CSE 515T (Spring 2015) Assignment 1

Due Wednesday, 28 January 2015

1. (Barber.) Suppose that a study shows that 90% of people who have contracted Creutzfeldt–Jakob disease ("mad cow disease") ate hamburgers prior to contracting the disease. Creutzfeldt–Jakob disease is incredibly rare; suppose only one in a million people have the disease.

   If you eat hamburgers, should you be worried? Does this depend on how many other people eat hamburgers?

2. (O'Hagan and Forster.) Suppose x has a Poisson distribution with unknown mean θ:

$$p(x \mid \theta) = \frac{\theta^x}{x!}\exp(-\theta), \qquad x = 0, 1, \dots$$

   Let the prior for θ be a gamma distribution:

$$p(\theta \mid \alpha, \beta) = \frac{\beta^\alpha \theta^{\alpha-1}}{\Gamma(\alpha)}\exp(-\beta\theta), \qquad \theta > 0$$

   where Γ is the gamma function. Show that, given an observation x, the posterior $p(\theta \mid x, \alpha, \beta)$ is a gamma distribution with updated parameters $(\alpha', \beta') = (\alpha + x, \beta + 1)$.

3. (Optimal *Price is Right* bidding.) Suppose you have a standard normal belief about an unknown parameter θ, $p(\theta) = \mathcal{N}(\theta; 0, 1^2)$. You are asked to give a point estimate $\hat{\theta}$ of θ, but are told that there is a heavy penalty for guessing too high. The loss function is

$$\ell(\hat{\theta}, \theta; c) = \begin{cases} (\theta - \hat{\theta})^2 & \hat{\theta} < \theta; \\ c & \hat{\theta} \geq \theta \end{cases},$$

   where $c > 0$ is a constant cost for overestimating. What is the Bayesian estimator in this case? How does it change as a function of c?

4. (Maximum-likelihood estimation.) Suppose you flip a coin with unknown bias θ, $\Pr(x = \mathrm{H}) = \theta$, three times and observe the outcome HHH. What is the maximum likelihood estimator for θ? Do you think this is a good estimator? Would you want to use it to make predictions?

   Consider a Bayesian analysis of θ with a beta prior $p(\theta \mid \alpha, \beta) = \mathcal{B}(\theta; \alpha, \beta)$. What is the posterior mean of θ? What is the posterior mode? Consider $(\alpha, \beta) = (1/5, 1/5)$. Plot the posterior density in this case. Is the posterior mean a good summary of the distribution?

5. (Gaussian with unknown mean.) Let $\boldsymbol{x} = \{x_i\}_{i=1}^N$ be independent, identically distributed real-valued random variables with distribution $p(x_i \mid \theta) = \mathcal{N}(x_i; \theta, \sigma^2)$. Suppose the variance σ^2 is known but the mean θ is unknown with prior distribution $p(\theta) = \mathcal{N}(\theta; 0, 1^2)$.

   - What is the likelihood of the full observation vector $p(\boldsymbol{x} \mid \theta)$?
   - After observing $\boldsymbol{x}$, what is the posterior distribution of θ, $p(\theta \mid \boldsymbol{x}, \sigma^2)$? (Note: you might find it more convenient in this case to work with the *precision* $\tau = \sigma^{-2}$.)
   - Interpret how the posterior changes as a function of N. What happens if $N = 0$? What happens if $N \to \infty$? Does this agree with your intuition?

6. (Spike and slab priors.) Suppose θ is a real-valued random variable that is expected to either be near zero (with probability π) or to have a wide range of potential values (with probability $(1-\pi)$). Such scenarios happen a lot in practice: for example, θ could be the coefficient of a feature in a regression model. We either expect the feature to be useless for predicting the output (and have a value close to zero) or to be useful, in which case we expect a value with larger magnitude but can't say much else.

   A common approach in this case is to use a so-called *spike and slab prior.* Let $f \in \{0, 1\}$ be a discrete random variable serving as a flag. We define the following conditional prior:

$$p(\theta \mid f, \sigma^2_{\text{spike}}, \sigma^2_{\text{slab}}) = \begin{cases} \mathcal{N}(\theta; 0, \sigma^2_{\text{spike}}) & f = 0 \\ \mathcal{N}(\theta; 0, \sigma^2_{\text{slab}}) & f = 1, \end{cases}$$

   where σ_{spike} is the width of a narrow "spike" at zero, and $\sigma_{\text{slab}} > \sigma_{\text{spike}}$ is the width of a "slab" supporting values with larger magnitude.

   In practice, we will never observe the flag variable f; instead, we must infer it or marginalize it, as required.

   - Suppose we choose a prior $\Pr(f = 1) = \pi = 1/2$, expressing no *a priori* preference for the spike or the slab. What is the marginal prior $p(\theta \mid \sigma^2_{\text{spike}}, \sigma^2_{\text{slab}})$? Plot the marginal prior distribution for $(\sigma^2_{\text{spike}}, \sigma^2_{\text{slab}}) = (1^2, 10^2)$.
   - Suppose that we can make a noisy observation x of θ, with distribution $p(x \mid \theta, \omega^2) = \mathcal{N}(x; \theta, \omega^2)$, with known variance ω^2. Given x, what is the posterior distribution of the flag parameter, $\Pr(f = 1 \mid x, \sigma^2_{\text{spike}}, \sigma^2_{\text{slab}}, \omega^2)$? Plot this distribution as a function of x. What observation would teach us the most about f? What teaches us the least?
   - Given an observation x as in the last part, what is the posterior distribution of θ, $p(\theta \mid x, \sigma^2_{\text{spike}}, \sigma^2_{\text{slab}}, \omega^2)$? (Hint: use the sum rule to eliminate f and use the result above.)
   - Suppose the noise variance is $\omega^2 = 0.1^2$ and we make an observation $x = 3$. Plot the posterior distribution of θ, using the parameters from the first part.