Homework 2

Handed out: Tuesday, September 11, 2018

Due: Friday, September 21, 2018 Midnight

Notes:

- We *highly* encourage typed (Latex or Word) homework. Compile as single report containing solutions, derivations, figures, etc.
- Submit all files including report pdf, report source files (e.g. .tex or .docx files), data, figures produced by computer codes and programs files (e.g. .py or .m files) in a **.zip** folder. Programs should include a Readme file with running instructions.
- Zipped folder should be turned in on Sakai with the following naming scheme: **HW2_LastName_FirstName.zip**
- Collaboration is allowed however all submitted reports, programs, figures, etc. should be an individual student's write ups. Direct copying could be considered cheating.
- Software resources for this Homework set can be downloaded from this link or on Sakai under the Resource folder.

Problem 1-20pts

Hierarchical Bayesian models (Rat tumor experiments): Suppose that a group of scientists conduct a number of experiments to investigate the development of tumors in rats. For a given experiment j, let y_j denote the number of rats in the experiment that are observed to develop a tumor. n_j is the total number of rats in experiment j. θ_j describes the probability that a given rat within experiment j develops a tumor. $J = 71$ experiments are conducted.

One can model y_j using a Binomial distribution:

$$y_j = \mathrm{Bin}(n_j, \theta_j). \tag{1}$$

We are uncertain about the value of θ_j and can choose to model this uncertainty by defining a beta distribution over it:

$$\theta_j \sim \mathrm{Beta}(\alpha, \beta), \tag{2}$$

where α and β are hyperparameters. One can write down the joint posterior distribution of the parameters as

$$p(\theta, \alpha, \beta|y) \propto p(\alpha, \beta)p(\theta|\alpha, \beta)p(y|\theta, \alpha, \beta). \tag{3}$$

The joint posterior distribution of all parameters is given by:

$$p(\theta, \alpha, \beta|y) \propto p(\alpha, \beta) \prod_{j=1}^{J} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta_j^{\alpha+y_i-1}(1-\theta_j)^{\beta+n_j-y_j-1} \tag{4}$$

Given (α, β) the components of θ have independent posterior densities that are of the form $\theta_j^A(1-\theta_j)^B$- that is, beta densities and the joint density is expressed as

$$p(\theta|\alpha,\beta,y) = \prod_{j=1}^{J} \frac{\Gamma(\alpha+\beta+n_j)}{\Gamma(\alpha+y_j)\Gamma(\beta+n_j-y_j)} \theta_j^{\alpha+y_j-1}(1-\theta_j)^{\beta+n_j-y_j-1}. \tag{5}$$

Suppose we define the following noninformative hyperprior:

$$p(\alpha,\beta) \propto (\alpha+\beta)^{-5/2}. \tag{6}$$

One can show that this is equivalent to the following density over transformed variables:

$$p\left(\log\left(\frac{\alpha}{\beta}\right), \log(\alpha+\beta)\right) \propto \alpha\beta(\alpha+\beta)^{-5/2}. \tag{7}$$

Provided is a data.csv file containing the data of number of rats with tumors y_j and total number of rats (sample size) n_j.

(a) Obtain analytic forms of:

- the posterior distribution of Eq. (3) and
- the marginal posterior distribution over α and β: $p(\alpha,\beta|y)$ by using Eq. (4), Eq. (5) and the hint provided[1].

(b) Plot the marginal posterior density $p(\alpha,\beta|y)$ as a function of the transformed variables $\log\frac{\alpha}{\beta}$ and $\log(\alpha+\beta) \in [(-1.3,-2.3);(1,5)]$. As seen in the plot above, the marginal posterior distribution of the hyperparameters is approximately symmetric about the mode ($\approx$ (-1.75,2.8)). Obtain the corresponding value of (α,β).

Problem 2-10pts

Jeffrey's prior and maximum entropy prior: Consider a random variable x described by a Poisson distribution:

$$x \sim p(x;\theta) = \frac{\theta^x e^{-\theta}}{x!}. \tag{8}$$

(a) Determine the Jeffreys prior π^J for θ. Is the scale invariant prior $\pi_0(\theta) = \frac{1}{\theta}$ preferable to π^J? Why?

(b) Find the maximum entropy prior for θ for the reference measure π^J subject to the constraints $\mathbb{E}^\pi[\theta] = 1$, $Var^\pi[\theta] = 1$. [Note: Just write it in terms of the constraints λ_1,λ_2 and in unnormalized form]

(c) Find the maximum entropy prior for θ for the reference measure π_0 subject to the constraints $\mathbb{E}^\pi[\theta] = 1$, $Var^\pi[\theta] = 1$. [Note: Just write it in terms of the constraints λ_1,λ_2 and in unnormalized form]

[1] Hint: For example, the marginal posterior distribution of ϕ can be computed algebraically using the conditional probability formula, $p(\phi|y) = \frac{p(\theta,\phi|y)}{p(\theta|\phi,y)}$. Where θ is the parameter and y is the fixed data.

