UPPSALA UNIVERSITY
Department of Mathematics
Silvelyn Zwanzig

Bayesian Statistics , 10hp, 1MS900
May 26, 2023

Permitted aids: pocket calculator, one hand-written sheet of formulae (2 pages)

Time: 5 hours. For a pass (mark 3) the requirement is at least 18 points. For the mark 4, 25-31 points are necessary. For an excellent test (mark 5) the requirement is at least 32 points. Every problem is worth 5 points. For the international ECTS the following main rules are valid: A: 36-40 points, B: 28-35 points, C: 23-27 points, D: 20-22 points, E: 18-19 points.

OBS: ***Please explain and interpret your approach carefully.*** *Don't try to write more than really needed, but what you write must be clear and well argued.*

1. Consider the following model. The parameter $\theta \in \{0, 1, 2\}$ describes the willingness of the workforce to go on strike. It is coded as follows: 0 means "no willingness", 1 "willingness", 2 "ready to strike". The outcome $x \in \{0, 1, 2, 3\}$ stands for $x = 0$ no strike, $x = 1$ warning strike, $x = 2$ strike for one day, $x = 3$ unlimited strike.

x	0	1	2	3
$\theta = 0$	0.8	0.1	0.1	0
$\theta = 1$	0.5	0.3	0.2	0
$\theta = 2$	0	0.2	0.5	0.3

A query gives that the probability of readiness to go on strike ($\theta = 2$) is 0.2 and the probability of no willingness ($\theta = 0$) is 0.2. We have one observation x.

(a) Calculate the posterior distribution.

(b) Calculate the Maximum likelihood estimator (MLE) and the maximum a posteriori estimator (MAP).

(c) Compare both estimators. Discuss the prior.

2. Suppose another simplified model for analyzing the strike risk. The parameter has only two levels $\theta \in \{0,1\}$ where 1 means "ready to strike". The outcome $x \in \{0,1,2\}$ stands for $x=0$ no strike, $x=1$ limited strike, $x=2$ unlimited strike. The probability $P_\theta(x)$ is given in the following table:

x	0	1	2
$\vartheta = 0$	0.8	0.2	0
$\vartheta = 1$	0.2	0.7	0.1

Assume that the prior probability of $\theta = 1$ is p. The employer defines the loss $L(\theta, d)$ of a decision d and true θ as

	$d=0$	$d=1$
$\theta = 0$	0	0.5
$\theta = 1$	1	0

   (a) Determine the posterior distribution.

   (b) Calculate the posterior expected loss.

   (c) Determine the Bayes estimator.

   (d) Calculate the Bayes risk.

   (e) Determine the least favourable prior.

3. Consider the stroke data given in the following table. We suppose that the treatment sample is independent of the control sample.

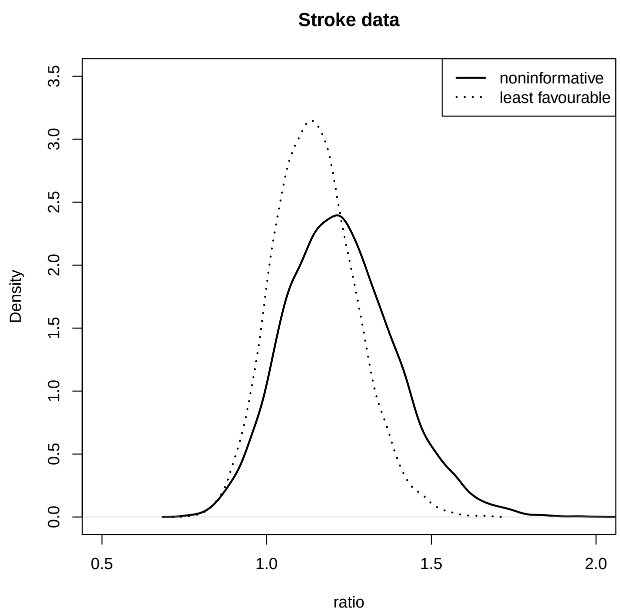
	success	size
treatment	119	11037
control	98	11034

Let p_1 be the probability of a success (to get a stroke) under the treatment and p_2 the probability of a success in the control group. We are interested in $\theta = \frac{p_1}{p_2}$, especially in testing $H_0 : \theta \geq 1$.

