UNIVERSITETET I OSLO

Matematisk Institutt

Exam in:	**STK 4021/9021 – Applied Bayesian Analysis and Numerical Methods Part II of two parts**
With:	**Nils Lid Hjort**
Auxilia:	**Calculator, plus one single sheet of paper with the candidate's own personal notes**
Time for exam:	**Part I: home project, 1–14/xii/2015; Part II: Wednesday 16/xii s.y., 9^{00}–13^{00}, written exam**

This exam set contains four exercises and comprises four pages.

Note: Please write **StudentWeb number** on the top of the first page of what you hand in today. In the marking process we need to connect your solutions of today with your project report.

Exercise 1

We start with some Bayesian questions related to the binomial distribution, which is fitting in that this is also where the Presbyterian minister Thomas Bayes (1702–1761) arguably started what is now known as Bayesian statistics, with a paper published two years after his death. Below you are free to utilise the formula

$$\int_0^1 x^a (1-x)^b \,\mathrm{d}x = \frac{a!\,b!}{(a+b+1)!},$$

valid for nonnegative integers a, b.

(a) Assume that $y_1, \ldots, y_n$ is a sequence of 0–1 variables that for given probability θ are independent with the same probability $\Pr(y_i = 1 \,|\, \theta) = \theta$. Write down the conditional mean and variance of y_i as well as for $z = \sum_{i=1}^n y_i$, the observed number of events in n trials.

(b) The event probability θ is most typically an unknown quantity, however, and we shall now assume as Bayes actually did that θ has a uniform prior density on $[0, 1]$.

  (i) Find the prior mean and prior variance of θ.
  (ii) Find the mean and variance of y_i in its marginal distribution.
  (iii) Find the covariance and correlation between y_i and y_j, where $i \neq j$.
  (iv) Find finally the marginal mean and marginal variance of z.

(c) Show that the marginal distribution of the full vector $y = (y_1, \ldots, y_n)$ is

$$f_n(y_1, \ldots, y_n) = \frac{z!\,(n-z)!}{(n+1)!} \quad \text{with } z = \sum_{i=1}^n y_i.$$

- Was the Norwegian 'no thanks', regarding the prospects of having the Olympics in Oslo for 2022, an unfortunate & regrettable decision, or not? I'm curious as to whether men and women are reacting differently to this question. I therefore impulsively interview ten suitably random-looking men along with ten equally random-looking women on Karl Johans gate (explaining my question carefully and patiently, so that we may regard their 'yes' or 'no' answers as responses to a well-defined criterion). Seven among the men, and four among the women, say 'yes, it was unfortunate & regrettable'.

(d) The question is now whether Model Zero or Model One has the higher probability, in view of the data, where Model Zero takes the same 'yes'-probability θ for both groups and Model One takes 'yes'-probability θ_A for the men and θ_B for the women. Compute the required posterior probabilities for Model Zero and Model One, under the additional assumptions (i) that the two models are a priori equally likely and (ii) that θ is taken uniform in Model Zero whereas θ_A and θ_B are taken independent and uniform in Model One.

Exercise 2

STATISTICIANS OFTEN NEED to estimate several different but perhaps similar quantities. There is then a choice of handling each quantity separately or to attempt to push the quantities in question under the same grander umbrella, thereby hopefully 'lending statistical strength' across situations. This exercise makes a brief excursion into one such situation.

(a) Suppose y given θ is a Poisson with parameter θ; thus the standard estimator of θ that does not take anything extra into account is $\theta^* = y$. Assume however that θ has a Gamma prior with parameters (a, b), i.e. with density $\{b^a/\Gamma(a)\}\theta^{a-1}e^{-b\theta}$ for θ positive. Its mean and standard deviation are a/b and $\sqrt{a}/b$ (which you do not need to show here). Show that the Bayes estimate under quadratic loss may be expressed as

$$\widehat{\theta} = \widehat{\theta}(a, b) = \frac{a + y}{b + 1}.$$

(b) Calculate the risk functions $R(\theta^*, \theta)$ and $R(\widehat{\theta}, \theta)$ for the two estimators (under quadratic loss), and demonstrate that the Bayes estimator has smaller risk in a certain neighbourhood around the prior mean.

