

Stat 543 Assignment 5 (due Monday March 7, 2016)
E-M Algorithm, (SEL) Bayes Estimation, Rejection Algorithm, Gibbs Sampling, Metropolis-Hastings Algorithm

1. Problems: 2.2.30, 2.2.35, 2.2.39 B&D
2. Consider a problem where observations Y_1, Y_2, \dots, Y_n are iid with mixture density

$$f(y|\alpha, \theta_1, \theta_2) = \alpha\phi(y - \theta_1) + (1 - \alpha)\phi(y - \theta_2)$$

(where ϕ is the standard normal pdf). Detail an EM algorithm for finding a maximum likelihood estimate for $(\alpha, \theta_0, \theta_1)$. (To do this, you may consider a model for $\mathbf{X} = (W, Y)$ where $W \sim \text{Ber}(p)$ and conditional on W , $Y \sim \text{N}(\theta_W, 1)$. For $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ iid with this distribution, apply the logic described in class to $S(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n) = (Y_1, Y_2, \dots, Y_n)$. What, for example, is the conditional probability computed under $(\alpha, \theta_0, \theta_1)_0$ that $W_i = 1$ given (Y_1, Y_2, \dots, Y_n) ?)

3. Write out an E-M algorithm for finding the MLE of \mathbf{p} based on all data collected in the three studies mentioned in the 3rd problem of Assignment 1.
4. Problems 3.2.1, 3.2.2, 3.2.3, 3.2.4 B&D
5. Suppose that I am interested in a continuous distribution with pdf $f(x) = K\phi(x)\sin^2(x)$ for $\phi(x)$ the standard normal pdf and K some appropriate normalizing constant (that to my knowledge can not be evaluated "in closed form" but that MathCad indicates is no more than 2.32).

- (a) Carefully and completely describe a rejection sampling algorithm for generating X from this distribution.
- (b) Then consider the matter of finding the mean and variance for this distribution, and $P[.3 < X < 1.2]$. Describe how you would use 10,000 values simulated as in a) to evaluate these.

6. A 10-dimensional random vector \mathbf{X} has a complicated (joint) pdf that is proportional to the function

$$h(\mathbf{x}) = \begin{cases} \exp\left(\prod_{i=1}^{10} \sin^2(x_i)\right) & \text{if each } x_i \in (0, 1) \\ 0 & \text{otherwise} \end{cases}$$

Notice that $h(\mathbf{x}) \leq e^1$.

- (a) Carefully describe an algorithm (using only iid Uniform $(0, 1)$ random variables) for generating a (10-dimensional) realization of \mathbf{X} .
- (b) Describe a simulation-based method of approximating EX_1X_2 .

7. Consider the following model. Given parameters $\lambda_1, \dots, \lambda_N$ variables X_1, \dots, X_N are independent Poisson variables, $X_i \sim \text{Poisson}(\lambda_i)$. M is a parameter taking values in $\{1, 2, \dots, N\}$ and if $i \leq M$, $\lambda_i = \mu_1$, while if $i > M$, $\lambda_i = \mu_2$. (M is the number of Poisson means that are the same as that of the first observation.) With parameter vector $\theta = (M, \mu_1, \mu_2)$ belonging to $\Theta = \{1, 2, \dots, N\} \times (0, \infty) \times (0, \infty)$ we wish to make inference about M based on X_1, \dots, X_N in a Bayesian framework.

As matters of notational convenience, let $S_m = \sum_{i=1}^m X_i$ and $T = S_N$.

- (a) If, for example, a prior distribution G on Θ is constructed by taking M uniform on $\{1, 2, \dots, N\}$ independent of μ_1 exponential with mean 1, independent of μ_2 exponential with mean 1, it is possible to explicitly find the (marginal) posterior of M given that $X_1 = x_1, \dots, X_N = x_N$. Don't actually bother to finish the somewhat messy calculations needed to do this, but show that this is possible (indicate clearly why appropriate integrals can be evaluated explicitly).
- (b) Suppose now that G is constructed by taking M uniform on $\{1, 2, \dots, N\}$ independent of (μ_1, μ_2) with joint density $g(\cdot, \cdot)$ on $(0, \infty) \times (0, \infty)$. Describe in as much detail as possible a "Gibbs Sampling" method for approximating the posterior of M , given $X_1 = x_1, \dots, X_N = x_N$. (Give the necessary conditionals up to multiplicative constants, say how you're going to use them and what you'll do with any vectors you produce by simulation.)
8. Consider the simple two-dimensional discrete (posterior) distribution for $\theta = (\theta_1, \theta_2)$ given in the table below.

		θ_2			
		1	2	3	4
θ_1	4	0	0	.2	.1
	3	0	0	.05	.05
	2	.2	.1	0	0
	1	.2	.1	0	0

- (a) This is obviously a very simple set-up where anything one might wish to know or say about the distribution is easily derivable from the table. However, for sake of example, suppose that one wished to use SSS to approximate $Q = P[\theta_1 \leq 2] = .6$. Argue carefully that Gibbs Sampling will fail here to produce a correct estimate of Q . In qualitative terms, what is it about this joint distribution that causes this failure?
- (b) Consider the very simple version of the Metropolis-Hastings algorithm where for each i , $J_i(\theta' | \theta)$ is uniform on the 8 vectors θ' where the posterior probability in the table above is positive. Make out an 8×8 table giving the conditional probabilities that (under the Metropolis-Hastings algorithm) $\theta_i^* = \theta'$ given $\theta_{i-1}^* = \theta$. If in fact θ_{i-1}^* has the distribution in the table, show that θ_i^* also has this distribution. (This is the important fact that the algorithm has the target distribution as its "stationary distribution.")