Elementary Bayesian Analysis - Instructions for exercise report

Aki Vehtari and Jarno Vanhatalo

September 23, 2008

1 General

Most of the exersices are from the course text book. However, all the exercises are given also at the end of this document. The exercises require basic skills on Matlab:

- Read some Matlab-guide
- Use Matlab online manual, e.g.: help stats, helpwin stats, doc stats
- Use pdf- and cdf-functions in stats toolbox

There are some m-files (additional to those provided by Matlab) that are needed in the exercises. The files are collected in codes.zip -file and can be downloaded from the course web page (from the Assignments and exercises page).

There are also template files for solving the exercise report problems. These templates are updated during the course and given in the Weekly exercises page in the course web page. In the report, you should add only the lines you have written in the template file.

2 The report and grading

It is recommended that you write the reports in groups of two. Return just one report with the names of both students. Describe your solution briefly, show the code you have used, and describe and discuss your results. You will get 0-5 points for the results and additional 0-1 points for meaningful discussion (as long as there are some results). In the discussion part you may comment if something was unclear or what questions or ideas did this exercise raise. You may also discuss the relevance of the exercise for this course. The assistant will help also with these exercises in the computer lab.

A latex template for exercise report can be downloaded from the course web page (Assignments page). The template is for pdflatex and in the unix computrers you can convert it straight into pdf-file by running a command 'pdflatex report_template.tex' in the folder where you have copied the tex file. Notice, there are two figures in the template, which need to be changed in ones that are in the same folder where you copied the template. See the Matlab template for exercise 2.9&2.17 for more info. For an introduction on latex see, for example, *The Not So Short Introduction to LaTeX2e*.

Some remarks conserning the report from the previous years:

- Position the figures and code together with the exercise discussion, not in the abstract or in attachments.
- Distinguish the code from the main text by using a different font for it. For example, in Latex you can use the field verbatim to do this (see the report template).
- Include a clear caption for all the figures. Reader must be able to understand the figure from the caption!
- Write a clear description of the used model.
  - Despite the fact that the models are given in the textbook, students have used wrong models in previous years. Thus, you need to show that you have used the right model.
  - Possible programming error is not that fatal for grade, if assistant can be sure that you have tried to implement right model.
- Write too detailed discussion rather than too short. Remember that assistant has tens of reports to grade and he does not have the time to guess, what have you meant by your writings.
- Remember that discussion is more than a statement "The exercise was nice, because..."

3 Submitting the report

The exercise report is submitted by email to course assistant-

4 Exercises and hints

The exercises to be included in the exercise report are listed below.

4.1 Exercise 2.9&2.17 in Gelman et. al.

2.9 Setting the parameters for a beta prior distribution: suppose your prior distribution for θ, the proportion of Californians who support the death penalty, is beta with mean 0.6 and standard deviation 0.3

- Determine the parameters α and β of your prior distribution. Sketch the prior density function.
- A random sample of 1000 Californians is taken, and 65% support the death penalty. What are your posterior mean and variance for θ? Draw the posterior density function.

2.17 Informative prior distribution: as a modern-day Laplace, you have more definite beliefs about the ration of male and to female births than reflected by his uniform prior distribution. In particular, if θ represents the proportion of female births in a given population, you are willing to place a Beta(100,100) prior distribution on θ

- Show that this means you are more than 95% sure that θ is between 0.4 and 0.6, although you are ambivalent as to whether it is greater or less than 0.5

- Now you observe that out of a random sample of 1,000 births, 511 are boys. What is you posterior probability that $\theta > 0.5$?

Do the exercises 2.9 and 2.17 with the following addition:

In both the exercises, also try a non-informative prior. Test the sensitivity of the posterior distribution to the expected value of the prior distribution when the amount of prior information increases. Comment the results.

Hints

Template file ex2_9.m. See Matlab functions: betacdf, betapdf. In addition to the book Chapter 2, characteristics of the Beta distribution are discussed on pages 576-577, 581-582.

