1. Consider the Laplacian distribution with location μ and scale σ. If a random variable is $Laplace(\mu, \sigma)$, it has pdf

$$p(x) = \frac{1}{2\sigma} \exp\left(-\frac{|x-\mu|}{\sigma}\right).$$

Use this to answer the following.

(a) (10 points) Let $X_i \sim Laplace(0, \sigma)$ for $i = 1, \ldots, n$. Find a conjugate prior family for σ. (Note: to establish conjugacy, you *must* determine the form of the posterior.)

(b) (10 points) Occasionally, for convenience, distributions are re-parameterized in terms of the precision instead of the scale. If $\tau = 1/\sigma$, then τ is the precision. The pdf is then

$$p(x) = \frac{\tau}{2} \exp\left(-\tau|x - \mu|\right).$$

Suppose $Z_i \sim \mathit{Laplace}(0, 1/\tau)$ for $i = 1, \ldots, n$. Find Jeffreys' prior for τ and determine the form of the resulting posterior.

(c) (5 points) Explain why placing placing the noninformative prior on τ in part (b) is equivalent to placing a noninformative prior on σ.

2. The geometric distribution has two parameterizations. Suppose Z counts the number of failures until the first success. Then for probability of success θ, the mass function is

$$p(Z = z) = (1-\theta)^z \theta, z = 0, 1, 2, 3, \ldots$$

with expectation $E(Z) = (1-\theta)/\theta$. If, on the other hand, we want to model the probability that the first success occurs on the z^{th} trial, then the mass function is

$$p(Z = z) = (1-\theta)^{z-1} \theta, z = 1, 2, 3, \ldots$$

with expectation $E(Z) = 1/\theta$.

(a) (8 points) Show that Jeffreys' prior on θ is the same for both parameterizations of the geometric model.

(b) (7 points) For both parameterizations, find the posterior distribution that results when using Jeffreys' prior for θ. Comment on any differences you observe.

3. (10 points) Let $N \sim Borel(\mu)$ where the mass function for a Borel distributed random variable is defined as

$$p(N = n) = \frac{1}{n!}(\mu n)^{n-1}e^{-\mu n}, n = 1, 2, 3, \ldots$$

and $\mu \in [0, 1]$ with expectation $E(n) = 1/(1 - \mu)$. Using either a conjugate prior or Jeffreys' prior, find a prior distribution that respects the support of μ. (Hint 1: One of these two approaches will result in a prior that is not consistent with the support of μ. Hint 2: If you settle on Jeffreys', you do not need to specify the posterior.)

4. (10 points) Suppose we take n iid random samples from the Erlang distribution where the density of one random variable, y_i, is

$$p(y_i) = \frac{\lambda^\alpha y_i^{\alpha-1} e^{-\lambda y_i}}{(\alpha-1)!}, y_i \geq 0,$$

and $\lambda > 0$, $\alpha \in \mathbb{N}$. A conjugate family exists for the Erlang distribution, but only if one of the parameters—either α or λ—is held fixed. Determine which parameter must be fixed and then find a conjugate family prior for the remaining parameter.

MATH 640: Midterm 1, Analysis (2017)

Instructions

1. You have until Thursday, February 16, at 11:59pm to finish this analysis.
2. Please submit your analysis electronically in PDF or DOC form to the assignment on Blackboard before 11:59pm. Your report must be typed (including any formulas or derivations). Late submissions will be accepted, however they will be penalized.
3. Do not consult with each other on this analysis. E-mail me if you have questions.
4. Your write-up should consist of four sections: a brief introduction, a methods section that details the models you will run, a results section that presents the results of your analysis, and a discussion section that discusses your findings.
5. Code should be placed in an appendix, not the body of the report.
6. The report should be no longer than four pages (not including appendices). If you need space, the derivations of the models can be placed into a second appendix.
7. Please see the MATH 640 Style Guide for more details under course documents.

Prompt

A study was conducted to determine risk factors for adult-onset, or Type II, diabetes in women. People with Type II diabetes either resist the effects of naturally produced insulin in the body or do not produce enough insulin to maintain normal glucose levels. Possible complications from Type II diabetes include coronary artery disease, stroke, and heart attack. Untreated or poorly treated Type II diabetes can also lead to damage of the kidneys, eyes, and feet that could result in kidney failure, blindness, and infections requiring the amputation of toes or even the whole foot. Early intervention may prevent these complications thus it is good to identify risk factors associated with being diabetic. To this end, researchers sampled 532 women in Arizona and determined whether or not the women were diabetic (specifically, adult-onset diabetes or Type II). They also collected a host of possible risk factors: `BMI` (levels `normal`, `overweight`, and `obese`), skin fold thickness (`Skinfold`, binary, 0 for "normal," 1 for above "normal"), and diastolic blood pressure, `DBP`. Data is in the file `azdiabetes.txt` on the website.

For each group, those who are diabetic and those who are not diabetic, build Bayesian models and conduct posterior inference, using $B = 100,000$ samples, to assess the following:

1. The proportions in each group that have normal, overweight, and obese BMI levels. (Note: BMI is one categorical variable with three levels.)
2. The proportion in each group that have above normal skinfold thickness.
3. The mean diastolic blood pressure in each group. (Note: the variance is also unknown.)

Your introduction should characterize the study and briefly touch on why we might want to study risk factors for diabetes. When describing the models, you must state the likelihood and prior you selected. You must also state the resulting posterior distribution. Your results section should provide descriptions of the results of the models. You may include tables and figures, but these must be labeled and appropriately referenced in the body of the text. In your discussion of the analysis, compare the results between the two groups and draw conclusions on potential risk factors for Type II diabetes in women.