ISTA 421/521 – Homework 4

Due: Tuesday, November 13, 10pm

20 pts total for Undergrads, 25 pts total for Grads

Emanuel Carlos de Alcantara Valente

Undergraduate

Instructions

In this assignment you are required to write 2 python scripts, for problems 6 and 7. This script for problem 6 should be named `plot_laplace_approx.py`, and the script for problem 7 should be named `pi_sample_estimate.py`.

All problems require that you provide some "written" answer, and in problem 6 this will include figures (and possibly elsewhere as well). (You can use LaTeX or any other system (including handwritten; plots, of course, must be program-generated) as long as the final version is in PDF.)

The final submission will include (minimally) the 2 python scripts and a PDF version of your written part of the assignment. You are required to create either a .zip or tarball (.tar.gz / .tgz) archive of all of the files for your submission and submit your archive to the d2l dropbox by the date/time deadline above.

NOTE: Problem 4 is required for Graduate students only; Undergraduates may complete it for extra credit equal to the point value.

(FCMA refers to the course text: Rogers and Girolami (2012), *A First Course in Machine Learning*. For general notes on using LaTeX to typeset math, see: `http://en.wikibooks.org/wiki/LaTeX/Mathematics`)

1. [2 points] Adapted from **Exercise 3.1** of FCMA p.133:

   For $\alpha, \beta = 1$, the beta distribution becomes uniform between 0 and 1. In particular, if the probability of a coin landing heads is given by r and a beta prior is placed over r, with parameters $\alpha = 1$, $\beta = 1$, this prior can be written as follows:

$$p(r) = 1 \quad (0 \leq r \leq 1)$$

   Using this prior, compute the posterior density for r if y heads are observed in N tosses (i.e., multiply this prior by the binomial likelihood and manipulate the result to obtain something that looks like a beta density).

   Solution. As our prior is a beta prior and our likelihood is a binomial likelihood, our posterior is conjugate and it will look like a beta density. So, we will have:

$$p(r|y_N) \propto p(y_N|r)p(r)$$

$$p(r|y_N) \propto \left[\binom{N}{y_N} r^{y_N}(1-r)^{N-y_N}\right] \times 1$$

   As we saw, the beta density, with parameters δ and γ , has the following general form:

   $Kr^{\delta-1}(1-r)^{\gamma-1}$, where K is a constant.

   Our posterior can be written in that way. So:

$$p(r|y_N) \propto \left[Kr^{\delta-1}(1-r)^{\gamma-1}\right]$$

   where: $K = \binom{N}{y_N}$; $\delta = y_N + 1$; $\gamma = N + 1 - y_N$

   Therefore

$$p(r|y_N) = \frac{\Gamma(2+N)}{\Gamma(1+y_N)\Gamma(1+N-y_N)} r^{y_N}(1-r)^{N-y_N}$$

2. [2 points] Adapted from **Exercise 3.2** of FCMA p.134:

   Repeat the previous exercise for the following prior, also a particular form of the Beta density:

$$p(r) = \begin{cases} 2r & 0 \leq r \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

   What are the values of the prior parameters α and β that result in $p(r) = 2r$?

   Solution.

   The values of the prior parameters α and β that result in $p(r) = 2r$ are 2 and 1 respectively.

   Proof: Substituting $\alpha = 2$ and $\beta = 1$ we will have:

$$p(r) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} r^{\alpha-1}(1-r)^{\beta-1} = \frac{\Gamma(2+1)}{\Gamma(2)\Gamma(1)} r^{2-1}(1-r)^{1-1} =$$

$$= \frac{\Gamma(3)}{\Gamma(2)\Gamma(1)} r(1-r)^0 = \frac{2!}{1!0!} r = 2r$$

As our prior is a beta prior and our likelihood is a binomial likelihood, our posterior is conjugate and it will look like a beta density. So, we will have:

$$p(r|y_N) \propto p(y_N|r)p(r)$$

$$p(r|y_N) \propto \left[\binom{N}{y_N} r^{y_N}(1-r)^{N-y_N}\right] \times 2r$$

$$p(r|y_N) \propto \left[2\binom{N}{y_N} r^{y_N+1}(1-r)^{N-y_N}\right]$$

As we saw, the beta density, with parameters δ and γ , has the following general form:

$Kr^{\delta-1}(1-r)^{\gamma-1}$, where K is a constant.

