An Application: Stochastic approximation and stochastic gradient descent (Robbins, Monro, 1951!)

Suppose I want to solve a nonlinear equation

\[ F(x) = 0 \quad x \in \mathbb{R} \]

There is only one solution

will assume \( F(x) > 0 \) for \( x > x^* \)
\( F(x) < 0 \) for \( x < x^* \).

Idea

\[ x_{n+1} = x_n - \alpha_n F(x_n) \]

stepsize
Robbins - Monro

Instead of $F(x_n)$ we can compute a r.v. $Z_{n+1}$ and perform

\[ X_{n+1} = X_n - a_n Z_{n+1}, \quad X_n = x_0 \]

\[ E[Z_{n+1} | X_0^n] = F(X_n) \]

Idea:

\[ X_{n+1} = X_n - a_n F(X_n) + \text{noise} \]

Does $X_n \to \star$ as $n \to \infty$?

Idea: Will assume $\| Z_n \|_C < C$ a.s.

\[ V_n := (X_n - x_\star)^2 \]

wts $V_n \xrightarrow{a.s.} 0$. 
\[ E[V_{n+1} | X_0] = E \{ (X_n - q_n Z_{n+1} - x) \} | X_0 \]
\[ = E \{ V_n - 2q_n z_{n+1} (X_n - x) + q_n z_{n+1}^2 | X_0 \} \]
\[ = V_n - 2q_n (X_n - x) f(X_n) + q_n E(Z_{n+1}^2 | X_0) \]
\[ \leq V_n - 2q_n |X_n - x| f(X_n) + q_n C^2 \]

want to get a sup MG.

\[ Y_n = V_n - C^2 \sum_{k=0}^{n-1} a_k \]

\[ E(Y_{n+1} | X_0) \leq Y_n \]

Assume \[ \sum_{k=0}^{\infty} Q_k < \infty \]

\[ Y_n \geq -C^2 \sum_{k=0}^{\infty} Q_k = -M \]
MG convergence

\[ Y_n \xrightarrow{a.s.} Y_0 \]

\[ V_n = Y_n + C^2 \sum_{k=0}^{n-1} A_k^2 \]

\[ V_n \xrightarrow{a.s.} V_0 \quad \text{wts} \quad V_0 = 0 \ a.s. \]

Also note that

\[ V_{n+1} - V_n = -2a_n (x_n - x_*) Z_{n+1} + a_n^2 Z_{n+1}^2 \]

\[ 2 \sum_{k=n}^{\infty} a_k (x_k - x_*) Z_{k+1} = V_n - V_0 + \sum_{k=n}^{\infty} a_k Z_{k+1}^2 \]

\[ \lim_{n \to \infty} \sum_{k=n}^{\infty} a_k (x_k - x_*) Z_{k+1} = 0 \quad (\#) \]
Suppose I can take expectation

$$\lim_{n \to \infty} \sum_{k=n}^{\infty} c_k E(X_k - x) f(x_k) = 0$$

$$\Rightarrow$$ if $$\sum_{k=1}^{\infty} c_k = 0$$

$$\Rightarrow$$

$$\sum_{k=1}^{\infty} \lim_{k \to \infty} E(X_k - x) f(x_k) = 0$$

[ In reality need to use (#) without taking expect ]

Where is this type of alg used?

ML.
minimize \[ U(x) = \frac{1}{n} \sum_{i=1}^{n} U_i(x) \]

equivalently solve

\[ f(x) = 0 \]

\[ f(x) = \nabla U(x) = \frac{1}{n} \sum_{i=1}^{n} f_i(x) \]

\[ f_i(x) = \nabla U_i(x) \]

Algorithm

\[ x_0 = x_0 \]

For \( k > 0 \), draw \( I_k \sim \text{Unif}(\{1, \ldots, n\}) \)

\[ x_{k+1} = x_k - \alpha_k f_{I_k}(x_k) \]

\[ Z_{k+1} \]