MIDTERM EXAM

This exam is due at 9AM Saturday, 3/17. As usual, you are not to discuss the exam with other students until after the exam is due. The exam must be entirely your own work. Answer all questions.

(1) Below are data on x and y, drawn from a process where

$$y \mid x \sim N(f(x), 1)$$

and f is an unknown function. Suppose that our prior makes

$$f(x_1) - f(x_2) \sim N(0, |x_1 - x_2|)$$

and makes the covariance of $f(x_1) - f(x_2)$ with $f(x_3) - f(x_4)$ zero if (x_1, x_2) and (x_3, x_4) do not overlap. In other words, $f(x)$ behaves like a Wiener process. Treat the prior on $f(0)$ as $N(0, 10)$.

The Data

	x	y
1	1.00	3.00
2	2.00	7.00
3	5.00	8.00
4	6.00	5.00
5	8.00	1.00
6	11.00	2.00

(a) Calculate the posterior expected value, given the observed x and y data, of $f(x)$ at 20 equally spaced x values points between 0 and 12.
(b) Add to the plot $\pm$ one standard deviation bands for the distribution of f values at these 20 x points.
(c) Add to the plot $\pm$ one standard deviation bands for a newly drawn y value at each of the 20 x points. This will characterize the distribution of forecast errors from our estimated f function, and of course will be wider than the bands for f.
(d) Plot the data as individual points, again on the same plot.

(2) Suppose we have 100 observations on y drawn i.i.d. from the exponential distribution, i.e. with pdf

$$e^{-y} .$$

Suppose also that the sampling has been subject to selection with the probability of selection w distributed on (0,1) with some pdf $g(w)$.
(a) Show that the sample average of the ratios y_i/w_i has expectation 1.0 and converges to this value as the sample size goes to infinity.
(b) Show that if w is independent of y, the sample average of y_i itself, without any use of w_i in the averaging, has expectation 1.0 and converges to this value as sample size goes to infinity.

Date: March 16, 2018.

(c) Compare the variance of the sample average of y_i to that of the sample average of y_i/w_i, maintaining the assumption of independence of w and y. Consider two cases: $g(w) \equiv 1$ and $g(w) = 2w$.

(d) Now suppose that w takes on only two values, .25 and .5, with each value having probability .5. Calculate the variance of the sample mean of the 100 draws of y_i/w_i for the case where w_i is independent of y_i, and also when instead $w_i = .5$ for all $y_i > \log(2)$ and $w_i = .25$ for all $y_i < \log(2)$. (Note that for exponentially distributed y, $P[y < \log(2)] = .5$.)