4.2 Exercise 2.11 in Gelman et. al. (modified)

Computing with a nonconjucate single-parameter model: suppose y_1, ..., y_5 are independent samples from a Cauchy distribution with unknown center θ and known scale 1: $p(y_i|\theta) \propto 1/(1 + (y_i - \theta)^2)$.Assume, for simplicity, that the prior distrbution for θ is uniform on [0, 100]. Given the observations $(y_1, ..., y_5) = (43, 44, 45, 46.5, 47.5)$:

- Compute the unnormalized posterior density function $p(\theta)p(y|\theta)$, on a grid of points $\theta = 0, 1/m, 2/m, ..., 100$, for some large integer m. Using the grid approximation, compute and plot the normalized posterior density function, $p(\theta|y)$, as a function of θ.
- Sample 1000 draws of θ from the posterior density and plot a histogram of the draws.
- Use the 1000 samples of θ to obtain 1000 samples from the predictive distribution of a future observation, y_6, and plot a histogram of the draws.

The purpose of the exercise is to teach the concepts of inverse CDF and grid sampling. The exercise also demonstrates a distribution, which has no finite variance and thus the central limit theorem does not apply (more of this later on the course).

Hints

Template file ex2_11.m

- You can make the grid in Matlab by the commands

```
theta=0:(100/M):100;
```

or

```
theta=linspace(0,100,M+1);
```

- When plotting the probability density, divide the density values by the step size (1/M) and scale the y-axis appropriately.

- Draw samples from the posterior distribution using the inverse CDF method (Gelman et al., p. 25).
- Note that when computing the CDF the posterior has to be normalized in such way that its cumulative sum is 1. Dividing by the step size was needed only for plotting purposes in 2.-above.
- Matlab hint: cumulative density values are ordered and so the value of theta corresponding to a given CDF value can be obtained by the command:

  ```
  theta(sum([0 cdfs]<=value))
  ```

  or

  ```
  theta(min(find(cdfs>value)))
  ```

  or in Matlab 7.*

  ```
  theta(find(cdfs>value,1))
  ```

- Matlab hint: Function `hist` accepts the number of histogram bins `n` as its second argument:

  ```
  hist(x,n)
  ```

- Cauchy distribution is equivalent to Student's-t distribution with one degree of freedom. You can use Student's-t related functions `trnd`, `tcdf` and `tpdf` for handling the Cauchy distribution.

4.3 Exercise 2.18 in Gelman et. al.

Predictive distribution and tolerance intervals: a scientist using an apparatus of known standard deviation 0.12 takes nine independent measurements of some quantity. The measurements are assumed to be normally distributed, with the stated standard deviation and unknown mean θ, where the scientist is willing to place a vague prior distribution on θ. If the sample mean obtained is 17.653, obtain limits between which a tenth measurement will lie with 99% probability. This is called a 99% tolerance interval.

4.4 Exercise 3.3 in Gelman et. al.

Estimation from two independent experiments: an experiment was performed on the effects of magnetic fields on the flow of calcium out of chicken brains. The experiment involved two groups of chicken: a control group of 32 chicken and an exposed group of 36 chickens. One measurement was taken on each chicken, and the purpose of the experiment was to measure the average flow μ_c in untreated (control) chickens and the average flow μ_t in treated chickens. The 32 measurements on the control group had a sample mean of 1.013 and a sample standard deviation of 0.24. The 36 measurements on the treatment group had a sample mean of 1.173 and a sample standard deviation of 0.20.

- Assuming the control measurements were taken at random from a normal distribution with mean μ_c and variance σ_c^2, what is the posterior distribution of μ_c? Similarly, use the treatment group measurements to determine the marginal posterior distribution of mu_t. Assume a uniform prior distribution on $(\mu_c, \mu_t, \log \sigma_c, \log \sigma_t)$

The problem of estimating two normal means with unknown ratio of variances is called the Behrens-Fisher problem.

Hints

- Function trnd creates standardized Student-t random numbers, that is, with scale 1.
- To get samples from a scaled t-distribution, multiply the standardized random numbers with the scale parameter and add the mean.
- Function prctile can be used to compute the posterior interval.
- Most common error here is to compute the scale wrong.
- (There should be difference)

4.5 Exercise 3.4 in Gelman et. al.

Inference for a 2x2 table: an experiment was performed to estimate the effect of beta-blockers on mortality of cardiac patients. A group of patients were randomly assigned to treatment and control groups: out of 674 patients receiving the control, 39 died, and out of 680 receiving the treatment, 22 died. Assume that the outcomes are independent and binomially distributed. With probabilities of death of p_0 and p_1 under the control and treatment, respectively.

