Introduction to Bayesian Data Analysis

Tutorial 4

(1) Show that the posterior predictive distribution $p(\tilde{y}|y)$ is the posterior expectation of $p(\tilde{y}|\theta)$.

(2) For the birth rate example in class - repeat the posterior predictive checking exercise for women age-40 with bachelor degrees

(3) Problem 4.2 (Hoff) Reconsider the tumor count data in Tutorial 3.

(a) For the prior distribution given in part a) of that exercise, obtain $Pr(\theta_B < \theta_A|\mathbf{y_A}, \mathbf{y_B})$ via Monte Carlo Sampling

(b) For a range of values of n_0, obtain $Pr(\theta_B < \theta_A|\mathbf{y_A}, \mathbf{y_B})$ for $\theta_A \sim \text{Gamma}(120, 10)$ and $\theta_b \sim \text{Gamma}(12 \times n_0, n_0)$. Describe how sensitive the conclusions about the event $\{\theta_B < \theta_A\}$ are to the prior distribution on θ_B.

(c) Repeat parts a) and b) replacing the event $\{\theta_B < \theta_A\}$ with the event $\left\{\tilde{Y}_B < \tilde{Y}_A\right\}$ where $\tilde{Y}_A$ and $\tilde{Y}_B$ are samples from the posterior predictive distribution. Describe how sensitive the conclusions about the event $\left\{\tilde{Y}_B < \tilde{Y}_A\right\}$ are to the prior distribution on θ_B, and compare to your observations in part (b).

(4) Problem 4.3 (Hoff) Let's investigate the adequacy of the Poisson model for the tumor count data. Generate posterior predictive data sets $\mathbf{y}_A^{(1)}$,...,$\mathbf{y}_A^{(1000)}$. Each $\mathbf{y}_A^{(s)}$ is a sample of size $n_A = 10$ from the Poisson distribution with parameter $\theta_A^{(s)}$, $\theta_A^{(s)}$ is itself a sample from the posterior distribution $p(\theta_A|\mathbf{y}_A)$, and $\mathbf{y}_A$ is the observed data.

(a) For each s, let $t^{(s)}$ be the sample average of the 10 values of $\mathbf{y}_A^{(s)}$, divided by the sample standard deviation of $\mathbf{y}_A^{(s)}$. Make a histogram of $t^{(s)}$, and compare it to the observed value of this statistic. Based on this statistic, assess the fit of the Poisson model for these data.

(b) Repeat the above goodness of fit evaluation for the data in population B.

(5) Problem 4.4 (Hoff) From the posterior density from Problem (4) (Tutorial 3)

(a) Make a plot of $p(\theta|y)$ or $p(y|\theta)p(\theta)$ using the mixture prior distribution and a dense sequence of θ-values. Can you think of a way to obtain a 95% quantile-based posterior confidence interval for θ? You might want to try some sort of discrete approximation.

(b) To sample a random variable z from the mixture distribution $wp_1(z) + (1-w)p_0(z)$, first toss a $w - coin$ and let x be the outcome (this can be done in R with `x<-rbinom(1,1,w)`). Then if $x = 1$ sample z from p_1 and if $x = 0$ sample z from p_0. Using this technique, obtain a Monte Carlo approximation of the posterior distribution $p(\theta|y)$ and a 95% quantile-based confidence interval, and compare them to the results in part (a).

(6) Consider a univariate posterior distribution, $p(\theta|y)$, which we wish to approximate and then calculate moments of, using importance sampling from an unnormalized density, $g(\theta)$. Suppose the posterior distribution is normal, and the approximation is t_3 with mode and curvature matched to the posterior density.

(a) Draw a sample of size S=100 from the approximate density and compute the importance ratios. Plot a histogram of the log importance ratios.

(b) Estimate $E[\theta|y]$ and $Var[\theta|y]$ using importance sampling. Compare to the true values.

(c) Repeat (a) and (b) for $S = 10,000$.