Homework #1

Stat4DS2+DS
https://elearning2.uniroma1.it/course/view.php?id=4951

deadline 23/03/2017 (23:55)

Your Last+First Name ________________________ Your Matricola _______________

1) Sample survey: Suppose we are going to sample 100 individuals from a county (of size much larger than 100) and ask each sampled person whether they support policy Z or not. Let $Y_i = 1$ if person i in the sample supports the policy, and $Y_i = 0$ otherwise.

a) Assume $Y_1, ..., Y_{100}$ are, conditional on θ, i.i.d. binary random variables with expectation θ. Write down the joint distribution of $Pr(Y_1 = y_1, ..., Y_{100} = y_{100}|\theta)$ in a compact form. Also write down the form of $Pr(\sum_{i=1}^n Y_i = y|\theta)$.

```
#
#
#
#
#
#
#
#
#
```

b) For the moment, suppose you believed that $\theta \in \{0.0, 0.1, ..., 0.9, 1.0\}$. Given that the results of the survey were $\sum_{i=1}^n Y_i = 57$, compute

$$Pr(\sum_{i=1}^{n} Y_i = 57|\theta)$$

for each of these 11 values of θ and plot these probabilities as a function of θ.

```
#
#
#
#
#
#
#
#
#
```

c) Now suppose you originally had no prior information to believe one of these θ-values over another, and so $Pr(\theta = 0.0) = Pr(\theta = 0.1) = ... = Pr(\theta = 0.9) = Pr(\theta = 1.0)$. Use Bayes' rule to compute $\pi(\theta| \sum_{i=1}^n Y_i = 57)$ for each θ-value. Make a plot of this posterior distribution as a function of θ.

```
#
#
#
#
#
#
#
```

```
#
#
```

d) Now suppose you allow θ to be any value in the interval $\Theta = [0, 1]$. Using the uniform prior density for $\theta \in [0, 1]$, so that $\pi(\theta) = I_{[0,1]}(\theta)$, plot $\pi(\theta) \times Pr(\sum_{i=1}^{n} Y_i = 57|\theta)$ as a function of θ.

```
#
#
#
#
#
#
#
#
#
```

e) As discussed in this chapter, the posterior distribution of θ is $Beta(1 + 57, 1 + 100 - 57)$. Plot the posterior density as a function of θ. Discuss the relationships among all of the plots you have made for this exercise.

```
#
#
#
#
#
#
#
#
#
```

2) Consider a normal statistical model with $X_i \sim N(\theta, \lambda = 1/\sigma^2)$ where the precision parameter is known. Use as a prior distribution on the (conditional) mean θ a Normal with prior mean μ and prior precison ν.

a) derive the general formula of the prior predictive distribution

```
#
#
#
#
#
#
#
#
#
```

b) derive the general formula of the posterior predictive distribution

```
#
#
#
#
#
#
#
#
#
```

c) assume that the known value of λ is 1/3 and suppose you have observed the following data

$$-1.25\ \ 8.77\ \ 1.18\ \ 10.66\ \ 11.81\ \ -6.09\ \ 3.56\ \ 10.85\ \ 4.03\ \ 2.13$$

Elicit your prior distribution on the unknown θ in such a way that your prior mean is 0 and you believe that the unknown theta is in the interval $[-5, 5]$ with prior probability 0.96

```
#
#
#
#
#
#
#
#
#
```

d) derive your posterior distribution and represent it graphically

```
#
#
#
#
#
#
#
#
#
```

e) derive your favorite point estimate and interval estimate and motivate your choices

```
#
#
#
#
#
#
#
#
#
```

3) As an alternative model for the previous 10 observations

$$-1.25\ \ 8.77\ \ 1.18\ \ 10.66\ \ 11.81\ \ -6.09\ \ 3.56\ \ 10.85\ \ 4.03\ \ 2.13$$

consider the following statistical model where $X_i|\theta$ are i.i.d with

$$X_i|\theta \sim f(x|\theta) = \frac{1}{20} I_{[\theta-10,\theta+10](x)}$$

Use the same prior elicitation for θ as in the model of the previous excercise

a) Provide a fully Bayesian analysis for these data explaining all the basic ingredients and steps for carrying it out. In particular, compare your final inference on the uknown $\theta = E[X|\theta]$ with the one you have derived in the previous point 2)

```
#
#
#
#
```

```
#
#
#
#
#
```

b) Write the formula of the prior predictive distribution of a single observation and explain how you can simulate i.i.d random drws from it. Use the simulated values to represent approximately the predictive density in a plot and compare it with the prior predictive density of a single observation of the previous model

```
#
#
#
#
#
#
#
#
#
```

c) Consider the same discrete (finite) grid of values as parameter space Θ for the conditional mean θ in both models. Use this simplified parametric setting to decide whether one should use the Normal model rather than the Uniform model in light of the observed data.

```
#
#
#
#
#
#
#
#
#
```

```
## This homework will be graded and it will be part of your final evaluation
##
##
```

```
## Last update by LT: Mon Mar 13 08:15:22 2017
```