- Set up a noninformative prior distribution on (p_0, p_1) and obtain posterior simulations.
- The odds ratio is defined as $(p_1/(1 - p_1))/(p_0/(1 - p_0))$. Summarize the posterior distribution for this estimand.
- Discuss the sensitivity of your inference to your choice of noninformative prior density.

Hints

`betarnd` draws from the beta distribution. Note how easy it is get samples from the posterior of the odds ratio, without need for computing analytic form of the distribution.

4.6 Exercise 3.11 in Gelman et. al.

Computation: in the bioassay example, replace the uniform prior density by a joint normal prior distribution on (α, β) with $\alpha \propto N(0, 22)$, $\beta \propto N(10, 102)$, and $\text{corr}(\alpha, \beta) = 0.5$.

- Repeat all the computations and plots of sections 3.7 wth this new prior distribution

- Check that your contour plot and scatterplot look like a compromise between the prior distribution and the likelihood (as displayed in Figure 3.4 in the book)
- Discuss the effect of this hypothetical prior information on the conclusions in the applied context

Hints

Template file ex3_11.m:

- Check that the range and spacing of A and B are sensible for the alternative prior
- Compute the log-posterior in a grid
- Scale the log-posterior by subtracting its maximum value before exponentiating
- Exponentiate
- Normalize the posterior
- Replace the direct 2D grid sampling with the factored grid sampling described in Gelman et al p. 92
- In addition to the plots, report $p(\beta > 0|x, y)$
- Note that this an example of regression model.
- Multivariate normal log-pdf `mnorm_lpdf.m`
- Check that shape of the vector x is correct (1xN).
- Useful functions: `catrand.m, binsgeq.m`

4.7 Exercise 4.2 in Gelman et. al.

Normal approximation: derive the analytic form of the information matrix and the normal approximation variance for the bioassay example.

- Your solution should include the first and second derivatives (full derivation), evaluation at the mode, and figures of your analytical solution according to esim5_1.m.
- See derivations hints in lecture slides.
- Check your analytic solution by editing bioassayfun.m and using fminunc. Include the edited function in your report. Use fminunc only for checking, do the actual minimization using analytic gradients and Newton's method.
- Useful functions: esim5_1.m, bioassayfun.m

4.8 Exercise 5.1 in Gelman et. al.

Hierarchical models and multiple comparisons:

- Reproduce the computations in Section 5.5 (Gelman et.al.) for the educational testing example. Use the posterior simulations to estimate (i) for each school j, the probability that its coaching program is the best of the eight; and (ii) for each pair of schools, j and k, the probability that the coaching program in school j is better than that in school k.
- Repeat (a), but for the simpler model with τ set to ∞ (that is, separate estimation for the eight schools). In this case, the probabilities (i) and (ii) can be computed analytically.
- Discuss how the answers in (a) and (b) differ
- In the model with τ set to 0, the probabilities (i) and (ii) have degenerate values, what are they?

Hints

- The model is presented on pages 132-137 in Gelman et al. Follow the instructions under "Computation" on page 137.
- In part b "the probabilities can be computed analytically" means that the integral can be expressed in a closed form (cf. hierarchical model, where the normalized joint posterior distribution is not in closed form). However, when computing the probability that the effect of school i is the largest, the integral needs to be evaluated numerically or via simulation.
- Do not implement pooled and separate estimates with hierarchical model and setting $\tau = 0$ or $\tau = \infty$. These were mentioned just to illustrate the relation. To better illustrate the differences between models all models need to be implemented separately.
- Note that there is slight inconsistency in the book on pages 139 and 140. The posterior variance and Chi^2 statistic were computed using the original data, $y = [28.39, 7.94, -2.75, 6.82, -0.64, 0.63, 18.01, 12.16]$;
$sd = [14.9, 10.2, 16.3, 11.0, 9.4, 11.4, 10.4, 17.6]$; but the table 5.2 shows rounded values. Use the rounded values in the excercise. This information was mentioned to avoid confusion in case that you compare your results to those mentioned in the book.

