Homework 4

Handed out: Wednesday, October 10, 2018
Due: Wednesday, October 24, 2018 Midnight

Notes:

- We *highly* encourage typed (Latex or Word) homework. Compile as single report containing solutions, derivations, figures, etc.
- Submit all files including report pdf, report source files (e.g. .tex or .docx files), data, figures produced by computer codes and programs files (e.g. .py or .m files) in a **.zip** folder. Programs should include a Readme file with running instructions.
- Zipped folder should be turned in on Sakai with the following naming scheme: **HW4_LastName_FirstName.zip**
- Collaboration is allowed however all submitted reports, programs, figures, etc. should be an individual student's write ups. Direct copying could be considered cheating.
- Software resources for this Homework set can be downloaded from this link or on Sakai under the Resource folder.

Problem 1-15pts (Accept-Reject)

Generate samples of a standard normal distribution, $f(x) \sim \mathcal{N}(0,1)$, using the accept-reject method with a double-exponential proposal distribution, $g(x|\alpha) = (\alpha/2)\exp(-\alpha|x|)$.

(a) Derive the upper bound for the likelihood ratio, $M = f(x)/g(x)$ and show that the ideal acceptance rate is obtained when $\alpha = 1$.

(b) Implement the accept-reject method and plot the true PDF and the proposal distribution for $\alpha = 1$ super-imposed on to the normalized histogram of your samples.

(c) Repeat part (b) but now use a *sub-optimal* proposal distribution with $\alpha = 2$, plot both distributions and your histogram. How do the acceptance rates compare?

Problem 2-10pts (Independent Metropolis-Hastings)

Traditionally in the Metropolis-Hastings algorithm the arbitrary proposal distribution is conditioned on the current state of the chain. Namely, one draws samples from $\boldsymbol{x}' \sim q(\boldsymbol{x}'|\boldsymbol{x}_t)$ where $\boldsymbol{x}_t$ indicates the state of the chain. Consider a proposal distribution that is independent of the chain's current state $q(\boldsymbol{x}')$. When such a distribution is used, this is referred to as the *Independent Metropolis-Hastings algorithm.*

Prove that the Independent Metropolis-Hastings accepts more than the Accept-Reject method when both have identical target ($f(\boldsymbol{x}')$) and proposal ($g(\boldsymbol{x}')$) distributions.

Problem 3-25pts (Accept-Reject & Metropolis-Hastings)

(a) Implement the *accept-reject* algorithm to calculate the mean of a gamma distribution $\mathcal{G}(4.3, 6.2)$ using a $\mathcal{G}(4, 7)$ candidate. Draw the true density function on top of the sampled histogram and plot the convergence.

(b) Implement the *Metropolis-Hastings* algorithm to calculate the mean of a gamma distribution $\mathcal{G}(4.3, 6.2)$ using the following candidate densities:

- A gamma $\mathcal{G}(4, 7)$ candidate distribution.
- A gamma $\mathcal{G}(5, 6)$ candidate distribution.

For both candidate distributions draw the true and candidate density functions on top of the sampled histogram. Plot the convergence using each candidate distribution on the same axis. How do the means compare?

Problem 4-25pts (Gibbs & Metropolis-Hastings)

Consider sampling from a 2D Gaussian. Suppose $x \sim \mathcal{N}(\mu, \Sigma)$ where $\mu = (1, 1)$ and $\Sigma = (1, -0.5; -0.5, 1)$.

(a) Derive the full conditional p($x_1|x_2$) and p($x_2|x_1$) . Implement the Gibbs algorithm for this case and plot the 1D marginals p(x_1) and p(x_2) as well as (superimposed) the computed histograms.

(b) Let us now consider block-wise Metropolis Hastings. For our proposal distribution, q(x) let us use a normal centered at the previous state/sample of the Markov chain/sampler, i.e: $q(x|x^{(t-1)}) \sim \mathcal{N}(x^{(t-1)}, I)$, where I is a 2-D identity matrix. Show the 2D target distribution and its sampled approximation.

(c) We now consider component-wise Metropolis Hastings approximation of the same problem. The proposal distribution q(x) is now a univariate Normal distribution with unit variance in the direction of the i-th dimension to be sampled. Show the sampled and exact target distribution.

Show your results and compare the convergence with that obtained with the block-wise, component-wise Metropolis-Hastings and Gibbs implementation.

Problem 5-25pts (Metropolis-Hastings)

Consider the braking data of Tukey (1977) (see also Chapter 7 of "Monte Carlo Statistical Methods" by C. Robert[1]). It corresponds to breaking distances $y_{i,j}$ of cars driving at speeds

[1] Christian Robert and George Casella. *Monte Carlo statistical methods*. Springer Science & Business Media, 2013. URL: `https://www.springer.com/us/book/9780387212395`.

x_i . It is thought that a good model for this dataset is a quadratic model:

$$y_{i,j} = \beta + \beta_1 x_i + \beta_2 x_i^2 + \epsilon_{i,j}, \tag{1}$$

where $\epsilon_{i,j} \sim \mathcal{N}(0, \sigma^2)$, $i = 1, ..., k$ and $j = 1, ..., n_i$.

If we assume that $\epsilon_{i,j} \sim \mathcal{N}(0, \sigma^2)$ are independent, then the likelihood function is:

$$\left(\frac{1}{\sigma^2}\right)^{(N/2)} e^{-\frac{1}{2\sigma^2}\sum_{i,j}(y_{ij}-\beta_0-\beta_1 x_i-\beta_2 x_i^2)^2}, \tag{2}$$

where $N = \sum_i n_i$. We can view this likelihood as a posterior distribution of $\beta_0, \beta_1, \beta_2, \sigma^2$ and we can sample from it with a Metropolis-Hastings algorithm.

(a) Obtain maximum likelihood estimates for $\beta_0, \beta_1, \beta_2, \sigma^2$

(b) Use the estimates to select a candidate distribution. Take normal for $\beta_0, \beta_1, \beta_2$, and inverted Gamma for σ^2

(c) Make histograms of the posterior distributions of the parameters. Monitor convergence.

Robustness considerations could lead to using an error distribution with heavier tails. If we assume that $\epsilon_{i,j} \sim \mathcal{T}(0, \sigma^2)$ independent, then the likelihood function is:

$$\left(\frac{1}{\sigma^2}\right)^{(N/2)} \prod_{i,j} \Big(1 + \frac{1}{\nu}\frac{(y_{ij} - \beta_0 - \beta_1 x_i - \beta_2 x_i^2)^2}{\sigma^2}\Big)^{(\nu+1)/2} \tag{3}$$

where ν are the degrees of freedom. For $\nu = 4$, use Metropolis-Hastings to sample $\beta_0, \beta_1, \beta_2, \sigma^2$ from the posterior distribution. Use either normal or $\mathcal{T}$ candidates for $\beta_0, \beta_1, \beta_2$, and inverted Gamma or half-$\mathcal{T}$ for σ^2