Our posterior can be written in that way. So:

$$p(r|y_N) \propto \left[Kr^{\delta-1}(1-r)^{\gamma-1}\right]$$

where: $\delta = y_N + 2$; $\gamma = N - y_N + 1$, and $K = 2\binom{N}{y_N}$

Therefore, as we know that $\alpha = 2$ and $\beta = 1$:

$$p(r|y_N) = \frac{2\Gamma(2+1+N)}{\Gamma(2+y_N)\Gamma(1+N-y_N)} r^{y_N+1}(1-r)^{N-y_N} =$$

$$= \frac{2\Gamma(3+N)}{\Gamma(2+y_N)\Gamma(1+N-y_N)} r^{y_N+1}(1-r)^{N-y_N}$$

3. [3 points] Adapted from **Exercise 3.5** of FCMA p.134:

   If a random variable R has a beta density

   $$p(r) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} r^{\alpha-1}(1-r)^{\beta-1},$$

   derive an expression for the expected value of r, $\mathbb{E}_{p(r)}\{r\}$. You will need the following identity for the gamma function:

   $$\Gamma(n+1) = n\Gamma(n).$$

   Hint: use the fact that

   $$\int_{r=0}^{r=1} r^{\alpha-1}(1-r)^{b-1}\ \mathrm{d}r = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}.$$

   Solution.

   We know the fact that $\mathbb{E}_{p(r)}\{r\}$ is given by:

   $$\mathbb{E}_{p(r)}\{r\} = \int_{r=0}^{r=1} r \left[\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} r^{\alpha-1}(1-r)^{\beta-1}\right]\ \mathrm{d}r$$

   Solving it, we will have:

$$\mathbb{E}_{p(r)}\{r\} = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_{r=0}^{r=1} r^{\alpha+1-1}(1-r)^{\beta-1}\ \mathrm{d}r$$

Consider: $\alpha^{'} = \alpha + 1$. So, our expectation can be write in this following form:

$$\mathbb{E}_{p(r)}\{r\} = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_{r=0}^{r=1} r^{\alpha^{'}-1}(1-r)^{\beta-1}\ \mathrm{d}r$$

By the hint given, we know that:

$$\int_{r=0}^{r=1} r^{\alpha^{'}-1}(1-r)^{\beta-1}\ \mathrm{d}r = \frac{\Gamma(\alpha^{'})\Gamma(\beta)}{\Gamma(\alpha^{'}+\beta)} = \frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+1+\beta)}$$

Therefore:

$$\mathbb{E}_{p(r)}\{r\} = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+1+\beta)} = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{\alpha\Gamma(\alpha)\Gamma(\beta)}{(\alpha+\beta)\Gamma(\alpha+\beta)} \rightarrow$$

$$\mathbb{E}_{p(r)}\{r\} = \frac{\alpha}{\alpha+\beta}$$

4. [5 points; **Required only for Graduates**] Adapted from **Exercise 3.12** of FCMA p.135:

   When performing a Bayesian analysis of the Olympics data, we assumed that σ^2 was known. If instead we assume that $\mathbf{w}$ is known and an inverse Gamma prior is placed on σ^2

   $$p(\sigma^2|\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}(\sigma^2)^{-\alpha-1}\exp\left\{-\frac{\beta}{\sigma^2}\right\},$$

   the posterior over σ^2 will also be inverse Gamma. Derive the posterior parameters.

   Solution.