4.9 Exercise: Hierarchical model with Gibbs sampling (12 points)

Implement a

- pooled
- separate
- and a hierarchical Gaussian model

for a simple data. There are 6 machines in a factory and we want to assess the quality of their products. Quality control measurements are expensive and time-consuming, so only 5 measurements were done for each machine. In addition to the existing machines, we are interested in the possible quality of a possible seventh machine. A quality measurement can have both positive and negative values. The units of the measurements are irrelevant to the exercise.

The data is in the file factory.dat in the course webpage. Each column contains the measurements for one machine.

The pooled and separate models can be implemented without Gibbs sampling and you can directly use the results in Chapter 3. In the pooled model all measurements are combined together and in the separate model each machine has its own model.

Use Gibbs sampling to sample from the posterior of the hierarchical model. The hierarchical model and its conditional distributions can be found in section 11.7. of the book. Make a hierarchical model for the means and common model for the variances of the machine qualities by following the instructions in the book.

You can get two extra points by making a hierarchical model also for the variances of the machine qualities. When your first model works, add individual variance to each of the machine and for each of these use a common Inv-Chi^2 prior, where the scale parameter σ_0^2 is unknown and the degrees of freedom is fixed (the data doesn't contain enough information for determining the degrees of freedom). The conditional distributions for the variances of the machine qualities can be easily computed by modifying the equations 11.10 and 11.11. The conditional distribution of σ_0^2 is not of simple form, but you can sample from its distribution, for example, using the grid sampling.

Remember to perform convergence diagnostics and remove the burn-in samples. Report the results of your convergence analysis either for this or for the Metropolis sampling exercise.

Your solution should contain

- m-files for the three models
- the posterior distribution of the mean of the quality of the sixth machine with the different models
- the predictive distribution of the quality of the sixth machine with the different models
- the posterior distribution of the mean of the quality of the seventh machine with the different models
- the predictive distribution of the quality of the seventh machine with the hierarchical model, both with an uniform prior on log-sigma and and inverse Chi^2 prior on σ^2 (note that the prior on τ^2 remains uniform)
- comments on the results

Hints

In the case of hierarchical model with separate variances, the first thing to do is to explicitly write down the expression of the full posterior distribution (up to the normalization constant). In order to implement the Gibbs sampler you need the following conditional distributions

- Conditional distributions of θ_j are given by the equations (11.6) and (11.7) except that the sigmas have the subscript j.
- The conditional distribution of mu remains the same (11.8,11.9).
- The conditional distribution of σ_j^2 can be obtained as a special case of the last equation on page 50. If you are unsure, you can always compute the conditional distribution on grid and compare to the formal distribution.
- The conditional distribution of τ^2 remains the same (11.12,11.13).
- The conditional distribution of the scale parameter σ_0^2 is easiest to sample with grid sampling.
- Separate and pooled models have analytic solution, see chapter 3. Do not use MCMC for these.
- Functions: `sinvchi2rand.m sinvchi2pdf.m psrf.m geyer_imse.m ksstat.m acorr.m`

4.10 Exercise 11.3 in Gelman et. al.

Metropolis algorithm: Replicate the computations for the bioassay example of section 3.7 (in Gelman et.al.) using the Metropolis algorithm. Be sure to define your starting points and your jumping rule. Run the simulations long enough for approximate convergence.

Hints

- Use simple proposal distribution.
- In this exercise there is no need to try to find optimal proposal distribution. See page 306 for instructions for how to choose proposal distribution. In real-life prerun could be made with automatic adaptive control to adapt proposal distribution.
- Metropolis is simple algorithm, you do not need many lines of code for it (less than ten lines).
- Compute with log-densities. Reasons are explained on page 281. Remember that $log(a/b) = log(a) - log(b)$. See also `bioassaylp.m`.
- Monitor convergence visually and using `psrf.m`.
- Use the
- geyer_imse to compute autocorrelation time.
- Thin the chain and use ksstat for additional convergence check.
- Report the results of your convergence analysis either here or for the Gibbs sampling exercise.

4.11 Exercise 22.1 in Gelman et. al.

Conditional probability and elementary decision theory: Oscar has lost his dog; there is a 70% probability it is in forest A and a 30% chance it is in the forst B. If the dog is in forest A and Oscar looks there for a day, he has a 50% chance of finding the dog. If the dog is in forest B and Oscar looks there for a day, he has an 80% change of finding the dog.