Problem 3-20pts

Laplace approximation: The data set $X = (X_1, \ldots, X_n)$ represents the number of wins of a football team in the past n home games. We can model this using

$$X_i \overset{i.i.d.}{\sim} g(x_i|\theta) = \theta(\theta+1)x_i^{\theta-1}(1-x_i), \quad x_i \in (0,1),$$

with parameter $\theta > 0$. Unfortunately, this model does not have any corresponding, useful, conjugate prior. But it is acceptable to impose a prior model on θ with Gamma distribution.

(a) Derive the posterior PDF of θ.

Hint: Make sure you do not get involved with constant terms. Just write them as a constant normalization factor.

(b) Using Laplace approximation, find a normal distribution that approximates the posterior distribution using $n = 20$,

$$\sum_{x=i} \log X_i = -4.59,$$

and $a = b = 1$ where a and b are the hyperparameters of the gamma distribution $\mathcal{G}amma(a, b)$.

Problem 4-20pts

Monte Carlo integration: Consider the following function,

$$f(x) = x^3 + 5x\cos(x)$$

(a) Calculate the integral $I = \int_a^b f(x)dx$ with $a = 3$ and $b = 4$ using Monte Carlo integration with $N = 10000$ samples. Compare this value with the exact solution.

(b) Check the relation between the number of samples N and solution accuracy by plotting the error for $N = [10, 1000]$.

(c) For $N = 100, 1000, 10000,$ and 100000 repeat the MC integration for m=10000 times. Plot the histogram of the results of MC integration for each N. Use the law of large numbers to justify the trend in the histograms.

Problem 5-20pts

Bayesian Information Criterion (BIC): Suppose we toss a biased coin where probability of heads ($x = 1$) is θ_1. However, we only know about the outcome through an unreliable friend of ours, Joey, who can be trusted with a probability θ_2. Let us call this report y. This means that we can write down $p(y|x, \theta_2)$ as

	y=0	y=1
x=0	θ_2	$1-\theta_2$
x=1	$1-\theta_2$	θ_2

(a) What is the joint probability distribution $p(x, y|\theta_1, \theta_2)$? Write your answer in a table.

(b) Consider we have the outcomes

$$x = (1, 1, 0, 1, 1, 0, 0),$$

$$x = (1, 0, 0, 0, 1, 0, 1).$$

Find the maximum likelihood estimate for θ_1 and θ_2

Hint: θ_2 is independent of θ_1 and x.

(c) We denote this model with $\mathcal{M}_2$, where index 2 stands for the number of parameters in the model. Find $p(\mathcal{D}|\hat{\theta}_1, \hat{\theta}_2, \mathcal{M}_2)$ where $\hat{\theta}$ denotes the MLE solution for parameter θ.

(d) If we also a model with 4 parameters $\bar{\theta} = (\theta_{0,0}, \theta_{0,1}, \theta_{1,0}, \theta_{1,1})$ that represents $p(x, y|\bar{\theta}) = \theta_{x,y}$. Find the MLE of $\bar{\theta}$.

(e) Find $p = (\mathcal{D}|\hat{\theta}, \mathcal{M}_4)$ where $\hat{\theta}$ denotes the MLE solution for parameters $\bar{\theta}$.

(f) Find the Bayesian Information Criterion for $\mathcal{M}_2$ and $\mathcal{M}_4$. Which model is preferred by this criterion?

Problem 6-20pts

Maximum Likelihood Estimation (MLE) and Maximum A Posteriori (MAP): Consider a random variable x described by an exponential distribution with parameter λ:

$$x \sim p(x; \lambda) = \lambda e^{\lambda x}. \tag{9}$$

We are uncertain about the value of λ and can choose to model this uncertainty by defining a Gamma distribution over it:

$$\lambda \sim \text{Gamma}(\alpha, \beta), \tag{10}$$

where the Gamma distribution is the conjugate prior for the exponential distribution.
Conjugate prior: If the posterior distributions $p(\lambda|x)$ are in the same probability distribution family as the prior probability distribution $p(\lambda)$, the prior is called a conjugate prior for the likelihood function. Therefore, the posterior can be expressed as:

$$p(\lambda|x) \propto \text{Gamma}(\alpha^*, \beta^*) \tag{11}$$

(a) Derive the maximum likelihood estimate (MLE) (λ_{MLE}) of Eq. (9)

(b) • Obtain an analytic form of the posterior distribution of Eq. (11) and

- Derive the maximum a posteriori estimator (MAP) λ_{MAP} as a function of α, β.

(c) Generate $N = 20$ samples drawn from an exponential distribution with parameter $\lambda = 0.2$. Fix $\beta = 100$ and vary α over the range $(1, 40)$ using a step-size of 1.

- Compute the corresponding MLE and MAP estimates for λ.
- For each α, compute the mean squared error [2] of both estimates compared against the true value and then plot the mean squared error as a function of α.
- Now, fix $\alpha = 30$, $\beta = 100$ and vary N over the range (1, 500) using a step-size of 1. Plot the mean squared error for each N of the corresponding estimates and explain under what conditions is the MAP estimator better.

[2] Mean square error (MSE) is defined as : MSE $= \frac{1}{N}\sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2$. Where Y_i is the true value and $\hat{Y}_i$ is the estimated value