1. This sequence of problems illustrates a case where we can break up a Markov chain into i.i.d.
components and use those components to analyze the chain. In a perfect world, we would be
able to use these sorts of techniques (called regenerative simulation) on every problem involving
MCMC. A second point of the exercise is to show one way to implement a rejection sampler
without knowing the exact suprema.

(a) Prove that if $f$ is a target density and $g$ is a candidate density and the supremum of $f/g$, say $C$, is finite (but unknown), then implementing rejection sampling with any positive constant $C$ yields i.i.d. variates from a density $\tilde{f}(x) \propto \min\{f(x), Cg(x)\}$. If $C$ happens to be greater than or equal to the exact supremum, $\sup f / g$, then $\tilde{f} = f$.

Tierney (1994 Annals of Statistics) proposed the use of $\tilde{f}$ as the candidate for an independence Metropolis algorithm. That is, a $C$ is chosen, candidates are drawn from $\tilde{f}$ and then these candidates are used in an independence Metropolis algorithm with invariant density $f$. Tierney calls this chain “a rejection sampling (independence Metropolis) chain”. In this homework, we will consider this chain.

(b) Let $S = \{x : f(x) \leq Cg(x)\}$. Let $y$ be a candidate generated from $\tilde{f}$ and let $x$ be the
current state of our Markov chain. Show the Metropolis acceptance probability for the
rejection sampling Metropolis chain is

\[
\alpha(x, y) = \begin{cases} 
1 & \text{for } x \in S, \\
\frac{c(y)}{f(x)} & \text{for } x \notin S, y \in S, \\
\min \left\{ \frac{f(y)g(x)}{f(x)g(y)}, 1 \right\} & \text{for } x \notin S, y \notin S
\end{cases}
\]

(c) A Markov chain is said to have “an atom”, say \(A\), if the transition kernel for the Markov chain satisfies

\[
k(x, y) = \nu(y)
\]

for any \(x \in A\). Argue that the \(S\) from the previous question is, in fact an atom.

The next couple of questions do not depend on our chain, instead they apply to chains that have an atom\(^1\). We let \(h\) be some function, and we are interested in \(\theta = E_f[h(X)]\). Assume our chain is started at stationarity.

(d) Let \(T_i\) be the indices where the chain enters \(S\). That is, \(X_{T_1}\) is the first point in \(S\), \(X_{T_2}\) is the second, and so on. Let \(B_i = \sum_{j=T_{i-1}+1}^{T_i} h(X_j)\) and \(N_i = T_i - T_{i-1}\). Here the \(B_i\) are sums between returns to \(S\) and the \(N_i\) is the length of time until the next return to \(S\). Argue (I’m not requiring formal proof) that the pairs \((B_i, N_i)\) are i.i.d. From here on out, we assume the \(B_i\) and \(N_i\) have finite second moments.

(e) If \(i = 1, \ldots, n\) argue that \(\hat{\theta}_n = \frac{\sum_{i=1}^{n} B_i}{\sum_{i=1}^{n} N_i}\) converges to \(E_f[h(x)] = \theta\) as \(n \to \infty\). You can assume “Wald’s Equation” which you might find useful. That is

\[
E[B_i] = E[N_i]E[h(X)]
\]

(f) Argue that

\[
\sqrt{n}(\hat{\theta} - \theta)
\]

converges to a \(N(0, \sigma^2)\) distribution and that a consistent estimate of \(\sigma^2\) is

\[
\frac{n \sum_{i=1}^{n}(B_i - \hat{\theta}_n N_i)^2}{(\sum_{i=1}^{n} N_i)^2}
\]

Notice the significance of what you proved. In parts 3 - 5 you showed that Markov chains that have an atom (and satisfy other regularity conditions we have swept under the rug) have an easy estimate of the Monte Carlo variance. In parts 1-2 you exhibited a specific chain that has an atom. In other words, we know how to estimate Monte Carlo error for this particular chain.

\(^1\)More generally, they apply to chains where you can locate “regeneration times”. See Mykland Tierney and Yu (JASA 1995)