68, 32, 48, 52, 56, 64, 66, 54, 70, 92, 93, 120, 85)$

     (the data is in a file downloadable from the Koppa-webpage).

- Compute the posterior distribution of the parameters in both models.
- Compute the posteriors also by using Winbugs, compare the results.
- Compute the posterior distribution also under the non-informative improper prior with

$$\pi(\beta) = \pi(\alpha) = \pi(\gamma) = \pi(b) = \pi(a) = 1, \quad \pi(\eta^2) = \frac{1}{\eta^2}, \pi(\sigma^2) = \frac{1}{\sigma^2}$$

- The Bayes' factor is given by

$$\frac{p(Y|X,\mathcal{M}_2)}{p(Y|X,\mathcal{M}_1)} = \frac{\int_0^\infty\int_{-\infty}^\infty\int_{-\infty}^\infty\int_{-\infty}^\infty p(Y|\alpha,\beta,\gamma,\eta^2,\mathcal{M}_1)\pi(\alpha|\eta^2)\pi(\beta\eta^2)\pi(\gamma|\eta^2)\pi(\eta^2)d\alpha d\beta d\gamma d\eta^2}{\int_0^\infty\int_{-\infty}^\infty\int_{-\infty}^\infty p(Y|a,b,\sigma^2,\mathcal{M}_1)\pi(a|\sigma^2)\pi(b|\sigma^2)\pi(\sigma^2)dbdad\sigma^2}$$

  The Bayes' factor depends on the choice of the prior. Remember that in the model choice setting in order to compare two models the parameter's prior must be a proper probability distribution. An exception is the case when the same parameter appears in both models. For example the conditional variances of the residuals, σ^2, η^2, could have the same improper prior in both models.

  i) Compute the Bayes factor by using normal approximation around the maximum of the posterior. Use the informative proper prior.

  ii) Compute the same Bayes factor by using Winbugs. You can adapt the winbugs code

  Pine: Bayes factors using pseudo priors, page 38 http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/Vol3.pdf

  iii) Exactly: by integrating the parameters analytically under the conjugate prior. This you can do by looking at the normalizing constant in the conjugate prior family.

  iv) Assume that the prior probabilities for the model are $\pi(\mathcal{M}_1) = \pi(\mathcal{M}_2) = 1/2$, compute the posterior probability $P(I = 2|Y,X)$ of the model indicator I.

  v) Compute the Bayes factor again under a less informative choice of the prior.

3. (How many components in the mixture ?)

   Let $\{\mathcal{M}_m \colon m \in \mathbb{N}\}$ a mixed gaussian model gaussian model for the data X, with conditional density

$$p(x|\theta_m,\mathcal{M}_m) = \sum_{\ell=1}^m \omega_\ell \phi\Big(\frac{x - x_\ell}{\eta_l}\Big)$$

   where

$$\phi(x) = \frac{1}{\sqrt{2\pi}}\exp\big(-\frac{x^2}{2}\big)$$

   is the standardi gaussian density.

$\theta_m = (\omega_\ell, x_\ell, \eta_\ell : \ell = 1, \dots, m)$
is the parameter of the model $\mathcal{M}_m$.

Here $\omega_k \in [0, 1]$ and $(\omega_1 + \cdots + \omega_m) = 1$, which means that ω is a discrete probability. The conjugate prior distribution for the discrete probability vactor in a multinomial model is given by the Dirichlet distribution with density

$$p(\omega_1, \dots, \omega_{m-1}, \omega_m) \propto \left\{ \prod_{k=1}^{m} \omega_k^{\alpha_k - 1} \mathbf{1}(\omega_k \in [0, 1]) \right\} \mathbf{1}(\omega_1 + \cdots + \omega_m)$$

In order to obtain an identifiable model assume $x_1 \leq x_2 \leq \cdots \leq x_m$.

Let be X be the data with sample size $N = 100$,

$$X = c(-0.0153, 0.0305, -0.0120, 0.0856, 0.1742, 0.0465, 0.1982, 0.1541, 0.3241, 0.0961, 0.0981, 0.5720$$
$$0.6387, 0.8265, 0.8685, 1.0255, 1.1858, 1.1593, 1.2203, 1.3839, 1.5694, 1.3769, 1.4005, 1.5517, 1.4777, 1.$$
$$1.6918, 1.6941, 1.5925, 1.7422, 1.9580, 2.2188, 2.2998, 2.6073, 2.7795, 2.7310, 3.0587, 3.3305, 3.5364, 3.$$
$$3.6299, 3.8725, 3.7669, 3.8347, 3.5253, 3.5071, 3.3509, 3.3257, 3.2347, 3.0698, 3.3924, 2.3479, 2.1231, 1.$$
$$1.4268, 1.3024, 1.0590, 0.8167, 0.9264, 0.5895, 0.8094, 0.8233, 0.7721, 0.6843, 0.5710, 0.6625, 0.5485, 0.$$
$$0.4660, 0.3813, 0.5022, 0.4332, 0.4025, 0.3460, 0.5640, 0.4776, 0.3380, 0.3575, 0.4999, 0.5250, 0.4405, 0.$$
$$0.2838, 0.1174, 0.0647, 0.1204, 0.1818, 0.1004, 0.0915, 0.1725, 0.0785, -0.0797, 0.1072)$$

(download the data file from Koppa web-page). Compute the Bayes factor by using Winbugsilla (an exact computation with sample size $N = 100$ would be too demanding, why ?)

$$\frac{p(X|\mathcal{M}_m)}{p(X|\mathcal{M}_1)}$$

where $\mathcal{M}_1$ is the model with only one component. Do this for $m = 2, 3, 4$.

Since we are comparing two particular models with the Bayes' factor, you don't need to consider the other models. You can assume as a model prior that

$$\pi(\mathcal{M}_1) = \pi(\mathcal{M}_m) = 1/2$$

4. You don't need to answer to this question if you don't want, and in any case your answer will not change the result of the exam.

   Tell your honest phylosophical opinion about Bayes' theory:

   - You became a sincere Bayesian statisician, who accepts that all quantities which are unknown to you are in the same way random and probability represents your partial information about the uknown.
   - Or you believe that the law of nature determine the probabilities of the events. Bayes theory provides just another useful method, which requires for technical reasons to assign a probability distribution to the unknown parameters, which in reality are not to be considered as "truly random" quantities.

   It is enough to give a Yes/No answer, longer answers are also welcome.