Homework 8: Sampling Accidents

36-350, Fall 2011

Due at 11:59 pm on Thursday, 10 November 2011

You know the rules by now for turning in work.

It is common to model the time between events like industrial accidents, earthquakes, or e-mails with exponential distributions. That is, the probability that the time between two events is between x and $x + dx$ is $\approx \lambda e^{-\lambda x} dx$. A complication, however, is that the rate λ could itself change. One way to model this is to have λ be random. Suppose that λ follows a gamma distribution with scale a and shape 1. Then

$$p(\lambda) = \frac{\lambda^{a-1} e^{-\lambda}}{\Gamma(a)} \quad (1)$$

$$p(x|\lambda) = \lambda e^{-\lambda x} \quad (2)$$

$$p(x) = \int_0^\infty d\lambda \frac{\lambda^{a-1} e^{-\lambda}}{\Gamma(a)} \lambda e^{-\lambda x} \quad (3)$$

It would be convenient to infer the rate λ from an observation of x. By Bayes's rule, the conditional distribution is

$$p(\lambda|x) = \frac{p(x|\lambda)p(\lambda)}{p(x)} \quad (4)$$

1. (20) Write a function to draw n values of X from the distribution in Eq. 3. It should take as its inputs the number of samples, and the value of a, and return a vector of length n. *Hint:* use `rgamma` and `rexp`.

2. (30) Following the Metropolis algorithm from lecture, write a function to draw Monte Carlo samples from the distribution $p(\lambda|x)$ in Eq. 4. It should take three arguments: the observed value of x, the posited prior value of a, and the number of samples to generate.

   - Draw the initial value of λ from the prior distribution.
   - Draw proposals from a uniform distribution of width 1, centered on the current value of λ.
   - Use `dexp` and `dgamma` to calculate $p(x|\lambda)$ and $p(\lambda)$.
   - Return the complete sequence of λ values generated.

You do not need to know $p(x)$ for this problem, just $p(\lambda)$ and $p(x|\lambda)$.

3. (10) Do the integral in Eq. 3 to find $p(x)$. *Hint:* use change of variables, the fact that $\int_0^\infty t^{a-1}e^{-t}dt = \Gamma(a)$, and that $\Gamma(a+1)/\Gamma(a) = a$.

4. (10) Set $a = 10$ and draw 1000 values from your function in Problem 1. Plot the histogram of this sample and compare it to the $p(x)$ you calculated in Problem 3. Do they match? Should they?

5. (10) Using your answer from problem 3, write a function to calculate the posterior density $p(\lambda|x)$, according to Eq. 4.

6. (10) Set $a = 10$ and $x = 5$. Draw 2000 values from your Monte Carlo sampler from Problem 2. Discard the first 1000, and plot the histogram of the rest. Compare this to the curve from your function from the previous problem. Do they match? Should they?

7. (10) If $p(\lambda)$ were not a gamma distribution but some other probability law, what would you have to change in your code in Problems 1 and 2?