Let $D = \{X_i : i \in \mathbb{N}\}$ be a data set of iid sequence from a random generator of distribution $\mathcal{N}(b, \sigma^2)$ for some unknown parameters b and σ. Our goal is to estimate b using so called stochastic approximation (SA) with a given learning rate $\alpha \in (0, 1)$:

- initialize b_0
- iterate $b_{n+1} = b_n + \alpha(x_k - b_k)$.

We want to examine the convergence $b_n \to b$. For simplicity, let's fix $\alpha = 0.01$.

1. Write pseudocode for SA.
2. Generate a data set $D = \{x_i : 1 \le i \le 10000\}$ with $\mathcal{N}(1, 4)$.
3. Write a code to implement SA on D, and demonstrate $b_n \to b$ as $n \to \infty$.
4. Prove that $\lim_n \mathbb{E} b_n = b$.
5. Can you prove or disprove that $b_n \to b$ in L^2?

Solution

1.
2.
3.
4. By taking expectation on both sides, it's enough to show that

   $$y_{n+1} = y_n(1 - \alpha),$$

   where $y_n = \mathbb{E}[b_n] - b$.
5. Let

   $$\epsilon_n = b_n - b.$$

   If we have L^2 convergence, then it means $\|\epsilon\|_2 \to 0$. Since, we can write

   $$\epsilon_{n+1} - (1 - \alpha)\epsilon_n = \alpha(x_k - b),$$

   this implies that

   $$\|\alpha(x_k - b)\|_2 = \|\epsilon_{n+1} - (1 - \alpha)\epsilon_n\|_2 \le 2\|\epsilon_{n+1}\|_2 + 2(1 - \alpha)^2\|\epsilon_n\|_2 \to 0.$$

   Hence, we have contradiction

   $$0 < \alpha^2\sigma^2 = \|\alpha(x_k - b)\|_2 \le 0.$$