STAT 391
Homework 6
Out Tuesday May 19, 2020
Due Tuesday May 26, 2020

mmp@stat.washington.edu

Problem 1 – Bayesian inference for the Normal distribution
The data are $x_{1:n}$, the model is $Normal(\mu, \sigma^2)$, with σ^2 known.

a. For this question and the next, the prior for μ is $Normal(0, s^2)$. Show that the posterior mean m^{new} is always smaller in absolute value than μ^{ML}, i.e. $|m^{new}| < |\mu^{ML}|$. *FYI: Making the prior mean $m = 0$ is therefore called shrinkage in statistics. It is a way to "smooth out" too extreme values that result from the data. Another intuitive way of looking at priors with zero mean is to interpret them as follows. If the data does not strongly indicate that $\mu > 0$, then we better assume μ is near 0. For models with many parameters, if we set our prior to have high mass near 0, we "promote" sparsity, i.e. high probability that the posterior also has high mass at or near 0. There are better priors than the normal to favor sparsity.*

b. One interpretation for the prior variance s^2 is as "strenght of belief" that the value $m = 0$ is correct. More exactly, when $1/s^2$ is large, the prior is concentrated around 0. We will define the *shrinkage* of μ^{ML} as

$$1 - \frac{m^{new}}{\mu^{ML}}.$$

Prove that the shrinkage increases with $1/s^2$ and decreases with the sample size n.

c. For the next two questions, the prior $p_0(\mu)$ is uniform in $[-a, a]$. Obtain the posterior distribution of μ, $p(\mu|x_{1:n}, \sigma^2)$. Using Bayes rule, obtain the expression of $p(\mu|x_{1:n}, \sigma^2)$ as a function of a and the data. Be careful to handle all cases. Give and explicit simple expression for the normalization constant. You are allowed to use special functions like $\Gamma, \sin, \cos$ and Φ, the CDF of the standard normal distribution.

d. Show that when $\mu^{ML} > 0$, the expectation $E[\mu]$ under $p(\mu|x_{1:n}, \sigma^2)$ satisfies $|E[\mu]| < \mu^{ML}$, hence, for this prior too, there is shrinkage. *Hint 1: Make a drawing/plot of the posterior $p(\mu|x_{1:n}, \sigma^2)$. Also ask yourself what happens when $\mu^{ML} = 0$? Hint 2: Prove that $E[\mu] > 0$. Hint 3: Show that $\int_{-a}^{a} (\mu - \mu^{ML}) p(\mu|x_{1:n}, \sigma^2) d\mu < 0$ when $\mu^{ML} > 0$ by writing the integral as a sum of two integrals, one on $[-a, b]$, the other on $[b, a]$ for a cleverly chosen b. The drawing will help. You can do this even before proving hint 2.*

Problem 2 – Dirichlet/Beta distribution (Read: Ch 11 from textbook)
For $m = 2$, $S = \{1, 2\}$, the Dirichlet distribution is known as the Beta distribu tion, that is $Diri(\theta_1, \theta_2; \alpha_1, \alpha_2) = Beta(\theta_1, \theta_2; \alpha_1, \alpha_2) = \frac{\Gamma(\alpha)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \theta_1^{\alpha_1 - 1} \theta_2^{\alpha_2 - 1}$, with $\alpha_{1,2} > 0$ and $\alpha = \alpha_1 + \alpha_2$.

[**a. Change of variable – Not graded**] The likelihood of a sample of size n from a Bernoulli(θ_1, θ_2) distribution is given by $L = P(\theta_1, \theta_2) = \theta_1^{n_1} \theta_2^{n_2}$. Change the variables θ_j to $\xi_j = \ln \theta_j$ for $j = 1 : m$ and express L as a function of $\xi_{1:m}$.

Now change the variables in the Beta density to $\xi_{1:m}$. Remember the chang e of variable formula in densities! Do you need to apply it to L as well?

Calculate the expression of the posterior in the variables $\xi_{1:m}$ and show that it is also a Dirichlet/Beta distribution. The parameters $\xi_{1:m}$ are ca lled the *natural parameters* of the multinomial/Bernoulli distribution.

b. Let $m = 2$, $S = \{1, 2\}$, and $\mathcal{D} = \{1, 1, 2, 1, 1\}$. For the following 3 Dirichlet priors, give the numerical values of the *fictitious sample siz e* α, and the posterior parameters $\alpha'_{1,2}$.

$$D(\theta_1, \theta_2; 10, 1), \quad D(\theta_1, \theta_2; 10, 10), \quad D(\theta_1, \theta_2; 0.1, 0.2). \tag{1}$$

c. For each of the 3 cases above, make a plot showing the prior, posterior as functions of θ_2, as well as the location of the ML estimate θ_2^{ML} on the θ_2 axis.

d. Assume now the prior is uniform. Show that the posterior of (θ_1, θ_2) is a Beta distribution and calculate its parameters for the data in **b**. The Beta distribution is given by

$$Beta(\theta_1, \theta_2; a_1, a_2) = \frac{\Gamma(a_1 + a_2)}{\Gamma(a_1)\Gamma(a_2)} \theta_1^{a_1 - 1} \theta_2^{a_2 - 1} \tag{2}$$

e. Same as **c.** for the uniform prior.