- If Oscar can search only one forest for a day, where should he look to maximize his probability of finding the dog? What is the probability that the dog is still lost after the search?
- Assume Oscar made the rational decision and the dog is still lost (and is still in the same forest as yesterday). Where should he search for the dog on the second day? What is the probability that the dog is still lost at the end of the second day?
- Again assume Oscar makes the rational decision on the second day and the dog is still lost (and is still in the same forest). Where should he search for the dog on the third day? What is the probability that the dog is still lost at the end of the third day?
- (Expected value of additional information.) You will now figure out the expected value of knowing, at the beginning, which forest the dog is in. Suppose Oscar will search for at most three days, with the following payoffs: -1 if the dog is found in one day, -2 if the dog is found on the second day, and -3 if the dog is found on the third day, and -10 otherwise.
  - what is Oscars expected payoff without the additional information?
  - what is Oscars expected payoff if he knows the dog is in forest A?
  - what is Oscars expected payoff if he knows the dog is in forest B?
  - Before the search begins, how much should Oscar be willing to pay to be told which forest his dog is in?

Hints

Read question carefully and think. Compute step-by-step with care. There is also a different story for this exercise in finnish (see appendix A).

4.12 Exercise 22.2 in Gelman et. al. (Design of Experiments)

This exercise is an example of Design of Experiments (DOE). In DOE our aim is to design an experiment such that we obtain as much information as possible from the measurement(s). This is extremely important if the individual experiments are slow and/or expensive. For example, it is economically unsound to stop the production processes of a factory for a long time. It is also possible to perform DOE stepwise by computing after each experiment what is the next most useful experiment.

Use the bioassay experiment data and model. The interesting quantity is the value of LD50. Consider a situation, where the current posterior uncertainty of LD50 is too high for public authorities. In order to avoid exposing animals to poison, compute

which dose level in a new experiment would reduce the posterior variance in LD50 most.

For simplicity, you may restrict the analysis to the same dose levels as in the book (you can analyze more levels if you like). Design only one experiment.

Compute with each dose level how posterior would change if
1) animal died
2) animal didn't die
Compute with each dose level what is the probability of that
1) animal dies
2) animal doesn't die
Combine the information into the expected utility (reduction of posterior variance)

You can do the exercise by using any of the integration methods taught in the course (e.g. grid sampling, Laplace approximation, MCMC). Alternatively you may also use the adaptive Simpson quadrature method, for which there is a draft m-file and two help functions: `design_quad.m`, `bioassayp.m` and `dblquadvec.m`.

Quadrature methods are suitable for fairly smooth 1-3 dimensional integrals and the method is often faster than grid sampling or MCMC. In this particular problem, for example, the Monte Carlo -error in the MCMC-integration may be so large that we would need fairly large number of samples in order to get high enough precision. With a quadrature algorithm, however, the integration is fast and the amount of code you have to write for implementing it is les than 10 lines. This way you can consentrate on the actual evaluations of the experiment desing problem.

Couple of Comments on the Design of Experiments

Design of experiments can be also done sequentially one experiment at time if the cost of single experiment is high. The next experiment is designed after the result of the first experiment is known. This way the maximum amount of information is gained. The experiments are continued until the desired accuracy is reached.

Alternatively it is possible to do multiple experiments at a time if the cost of the experiments is not significantly higher this way. In an animal experiment one can think that each animal has high cost, but doing the experiments sequentially would take more time. If the results are needed quickly it is reasonable to do multiple experiments at a time.

4.13 Exercise 6.1.

Posterior predictive checking:

- On page 140 (Gelman et.al.), the data from the SAT coaching experiments were checked against the model that assumed identical effext in all eight school: the expected order statistics of the effect sizes were (26, 19, 14, 10, 6, 2, -3, -9), compared to observed data of (28, 18, 12, 7, 3, 1, -1, -3). Express this comparison formally as a posterior predictive check comparing this model to the data. Does the model fir the aspect of the data tested here?

- Explain why, even though the identical-schools model fits under this test, it is still unacceptable for some practical purposes.