   We know that

   $$p(\sigma^2|t,X,w) \propto p(t|w,X,\sigma^2)p(\sigma^2|\alpha,\beta)$$

   and our likelihood is going to be the same. So:

   $$p(\sigma^2|t,X,w) \propto \frac{1}{2\pi^{\frac{N}{2}}|\sigma I|} exp\left\{\frac{-1}{2}(t-Xw)^T(\sigma^2 I)(t-Xw)\right\}\frac{\beta^\alpha}{\Gamma(\alpha)}(\sigma^2)^{-\alpha-1}\exp\left\{-\frac{\beta}{\sigma^2}\right\}$$

   $$p(\sigma^2|t,X,w) \propto \frac{\sigma^{-N}\beta^\alpha\sigma^{-2\alpha-2}}{2\pi^{\frac{N}{2}}|\sigma I|\Gamma(\alpha)} exp\left\{\frac{-1}{2\sigma^2}\sum_{d=1}^{N}(t_d - x_d w)^2 - \frac{\beta}{\sigma^2}\right\}$$

   $$p(\sigma^2|t,X,w) \propto \frac{\beta^\alpha\sigma^{-(N+2\alpha+2)}}{2\pi^{\frac{N}{2}}\Gamma(\alpha)} exp\left\{\frac{-1}{\sigma^2}\left(\frac{1}{2}\sum_{d=1}^{N}(t_d - x_d w)^2 + \beta\right)\right\}$$

$$p(\sigma^2|t, X, w) \propto \frac{\beta^{\alpha}(\sigma^2)^{-(N/2+\alpha+1)}}{2\pi^{\frac{N}{2}}\Gamma(\alpha)} exp\left\{\frac{-1}{\sigma^2}\left(\beta + \frac{1}{2}\sum_{d=1}^{N}(t_d - x_d w)^2\right)\right\}$$

In fact, $p(\sigma^2|t, X, w)$ looks like an inverse Gamma distributon where the new parameters are:

$$\alpha_{new} = \frac{N}{2} + \alpha$$

and

$$\beta_{new} = \beta + \frac{1}{2}\sum_{d=1}^{N}(t_d - x_d w)^2$$

5. [5 points] Adapted from **Exercise 4.2** of FCMA p.165-166:

   In Chapter 3, we computed the posterior density over r, the probability of a coin giving heads, using a beta prior and a binomial likelihood. Recalling that the beta prior, with parameters α and β, is given by

$$p(r|\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} r^{\alpha-1}(1-r)^{\beta-1}$$

   and the binomial likelihood, assuming y heads in N throws, is given by

$$p(y|r, N) = \binom{y}{N} r^y (1-r)^{N-y}$$

   compute the Laplace approximation to the posterior. (Note, you should be able to obtain a closed-form solution for the MAP value, $\hat{r}$, by setting the log posterior to zero, differentiating, equating to zero and solving for r.)

   Solution.

   We can write the posterior as following:

$$p(r|y) \propto p(y|r, N)p(r, \alpha, \beta) = \binom{N}{y} r^y(1-r)^{N-y} \times \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} r^{\alpha-1}(1-r)^{\beta-1}$$

$$= \binom{N}{y}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} r^{y+\alpha-1}(1-r)^{N-y+\beta-1}$$

   Finding the mode $\hat{r}$ by setting the log posterior to zero, differentiating, equating to zero and solving for r:

$$log(p(r|y)) = log\binom{N}{y}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} r^{y+\alpha-1}(1-r)^{N-y+\beta-1}$$

$$log\binom{N}{y}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} + log(r^{y+\alpha-1}) + log((1-r)^{N-y+\beta-1})$$

   Differentiating:

$$\frac{\partial \log(p(r|y)}{\partial r} = \frac{(y+\alpha-1)}{r} + \frac{N-y+\beta-1}{r-1}$$

Setting to zero

$$\frac{(y+\alpha-1)}{r} + \frac{N-y+\beta-1}{r-1} = 0 \leftrightarrow r = \frac{y+\alpha-1}{\alpha+N+\beta-2}$$

Therefore:

$$\hat{r} = \frac{y+\alpha-1}{\alpha+N+\beta-2}$$

Next, we will find the variance, σ^2, of the approximating Gaussian. We need the second derivative of $log p(r|y)$ with respect to r.