   (a) Which distribution model is valid?

   (b) Suppose the Jeffreys prior for p_1 and p_2.

   (c) Suppose the least favourable prior for p_1 and p_2.



(d) Derive the posterior distributions for p_1 and p_2 in both cases.

(e) Give the main steps of an algorithm generating a sample of size N from the posterior distributions of θ.

(f) In Figure 1 the density estimators of the posterior distributions of θ are plotted. Can we reject $H_0 : \theta \geq 1$?

4. Consider an i.i.d sample $X_1, \ldots, X_n$ from $X \sim \mathsf{Gamma}(\alpha, \theta)$, with density

$$f(x|\theta) = \frac{\theta^\alpha}{\Gamma(\alpha)} \, x^{\alpha-1} \exp(-\theta x) \qquad (1)$$ {2}

where $\alpha > 2$ is known. The prior is $\theta \sim \mathsf{Gamma}(\alpha_0, \beta_0)$.

(a) Calculate the posterior distribution.

(b) Consider the one point test problem $H_0 : \theta = 5$ versus $H_1 : \theta \neq 5$. Calculate the Bayes factor B_{01}.

(c) Suppose the Bayes factor $\mathsf{B}_{10} = 200$. Which conclusion is possible?

5. Consider an i.i.d sample $X_1, \dots, X_n$ from X with density $f(x|\theta)$ of a Weibull distribution with parameter $\theta = (\lambda, k)$ with $\lambda > 0$ and $k > 0$:

$$f(x|\theta) = \lambda\, k(\lambda\, x)^{k-1} \exp(-(\lambda x)^k) \text{ for } x \geq 0. \qquad (2) \quad \{1\}$$

   (a) Belongs the sample distribution to an exponential family?

   (b) Set $k = 1$. Determine the sufficient statistic and the natural parameter.

   (c) Set $k = 1$. Calculate the Fisher information.

   (d) Set $k = 1$. Derive a conjugate family. Which family of distributions is this conjugate family?

   (e) Set $k = 1$. Give the conjugate posterior distribution $p(\lambda \mid X_1, \dots, X_n)$.

   (f) Set $k = 1$. Derive Jeffreys prior.

   (g) Set $k = 1$. Does the Jeffreys prior belong to the conjugate family?

6. The wind power industry needs the data analyse of wind speed data. Consider daily averages of wind speed taken in autumn on Mondays on Gotland. We assume that the data are generated from an i.i.d. sample $X_1, \dots, X_n$ of a Weibull distribution, with density given in (2) and

$$\mathsf{E}X = \lambda\Gamma(1 + \frac{1}{k}),\ \ \mathsf{E}X^2 = \lambda^2\Gamma(1 + \frac{2}{k}),\ \text{median} = \lambda \ln(2)^{\frac{1}{k}}.$$

   Further let us assume that the data are scaled such hat $\lambda = 1$.

   (a) Propose a subjective prior distribution for $\theta = k$. Have in mind that in autumn on Gotland the chance for high wind speeds is high.

   (b) Give the main steps of an MCMC algorithm for determining the Bayes estimator. (L2-loss is supposed.)

7. Consider a multiple regression model

$$y_i = \beta_0 + x_i\beta_1 + z_i\beta_2 + \varepsilon_i, \quad i = 1, \dots, n$$

   where ε_i are i.i.d. normally distributed with expectation zero and variance $\sigma^2 = 0.25$. The unknown three dimensional parameter $\beta =$

$(\beta_0, \beta_1, \beta_2)^T$ is normally distributed with mean $\mu = (1, 1, 1)^T$ and covariance matrix

$$\Sigma = \begin{bmatrix} 1 & 0.5 & 0 \\ 0.5 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

(a) Write the model in matrix form.

(b) Determine the posterior distribution of β.

(c) Give the criterion for the MAP estimator.

(d) Give the formulary of the MAP estimator.

(e) Which name has the regularized estimator, which coincides with the MAP?

8. Consider the following R code:

```
rand<-function(N,M){rand<-rep(NA,N);
for(i in 1:N){ L<-TRUE;
while(L){rand[i]<-rcauchy(1,thetaM);
r<-prod(dcauchy(x,rand[i]))*dnorm(rand[i],2,1)/(M*dcauchy(rand[i],thetaM));
if(runif(1)<r){L<-FALSE}}}; return(rand)}
M1<-5
thetaM<-3
R1<-rand(10000,M1)
mean(R1)
```

(a) Which Bayesian model is considered? Determine $\pi(\theta)$ and $f(x|\theta)$.