Hints

- The parameters of the identical effect model are given on page 139 in Gelman et al. Use these and the sigmas on page 140 to generate replicated data sets. Sort the replicates (with sort) and compare the smallest (largest, 3rd largest etc.) elements with the corresponding element from the observed data.
- Note that there is couple of errors in the book conserning this exercise:
  - On page 192 in the exercise 6.1 the observed data is different than on table 5.2 on page 140. On page 192 effect of the school B is erroneously listed as 3 instead of the correct value 8.
  - The posterior parameters (7.9 and 17.4) are wrong in the page 139 and 143. You should re-evalute the posterior parameters of theta using the table 5.2, equation (5.13) and the equation in the last line of page 139.

4.14 DIC

Compute DIC and p_eff for the SAT and factory datasets (exercises 5.1 and Gibbs sampling) using a

- pooled
- separate
- and an hierarchical Gaussian model

Return

- DIC and p_eff values for the different models
- an assessment whether there are differences between the models and if so, which model should be selected according to DIC
- comments on the results

Hints

- You can use the explicit deviance formula on page 183 if you wish, but it is simpler to use equation 6.6. (page 180) directly with norm_lpdf.
- The approximation used in the derivation of DIC is exact only when the posterior distributions of the parameters are Gaussian. Thus, when computing the plug-in estimate for sigma, use exp(mean(log(sigma))) instead of mean(sigma).
- Remember to remove the burn-in samples from your Gibbs simulations!

A Vaihtoehtoinen tarina tehtävään 22.1

Jotkut opiskelijat pitivät kirjan 22.1 tehtävän tarina epämotivoivana. Numeerisesti sama tehtävä voidaan esitää myös toisella taustatarinalla joka voisi paremmin vastata todellista tilannetta, jossa päätösanalyysia käytettäisiin.

Lentokoneessa (tai tehdaskoneessa) on useita osia. Oletetaan tilanne missä vikadiagonstiikka ilmoittaa, että vikaa on, mutta ei pysty tarkemmin ilmaisemaan missä osassa. Osat itsessään voivat olla sen verran monimutkaisia järjestelmiä, että yhden osan tutkiminenkin voi viedä pidemmän ajan.

Lentokoneen/tehdaskoneen vikadiagonstiikka ilmoittaa viasta, joka voi olla osassa A tai B. Kone pitäisi saada käyttökuntoon mahdollisimman pian. Aiemman vikaantumis- ja huoltohistorian perusteella tiedetään, että on 70% todennäköisyys, että vika on osassa A ja 30% todennäköisyys, että vika on osassa B. Jos vika on osassa A, tiedetään että mekaanikko löytää tunnin etsimisellä vian 50% todennäköisyydellä. Jos vika on osassa B, tiedetään että mekaanikko löytää tunnin etsimisellä vian 80% todennäköisyydellä.

- Jos mekaanikko tutkii yhtä osaa tunnin, kumpaa osaa pitäisi tutkia? Mikä on todennäköisyys, että vikaa ei ole löytynyt tunnin jälkeen?
- Olettaen, että edellä tehtiin rationaalinen päätös ja vikaa ei löytynyt. Kumpaa osaa mekaanikon pitäisi tutkia toisen tunnin aikana? Mikä on todennäköisyys, että vikaa ei ole löytynyt toisen tunnin jälkeen?
- Olettaen, että edellä tehtiin rationaalinen päätös ja vikaa ei löytynyt. Kumpaa osaa mekaanikon pitäisi tutkia kolmannen tunnin aikana? Mikä on todennäköisyys, että vikaa ei ole löytynyt kolmannen tunnin jälkeen?
- Informaation odotettu hyöty. Oletetaan, että jos 3 tunnin vian etsimisen jälkeen vikaa ei löydy, vaihdetaan molemmat osat, mikä vie 7 tuntia aikaa. Eli jos vikaa ei löydy 3 tunnin etsimisellä, kuluu mekaanikolta aikaa yhteensä 10 tuntia.
  - Mikä on odotettu huoltoaika (ilman lisäinformaatiota)?
  - Mikä on odotettu huoltoaika jos mekaanikko tietäsi vian olevan osassa A?
  - Mikä on odotettu huoltoaika jos mekaanikko tietäsi vian olevan osassa B?
  - Kuinka paljon aikaa säästyisi jos vikadiagnostiikka pystyisi heti alussa kertomaan kummassa osassa vika on?