118B HW1 - Due Thurs Oct 10 11:59PM

Please Note - The Programming Part of HW 1 is on DataHub and will be submitted there.

Please submit written solutions to these questions on Gradescope (linked off canvas)

(1) [6 points]

Bishop problem 1.3.

(2) [3 points]

Bishop problem 2.1 all parts except the entropy.

(3) [4 points]

Bishop problem 2.6. Note that the mode does not require the use of Bishop equation (2.265), only a little bit of calculus. Assume that a>1 and b>1.

(4) [6 points total]

(In this problem we will be covering the material from Bishop section 2.1.1. This material was also covered in class as the Bayesian parameter estimation problem from Tues OCT 1)

We will be following the use of Bayes' theorem as expressed in Bishop equation (1.44):

$$\text{posterior} \propto \text{likelihood} \times \text{prior}$$

Assume that we observe a series of N IID coin flips $D = \{x^{(1)}, \dots, x^{(N)}\}$, where each coin flip is modelled as a Bernoulli random variable:

$$p(x^{(i)}|\mu) = \mu^{x^{(i)}}(1-\mu)^{(1-x^{(i)})} \text{ for } i = 1 \dots N$$

The above assumptions lead to the form of the likelihood function seen in Bishop equation (2.5). Make sure you understand this process.

For the prior on μ, we assume a Beta distribution with $a = 1, b = 1$:

$$p(\mu; a = 1, b = 1) = \frac{\Gamma(1+1)}{\Gamma(1)\Gamma(1)}\mu^{1-1}(1-\mu)^{1-1} = 1$$

a) [2 points] By plugging the above forms for the likelihood function and the prior distribution over μ into Bishop equation (1,44), show that the posterior distribution has the form seen in Bishop (2.17):

$$p(\mu|D) = p(\mu; m, l, a, b) \propto \mu^{m+a-1}(1-\mu)^{l+b-1}$$

where $m = \sum_{i=1}^{N} x^{(i)}$ (the number of heads) and $l = \sum_{i=1}^{N}(1 - x^{(i)})$ (the number of tails). Make particular note of the interpretation of the hyperparameters a and b as "effective observations" mentioned on Bishop page 72.

Suppose we now want to predict the outcome of the next trial (i.e. flip of the coin). The Bayesian way to do this is to evaluate the predictive distribution of the next coin flip x, given the observed data set D. The steps to do this are described on Bishop page 73, from equation (2.19) to (2.20).

b) [4 points] Starting with $p(x = 1|D)$ from the start of Bishop equation (2.19), carefully write out and explain all the steps needed to yield Bishop equation (2.20)

(5) [3 points]

Bishop problem 1.11 for μ only.

(Other) [Not graded]

The following problems from Bishop cover material that you should know, and that may be on the midterm: 1.5, 1.6, 1.9, 1.10, 1.11 σ^2 part. We will provide solutions but not grade these questions.

(Bonus) [3 points] A Probability Puzzle. Radford Neal, 1999.

A couple you've just met invite you over to dinner, saying "come by around 5pm, and we can talk for a while before our three kids come home from the school at 6pm".

You arrive at the appointed time, and are invited into the house. Walking down the hall, your host points to three closed doors and says, "those are the kids' bedrooms". You stumble a bit when passing one of these doors, and accidently push the door open. There you see a dresser with a jewelry box, and a bed on which a dress has been laid out. "Ah," you think to yourself, "I see that at least one of their three kids is a girl".

Your hosts sit you down in the kitchen, and leaves you there while they go off to get goodies from the stores in the basement. While they're away, you notice a letter from the principal of the local school tacked up on the refrigerator. "Dear Parent", it begins, "Each year at this time, I write to all parents, such as yourself, who have a boy or boys in the school, asking you to volunteer your time to help the boys' hockey team..." "Umm", you think, "I see that they have at least one boy as well".

That, of course, leaves only two possibilities: Either they have two boys and one girl, or two girls and one boy. What are the probabilities of these two possibilities? Show your work.

NOTE: This isn't a trick puzzle. You should assume all things that it seems you're meant to assume, and not assume things that you aren't told to assume. If things can easily be imagined in either of two ways, you should assume that they are equally likely. For example, you may be able to imagine a reason that a family with two boys and a girl would be more likely to have invited you to dinner than one with two girls and a boy. If so, this would affect the probabilities of the two possibilities. But if your imagination is that good, you can probably imagine the opposite as well. You should assume that any such extra information not mentioned in the story is not available.

Hint:
Set the problem up and solve as we did for the Monty Hall Problem (see the lec 1 slides)