(b) Which method is carried out?

(c) What is generated?

(d) Give the main steps of the algorithm.

Good Luck!

Appendix

Following distributions of a discrete random variable X with sample space $\mathcal{X} \subseteq \mathbb{Z}$ are listed: Poisson distribution $\mathsf{Poi}(\lambda)$, binomial distribution $\mathsf{Bin}(n, p)$, negative binomial distribution $\mathsf{NB}(r, p)$, geometric distribution $\mathsf{Geo}(p)$.

name	$\mathsf{P}(X = k)$	$\mathsf{E}X$	$\mathsf{Var}(X)$
$\mathsf{Poi}(\lambda)$	$\frac{\lambda^k}{k!}\exp(-\lambda)$	λ	λ
$\mathsf{Bin}(n, p)$	$\binom{n}{k}p^k(1-p)^{n-k}$	np	$np(1-p)$
$\mathsf{NB}(r, p)$	$\binom{k+r-1}{k}p^r(1-p)^k$	$\frac{r(1-p)}{p}$	$\frac{r(1-p)}{p^2}$
$\mathsf{Geo}(p)$	$p(1-p)^k$	$\frac{(1-p)}{p}$	$\frac{(1-p)}{p^2}$

Following distributions of a real valued random variable X with sample space $\mathcal{X} \subseteq \mathbb{R}$ are listed: normal distribution $\mathsf{N}(\mu, \sigma^2)$, t–distribution $\mathsf{t}_1(f, \mu, \sigma^2)$, F–distribution F_{f_1, f_2}, exponential distribution $\mathsf{Exp}\,(\lambda)$, Cauchy distribution $\mathsf{C}(m, \sigma)$, Laplace distribution $\mathsf{La}(\mu, \sigma)$, beta distribution $\mathsf{beta}(\alpha, \beta)$, gamma distribution $\mathsf{Gamma}(\alpha, \beta)$, inverse–gamma distribution $\mathsf{InvGamma}(\alpha, \beta)$.

name	$f(x\|\theta) \propto$	$\mathsf{E}X$	$\mathsf{Var}(X)$
$\mathsf{N}(\mu, \sigma^2)$	$\exp\left(-\frac{1}{2}(\frac{x-\mu}{\sigma})^2\right)$	μ	σ^2
$\mathsf{t}_1(f, \mu, \sigma^2)$	$\left(1 + \frac{1}{f}(\frac{x-\mu}{\sigma})^2\right)^{-\frac{f+1}{2}}$	μ	$\frac{f}{f-2}\sigma^2$ $f > 2$
F_{f_1, f_2}	$x^{\frac{f_1}{2}-1}\left(1 + \frac{f_1}{f_2}x\right)^{-\frac{f_1+f_2}{2}}$	$\frac{f_2}{f_2-2}$ $f_2 > 2$	$\frac{2f_2^2(f_1+f_2-2)}{f_1(f_2-2)^2(f_2-4)}$ $f_2 > 4$
$\mathsf{Exp}\,(\lambda)$	$\lambda \exp(-\lambda x)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
$\mathsf{C}(m, \sigma)$	$\left(1 + (\frac{x-m}{\sigma})^2\right)^{-1}$	–	–
$\mathsf{La}(\mu, \sigma)$	$\exp\left(-\left\|\frac{x-\mu}{\sigma}\right\|\right)$	μ	$2\sigma^2$
$\mathsf{beta}(\alpha, \beta)$	$x^{\alpha-1}(1-x)^{\beta-1}$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
$\mathsf{Gamma}(\alpha, \beta)$	$x^{\alpha-1}\exp(-x\beta)$	$\frac{\alpha}{\beta}$	$\frac{\alpha}{\beta^2}$
$\mathsf{InvGamma}(\alpha, \beta)$	$x^{-\alpha-1}\exp(-\frac{\beta}{x})$	$\frac{\beta}{\alpha-1}$ $\alpha > 1$	$\frac{\beta^2}{(\alpha-1)^2(\alpha-2)}$ $\alpha > 2$