We calculated:

$$\frac{\partial \log(p(r|y)}{\partial r} = \frac{(y+\alpha-1)}{r} + \frac{N-y+\beta-1}{r-1}$$

So:

$$\frac{\partial^2 \log(p(r|y)}{\partial r^2} = \frac{-(r-1)^2(y+\alpha-1) - r^2(N-y+\beta-1)}{r^2(r-1)^2}$$

The variance will be equal to the negative inverse of this quatity evaluated at $r = \hat{r}$. In particular:

$$\sigma^2 = \frac{-\hat{r}^2(\hat{r}-1)^2}{(\hat{r}-1)^2(-y+-\alpha+1) + \hat{r}^2(-N+y-\beta+1)}$$

where:

$$\hat{r} = \frac{y+\alpha-1}{\alpha+N+\beta-2}$$

Finally,

$$p(r|y) \approx \mathcal{N}(\mu, \sigma^2)$$

where:

$$\mu = \hat{r} = \frac{y+\alpha-1}{\alpha+N+\beta-2}$$

$$\sigma^2 = \frac{-\hat{r}^2(\hat{r}-1)^2}{(\hat{r}-1)^2(-y+-\alpha+1) + \hat{r}^2(-N+y-\beta+1)}$$

6. [4 points] Adapted from **Exercise 4.3** of FCMA p.166:

   In the previous exercise you computed the Laplace approximation to the true beta posterior. In this problem, plot both the true beta posterior and the Laplace approximation for the three sets of values:

   1. $\alpha = 5$, $\beta = 5$, $N = 20$, and $y = 10$,
   2. $\alpha = 3$, $\beta = 15$, $N = 10$, and $y = 3$,
   3. $\alpha = 1$, $\beta = 30$, $N = 10$, and $y = 3$.

   Be sure to clearly indicate the values in your plot captions. Include how the two distributions (the true beta posterior and the Laplace approximation) compare in each case. Include the python script you use to generate these plots; the script should be named `plot_laplace_approx.py`. **Suggestion**: for plotting the beta and Gaussian (Normal) distributions, use `scipy.stats.beta` and `scipy.stats.normal` to create the beta and Gaussian random variables, and use the `pdf(x)` method for each to generate the curves. Note that for `scipy.stats.normal`, the mean is the location (`loc`) parameter, and the sigma is the `scale` parameter. Also, `scipy.stats.normal` expects the scale parameter to be the standard deviation (i.e., take the square root: `math.sqrt(x)`) of the variance you'll compute for the Laplace approximation.

   Solution.

   The figures plotted are Figure 1, 2, and 3.

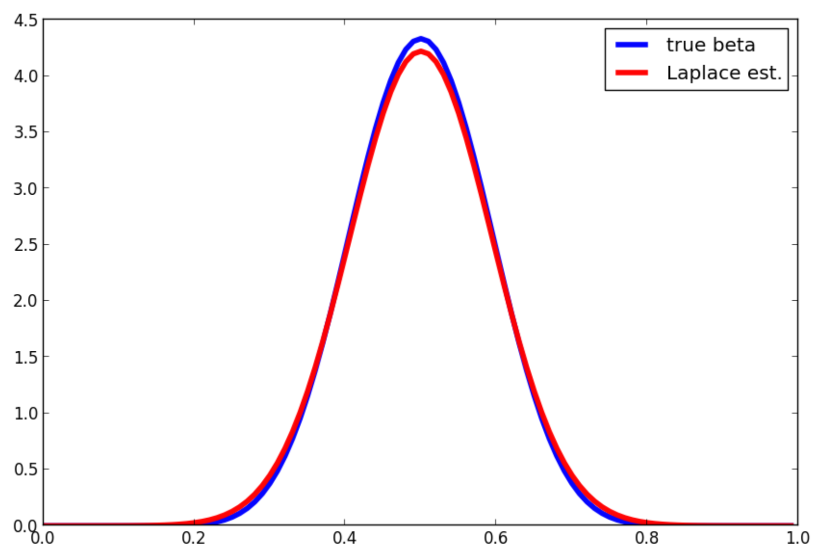

Figure 1: True Beta Posterior and the Laplace approximation: $\alpha = 5$, $\beta = 5$, N = 20, y = 10

