10-602: Statistical Approaches to Learning and Discovery

Problem Set 5
Date: March 22, 2001 **Due: start of class April 5, 2001**

---

1. This problem arises in Positron Emission Tomography (www.biomed.org/pet.html), a unique technology for diagnosing diseases and tumors. A radioactive substance is injected into the patient and we watch where it goes by counting the emissions and triangulating where they come from. Discretize the problem by dividing the patient's volume into J small boxes. The emissions from box j are assumed to follow a Poisson distribution with unknown rate λ_j. We have a ring of detectors around the patient. If box j emits a particle, there is a known probability p_{ij} that it will register on the ith detector, based on their relative position. We run the detectors for one unit of time get a total count n_i from each detector i. From this data $D = \{n_1, ..., n_I\}$ and the p_{ij} we want to determine the emission rates λ_j.

   (a) Given all p_{ij} and λ_j, what is the distribution of n_i?

   (b) Write down the log-likelihood for $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_J)$ given D. Show that this function is concave down in $\boldsymbol{\lambda}$ and therefore unimodal. Hint: $\log(x) - x$ is concave down.

   (c) Apply Jensen's inequality for the logarithm to get a lower bound on the log-likelihood. Use this bound to derive an EM algorithm for the MLE of $\boldsymbol{\lambda}$.

   (d) Introduce hidden variables n_{ij} such that $n_i = \sum_j n_{ij}$. The variable n_{ij} is the number of emissions detected at i which originated from box j. What is the distribution of n_{ij} given n_i and $\boldsymbol{\lambda}$, i.e. $p(n_{ij}|n_i, \boldsymbol{\lambda})$?

   (e) If we knew n_{ij}, the MLE of $\boldsymbol{\lambda}$ would be trivial. Use this fact to derive an EM algorithm for $\boldsymbol{\lambda}$ (the hidden variable method).

2. This problem compares importance sampling with different importance distributions (Tanner chap 1, exercise 5). The best importance distribution is not always the obvious one. Suppose y has a Cauchy distribution with density

$$p(y) = \frac{1}{\pi(1 + y^2)} \tag{1}$$

   We wish to estimate

$$\begin{aligned} p(y > 2) &= \int_{-\infty}^{\infty} \Theta(y - 2) p(y) dy && (2) \\ \Theta(x) &= \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases} && (3) \end{aligned}$$

The true value is 0.1476. Define $f(y) = \Theta(y-2)p(y)$. By sampling from $g(y)$, we can obtain the Monte Carlo estimate

$$\hat{I} = \frac{1}{n}\sum_{i=1}^{n} f(y_i)/g(y_i) \tag{4}$$

The standard deviation of $\hat{I}$ can be approximated by the sample standard deviation of $f(y)/g(y)$ divided by $\sqrt{n}$.

For each of the following three approaches, give $\hat{I}$ and its approximate standard deviation. Use $n = 10,000$ samples.

(a) $g(y) = p(y)$. You can sample from $g(y)$ using the transformation $y = \tan(\pi u)$, where $p(u) \sim U(0,1)$.

(b) $g(y) = 2p(y)\Theta(y)$. This is a half-Cauchy distribution. You can sample from $g(y)$ using the transformation $y = |\tan(\pi u)|$.

(c) $g(y) = 2y^{-2}\Theta(y-2)$. This is a Pareto distribution. You can sample from $g(y)$ using the transformation $y = 2/u$.

3. In this problem you will do changepoint analysis as in Tanner sec 6.2.3. Download the British coalmining data from the course website. The model is that the data up to time k had distribution $y_i \sim Po(\lambda_1)$ and after time k had distribution $y_i \sim Po(\lambda_2)$.

   (a) A good place to start is to plot the data. Does it look like there was a sudden change in the Poisson rate?

   (b) What is the joint MLE for $(k, \lambda_1, \lambda_2)$?

   (c) For the Bayesian analysis, use the priors $(j = 1, 2)$

$$p(\lambda_j) \sim \mathcal{G}(a_j, b_j) = \frac{\lambda_j^{a_j-1}}{\Gamma(a_j)b_j^{a_j}}\exp(-\frac{\lambda_j}{b_j}) \tag{5}$$

$$p(b_j) \sim \mathcal{IG}(c_j, d_j) = \frac{d_j^{c_j}}{\Gamma(c_j)b_j^{c_j+1}}\exp(-d_j/b_j) \tag{6}$$

   Note that these are different parameterizations than used by Tanner. Find the conditionals needed for Gibbs sampling.

   (d) Starting from the MLE for k, run the Gibbs sampler for T=10 iterations. Use $a_j = 1, c_j = 1, d_j = 1$. You will need code to sample from a Gamma distribution and from an arbitrary discrete distribution.

   (e) Plot the Monte Carlo approximation to the posterior for k:

$$p(k|D) = \frac{1}{T}\sum_{t=1}^{T} p(k|\lambda_1^t, \lambda_2^t, D) \tag{7}$$

   This is simply the average of the conditionals for k that you computed during Gibbs sampling. Does it look like figure 6.2 in Tanner? Is it stable over multiple runs? How does it compare to making a histogram of k over the 10 samples?