(c) Assume now that the statistician actually needs to estimate a collection of 50 such probabilities $\theta_1, \ldots, \theta_{50}$, from sets of raw data $y_i \sim \text{Pois}(\theta_i)$ that are conditionally independent given the θ_i. When these rate parameters are not widely dissimilar it may make sense to model them as having come from the same distribution, which we here take to be the Gamma (a, b). Use formulae for the mean and variance of y_i to give a recipe for finding reasonable estimates of (a, b) from the data, based on overall empirical mean and empirical standard deviation $\bar{y} = (1/50)\sum_{i=1}^{50} y_i$ and $s^2 = (1/49)\sum_{i=1}^{50}(y_i - \bar{y})^2$.

(d) Find an expression for the marginal distribution of y_i, in terms of the hyperparameters (a, b). Give also an expression for the marginal log-likelihood function for the observations, under the above circumstances, and show that the marginal maximum likelihood estimates $(\widehat{a}, \widehat{b})$ must satisfy $\bar{y} = \widehat{a}/\widehat{b}$.

(e) Conclude by giving such a 'combining strength' strategy for estimating the ensemble of θ_i parameters. Comment briefly on how you would expect this strategy to work, compared e.g. to the standard raw data method that merely uses $\theta_i^* = y_i$.

Exercise 3

STATISTICAL CLASSIFICATION involves the task of allocating objects to the right categories. There are several versions of such problems, as expected, and the following is one such type of problem formulation. Suppose objects a priori belong to one of ten classes $1, \ldots, 10$, and that a certain measurement y (typically a vector with several components) is extracted for each object. Assume further that the density of y, given that it comes from an object of type j, is equal to $f_j(y)$ (assumed in this little exercise to be known, i.e. with no unknown parameters). Finally assume that the classes have prior probabilities $\pi_1, \ldots, \pi_{10}$ (the frequencies with which they will occur in the long run, if we continue to observe many objects under a given set of circumstances).

(a) Consider the random pair (c, y), where c is the class label and y is the measurement extracted from the object. Show that

$$\Pr(c = j \mid y) = \frac{\pi_j f_j(y)}{\sum_{k=1}^{10} \pi_k f_k(y)} \quad \text{for } j = 1, \ldots, 10.$$

What is the marginal distribution of y?

(b) The challenge is to attempt to allocate a given object, for which we extract measurement y, to its correct class c. Let $\widehat{c} = \widehat{c}(y)$ be such a classification, with $\widehat{c}(y) \in \{1, \ldots, 10\}$. We work with the loss function

$$L(c, \widehat{c}) = I\{\widehat{c} \neq c\} = \begin{cases} 1 & \text{if } \widehat{c} \neq c, \\ 0 & \text{if } \widehat{c} = c. \end{cases}$$

Comment briefly on this type of loss function. For a given method or recipe $\widehat{c} = \widehat{c}(y)$, give a formula and an interpretation of its risk function.

(c) For the given object under study we view c as an unknown parameter, for which the prior distribution is defined via $\pi_1, \ldots, \pi_{10}$ defined above. Explain how the Bayes rule works for this classification problem. Explain also precisely which optimality property this Bayes strategy has.

(d) Sometimes the statistician would like to reserve judgement for some of the objects that are to be classified. Introduce therefore the extra category D, for 'doubt', so that the action space after having observed the y for a given object under study is $\{1, \ldots, 10, D\}$. Assume also that the revised loss function is as above, modulo the modification that the loss is equal to some threshold value k if the decision made is D; that is,

$$L(c, \widehat{c}) = \begin{cases} 0 & \text{if } \widehat{c} = c, \\ 1 & \text{if } \widehat{c} \neq c \text{ and } \widehat{c} \in \{1, \ldots, 10\}, \\ k & \text{if } \widehat{c} = D. \end{cases}$$

With the same prior probabilities as above, characterise the revised Bayes strategy.

Exercise 4

Consider the measurements

0.257 4.442 6.889 1.212 6.931 5.049 5.375 3.866 2.836 2.763

These are taken to be an i.i.d. sample from a uniform distribution on $[0, \theta]$, with θ an unknown parameter.

(a) Write down the likelihood function for these data, under the assumed uniform model. What is the maximum likelihood estimate?

(b) With the prior $1/\theta$ for the unknown parameter, give an explicit formula for the posterior distribution, and compute the Bayes estimate under absolute loss (i.e. $L(\theta, \widehat{\theta}) = |\widehat{\theta} - \theta|$).