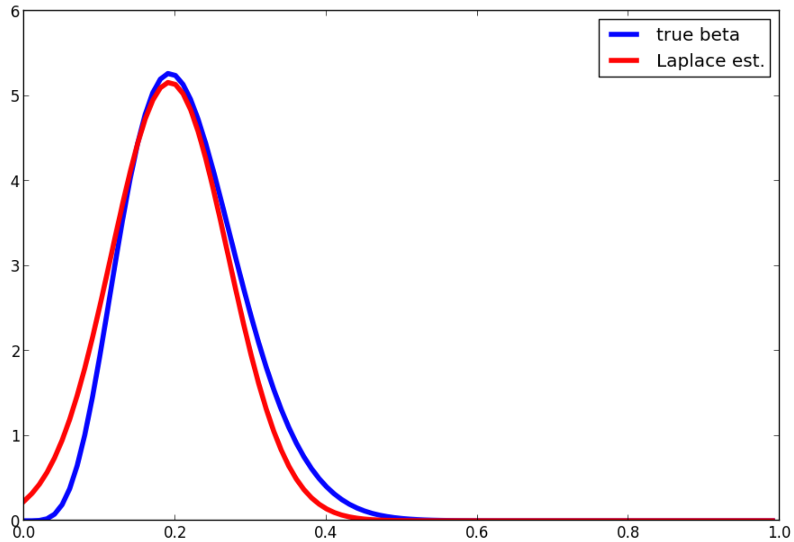

Figure 2: True Beta Posterior and the Laplace approximation: $\alpha = 3$, $\beta = 15$, N = 10, y = 3

7. [4 points] Adapted from **Exercise 4.4** of FCMA p.166:

   Given the expression for the area of a circle, $A = \pi r^2$, and using only uniformly distributed random variates, devise a sampling approach for computing π. Describe your method in detail and provide you script to do the estimation – this script should be called `pi_sample_estimate.py`.

   Solution.

   I chose using the relation between the are of a square and the are of a circle. This relation is given by:

   If a square has each side equals to $2R$, we will have:

$$S_{square} = 2R \times 2R = 4R^2$$

   If a circle has its radius equals to R, we will have:

$$S_{circle} = \pi R^2$$

   The relation between the areas going to be:

$$\frac{S_{square}}{S_{circle}} = \frac{4R^2}{\pi R^2} = \frac{4}{\pi} \rightarrow \pi = \frac{4S_{circle}}{S_{square}}$$

   Our method samples points (using an uniform distribution) inside a square with a circle inscribed in it. In the end, we pick up the points that are inside of the circle and we divide this points by the total points (area of the square).

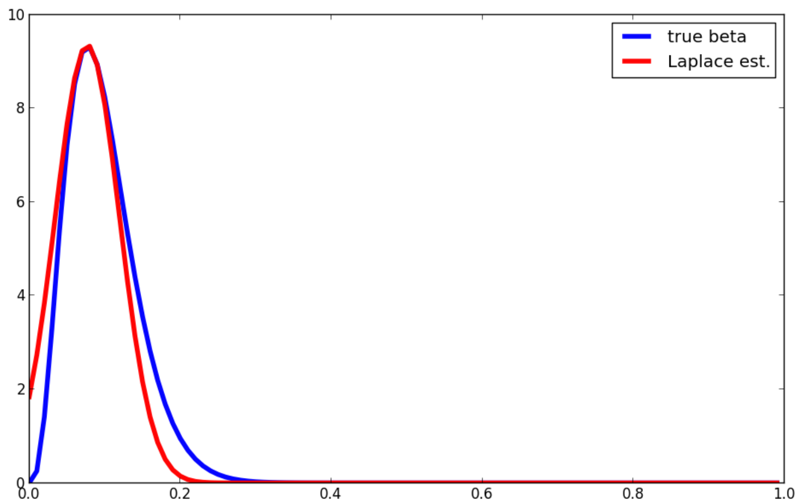

Figure 3: True Beta Posterior and the Laplace approximation: $\alpha = 1$, $\beta = 30$, N = 10, y = 3