STA 114: Statistics

HW 7

Due Wed Oct 26 2011

1. Consider scalar data $X = X_1$ and a future scalar variable $X^* = X_2$ modeled as $X_1, X_2 \overset{\text{IID}}{\sim} g(x_i|\theta)$, $\theta \in (-\infty, \infty)$ where the $g(y|\theta)$ is the following pmf: $g(y|\theta) = 0.5$ if $y = \theta - 1$ or $y = \theta + 1$, and $g(y|\theta) = 0$ otherwise. Describe the pmf $f^*(x^*|x)$ of X^* given $X = x$ under the prior $\xi(\theta) = \mathsf{Normal}(0, 1)$. [For any arbitrary x give the exact formula of $f^*(x^*|x)$ as a function of x^*.]

2. A soporific drug, given to a certain kind insomniac patients, enhances nightly sleep hours by an amount μ with some patient-to-patient variability σ. Suppose the drug is given to n such patients, and their 30-day average increase in nightly sleep hours are measured as $X = (X_1, \cdots, X_n)$, and modeled $X_i \overset{\text{IID}}{\sim} \mathsf{Normal}(\mu, \sigma^2)$. The model parameters (μ, σ^2) are assigned a prior pmf $\xi(\mu, \sigma^2)$ described by the following table ($\xi(\mu, \sigma^2) = 0$ otherwise):

μ	$\xi(\mu, 0.5^2)$	$\xi(\mu, 1)$	$\xi(\mu, 1.5^2)$	$\xi(\mu, 2^2)$
0.0	$\frac{1}{30}$	$\frac{2}{30}$	$\frac{2}{30}$	$\frac{1}{30}$
0.5	$\frac{2}{30}$	$\frac{2}{30}$	$\frac{2}{30}$	$\frac{2}{30}$
1.0	$\frac{3}{30}$	$\frac{2}{30}$	$\frac{2}{30}$	$\frac{2}{30}$
1.5	$\frac{2}{30}$	$\frac{1}{30}$	$\frac{1}{30}$	$\frac{0}{30}$
2.0	$\frac{2}{30}$	$\frac{1}{30}$	$\frac{0}{30}$	$\frac{0}{30}$

   (a) Give the posterior pmf $\xi(\mu, \sigma^2|x)$ based on observation $X = x$ for which $n = 10$, $\bar{x} = 1.6$ and $s_x = 1.23$.

   (b) Suppose the drug is to be given to a hypothetical future patient with similar conditions. Letting X^* denote his 30-day average increase in nightly sleep hours, calculate the probability $P(X^* > 1|X = x) = \int_1^\infty f^*(x^*|x)dx^*$ based on x as in part (a). State your model assumptions on X^*. [Hint. This is not a conjugate model, so don't copy the conjugate norma formulas from notes, they don't apply. And note $P(X^* > a|X = x) = \int_\Theta P(X^* > a|\theta)\xi(\theta|x)d\theta$.]

3. Consider data $X = (X_1, \cdots, X_n)$ and a future variable $X^* = X_{n+1}$ modeled as $X_1, \cdots, X_n, X_{n+1} \overset{\text{IID}}{\sim} \mathsf{Normal}(\mu, \sigma^2)$.

   (a) When σ^2 is fixed and $\mu \in (-\infty, \infty)$ is assigned the Jeffreys prior $\xi(\mu) = \text{const.}$, identify the central $100(1-\alpha)\%$ credible interval for the posterior predictive pdf $f^*(x^*|x)$ given $X = x$.

   (b) When $(\mu, \sigma^2) \in (-\infty, \infty) \times (0, \infty)$ is assigned the reference prior $\xi(\mu, \sigma^2) = \text{const.}/\sigma^2$, identify the central $100(1-\alpha)\%$ credible interval for the posterior predictive pdf $f^*(x^*|x)$ given $X = x$.

4. Suppose the duration (in seconds) of a smile of a certain eight week old baby follows a $\mathsf{Uniform}(0, \theta)$ distribution with $\theta > 0$ unknown. The data X consists of n observed durations $X_1, \cdots, X_n$ from the baby. We are interested in a future observation $X^* = X_{n+1}$.

   (a) Based on observations (10.4, 19.6, 12.8, 14.8, 1.3, 0.7, 5.8, 6.9, 8.9, 9.4), what is the probability that X^* would exceed 15 under the plug-in predictive pdf $\hat{f}^*(x^*|x)$ when θ is estimated by its MLE?

   (b) Based on the same observations, what is the probability that X^* would exceed 15 under the posterior predictive pdf $f^*(x^*|x)$ when θ is assigned a $\mathsf{Pareto}(2, 15)$ prior? [Hint. $P(X^* > a|X = x) = \int_\Theta P(X^* > a|\theta)\xi(\theta|x)d\theta$.]

5. Time gaps between successive eruptions of a geyser are modeled as independently distributed according to $\mathsf{Exponential}(\lambda)$ with $\lambda \in (0, \infty)$ assigned a $\mathsf{Gamma}(a, b)$ prior. Our data consists of n time gaps $X = (X_1, \cdots, X_n)$ and we are interested in a future time gap $X^* = X_{n+1}$.

   (a) Show that the posterior predictive pdf $f^*(x^*|x)$ of X^* equals

$$f^*(x^*|x) = \frac{a'(b')^{a'}}{(b' + x^*)^{a'+1}}, \text{ for } x^* > 0; \quad f(x^*|x) = 0 \text{ otherwise,}$$

   where $a' = a + n$ and $b' = b + n\bar{x}$ are the usual updates for the exponential-gamma conjugate model. [Hint. $\int_0^\infty \lambda^r \exp(-s\lambda)d\lambda = \Gamma(r + 1)/s^{r+1}$ and $\Gamma(r + 1) = r\Gamma(r)$.]

   (b) Show that the posterior predictive distribution of $Y^* = 1 + \frac{X^*}{b'}$ is a Pareto distribution. Identify its parameters.

6. To analyze how well undergraduates manage of pursue non-curricular interests and hobbies, data $X_1, \cdots, X_n$ will be recorded from n students, with X_i giving the weekly amount of time, averaged over last 5 weeks, the i-th student spent on such activities. Consider similar records Y_1, Y_2 from two hypothetical students randomly selected from the undergraduate population.

   (a) Quantify your personal beliefs about undergraduates' pursuit of such activities by answering the following questions on your belief about Y_1, Y_2

   - Give a number q_1 that you think Y_1 is equally likely to be smaller or larger than.
   - Suppose you were told that Y_1 exceeds q_1 of part (a). Give a number q_2 that you think Y_1 is equally likely to be smaller or larger than.
   - Suppose you were told that Y_1 exceeds q_2 of part (b). Give a number q_3 that you think Y_2 is equally likely to be smaller or larger than.
   - Give a number q_4 that you think $|Y_1 - Y_2|$ is equally likely to be smaller or larger than.

   (b) Are there numbers $-\infty < m < \infty, k > 0, r > 0, s > 0$ such that the Bayesian model $X_1, \cdots, X_n \overset{\text{IID}}{\sim} \mathsf{Normal}(\mu, \sigma^2)$, with prior $\xi(\mu, \sigma^2) = \mathsf{N}\chi^{-2}(m, k, r, s)$ is consistent with your belief quantification above? If not, explain why not. If yes, give values of m, k, r, s.