STA 250: Statistics

HW 4

Due Tue Oct 02 2013

1. A machine goes through 4 hazard levels θ, coded 0 through 3 (from low hazard to high hazard) with use over time. The hazard level can be measured by frequency of hazardous incidents X, again coded 0 through 3 (low frequency to high frequency). Suppose X is modeled with pmfs $f(x|\theta)$, $\theta \in \Theta = \{0, 1, 2, 3\}$ as given by the rows of the following table.

θ	$f(0\|\theta)$	$f(1\|\theta)$	$f(2\|\theta)$	$f(3\|\theta)$
0	$\frac{4}{10}$	$\frac{3}{10}$	$\frac{2}{10}$	$\frac{1}{10}$
1	0	$\frac{3}{6}$	$\frac{2}{6}$	$\frac{1}{6}$
2	0	0	$\frac{2}{3}$	$\frac{1}{3}$
3	0	0	0	1

   (a) For the discrete uniform prior pmf on θ, i.e., $\xi(0) = \xi(1) = \xi(2) = \xi(3) = 1/4$, and $\xi(\theta) = 0$ otherwise, calculate the posterior probability $P(\theta \in \{0, 1\} \mid x)$ for each of $x = 0, 1, 2, 3$.

   (b) The machine under consideration is old, and is a lot more likely to be in a high hazard level than a low one, though no levels can be ruled out a-priori. Give a prior pmf that is consistent with this belief and repeat (a) for this prior pmf.

2. Consider two observations $X_1, X_2 \overset{\text{IID}}{\sim} g(x_i|\theta)$, $\theta \in (-\infty, \infty)$, where $g(y|\theta)$ is the following pmf: $g(y|\theta) = 0.5$ if $y = \theta + 1$ or $y = \theta - 1$, and $g(y|\theta) = 0$ for all other y. Suppose θ is assigned the $\mathsf{Normal}(0, 1)$ prior pdf. Given observed data $x = (x_1, x_2)$, do we get a posterior pdf or a posterior pmf $\xi(\theta|x)$ for θ? Explain. Give precise expressions for $\xi(\theta|x)$ for an arbitrary x.

3. In handout 7 we saw that for the model $X \sim \mathsf{Binomial}(n, p)$, $p \in (0, 1)$ if we assign p the uniform prior pdf: $\xi(p) = \mathsf{Uniform}(0, 1)$ then the posterior is a beta pdf: $\xi(p|x) = \mathsf{Beta}(x + 1, n - x + 1)$. The $\mathsf{Uniform}(0, 1)$ is in fact same as the $\mathsf{Beta}(1, 1)$ pdf. Show that for any $a > 0, b > 0$, if we had used the prior $\xi(p) = \mathsf{Beta}(a, b)$ then the posterior will be $\xi(p|x) = \mathsf{Beta}(x + a, n - x + b)$.

4. In a survey of $n = 500$ randomly chosen college students, $X = 200$ said they support a certain federal policy. Let $p \in [0, 1]$ denote the fraction of all students in the college who support the policy.

(a) Consider three different prior beliefs about p:

i. "No reason to prefer one value of p over another"

ii. "The fact that the policy was long debated at the senate was perhaps because it split the general public roughly in 50-50"

iii. "Due to partisan effect, any given college is likely to have either large support or little support, but not much in-between";

and three possible choices for a prior pdf $\xi(p)$: $\mathsf{Beta}(0.5, 0.5)$, $\mathsf{Uniform}(0, 1) = \mathsf{Beta}(1, 1)$ and $\mathsf{Beta}(10, 10)$. Discuss which prior belief is best matched by which choice of $\xi(p)$ (make plots of the pdfs to make your point).

(b) For each choice of $\xi(p)$, evaluate the prior and posterior probabilities of $p > 0.5$ (i.e., policy more popular than not).

(c) For each choice of $\xi(p)$ present the prior and posterior 95% ranges for p.

5. The swing (difference between the maximum and minimum values) of S&P500 on any one day is distributed with pdf $f_1(x) = \frac{2}{20(1+x/20)^3}$, $x > 0$ if the market that day is "stable" or as $f_2(x) = \frac{2}{200(1+x/200)^3}$ if the market encounters high volatility on that day. At the beginning of Dec 14, 2012, a market analyst assigns 5% chance for that day to see high volatility. The swing that day is recorded to be 20. Suppose market state (stable/volatile) persists the next day with probability 0.8 and with probability 0.2 switches to the other state. What is the posterior predictive probability that on Dec 15, 2012 S&P500 will swing more than 20 points? [Hint: The two pdfs are of the form $\frac{2}{b(1+x/b)^3}$ and have cdfs of the form $1 - \frac{1}{(1+x/b)^2}$.]