Homework 1

Instructor: Hedibert Freitas Lopes
Course: STP 598 Advanced Bayesian Statistical Learning (Class # 31199)
Semester: Spring 2022
Due date: 1:30pm, February 14th, 2022.

Bernoulli model. In class we talked about the Bernoulli model where, conditionally on θ, $x_1,\ldots,x_n$ are i.i.d. $Bernoulli(\theta)$, for $0 \leq \theta \leq 1$, such that $p(x_i|\theta) = \theta^{x_i}(1-\theta)^{1-x_i}$, for $x_i = 0,1$ and $i = 1,\ldots,n$. Let $s_n = \sum_{i=1}^n x_i$ be the number of successes out of n trials. It is relatively easy to show that $s_n|\theta \sim Binomial(n,\theta)$, i.e.

$$Pr(s_n|\theta) = \frac{n!}{k!(n-s_n)!}\theta^{s_n}(1-\theta)^{n-s_n}, \text{ for } s_n = 0,1,2,\ldots,n,$$

and that $s_n = \sum_{i=1}^n x_i$ is a sufficient statistics for θ. Obviously, the likelihood function is

$$L(\theta|n,s_n) \propto \theta^{s_n}(1-\theta)^{n-s_n},$$

which resembles the kernel of a Beta distribution with parametes s_n+1 and $n-s_n+1$. Let us now consider three prior specifications:

- Prior A: $\theta \sim Uniform(0,1)$ (this is actually Prior B with $a=b=1$),
- Prior B: $\theta \sim Beta(a,b)$ (below we will assume the values $a=4$ and $b=2$),
- Prior C: $\log(\theta/(1-\theta)) \sim Normal(\mu,\sigma^2)$ (below we will use $\mu=0$ and $\sigma^2=3$).

It is easy to show that $p(\theta|A) = 1$, $p(\theta|B) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}$, and that

$$p(\theta|C) = (2\pi\sigma^2)^{-1/2}\exp\left\{-\frac{1}{2\sigma^2}\left[\log\left(\frac{\theta}{1-\theta}\right)-\mu\right]^2\right\}\frac{1}{\theta(1-\theta)}.$$

When working on the following questions, you should assume that $n = 10$ Bernoulli trials lead to $s_{10} = 7$ successes.

a) Derive $p(\theta|s_n, A)$, which is a Beta distribution (just like the prior). This should be fairly easy!

b) Derive $p(\theta|s_n, B)$, which is also a Beta distribution. This should also be easy!

c) Derive $Pr(s_n|A)$ and $Pr(s_n|B)$. These quantities are known scalars that represents the predictive densities of the observed data under the Bernoulli model and Prior A or Prior B, respectively. Since both Prior A and Prior conjugate with the likelihood function, the easiest way to compute both predictive densities is

$$Pr(s_n|A) = \frac{Pr(s_n|\theta)p(\theta|A)}{p(\theta|s_n, A)} \qquad \text{and} \qquad Pr(s_n|B) = \frac{Pr(s_n|\theta)p(\theta|B)}{p(\theta|s_n, B)},$$

d) Derive $Pr(s_n|C)$ first and then $p(\theta|s_n, C)$. These are the complicated ones, as far as computation is concerned, since the prior density and likelihood function fail to conjugate. Therefore, you will need to approximate the denominator of

$$p(\theta|s_n, C) = \frac{p(s_n|\theta))p(\theta|C)}{\int_0^1 p(s_n|\theta)p(\theta|C)d\theta} = \frac{p(s_n|\theta))p(\theta|C)}{Pr(s_n|C)}$$

Approximate $Pr(s_n|C)$ by a simple fine grid for the interval $(0, 1)$. In a couple of classes, we will show that a simple Monte Carlo integration is equally easy and more suitable when the dimension of θ is greater than 2 or 3).

e) Compute $E(\theta|s_n, A)$ and $V(\theta|s_n, A)$. Repeat for Priors B and C.

f) Graphically compare the three priors and posteriors.