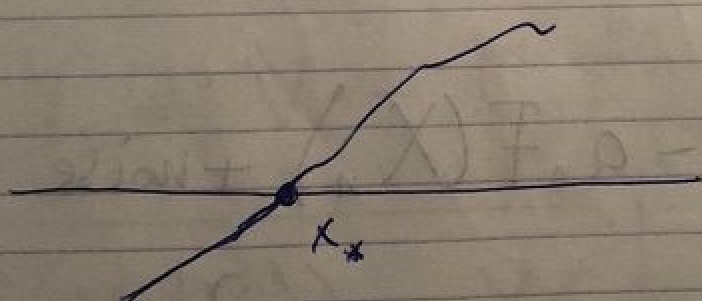
An Application: Stochastic approximation and stochastic gradient descent

(Robbins, Monro, 1951 !)

Suppose I want to solve a nonlinear equation

$$F(x) = 0 \qquad x \in \mathbb{R}$$



only one solution

will assume $F(x) > 0$ for $x > x_*$

$F(x) < 0$ for $x < x_*$

Idea

$$x_{n+1} = x_n - a_n F(x_n)$$

stepsize

Robbins - Monro

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

$$X_{n+1} = X_n - a_n Z_{n+1} \qquad X_n = x_0$$

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

idea:

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

Does $X_n \to x_*$ as $n \to \infty$?

Will assume $|Z_n| \leq C$ a.s.

$$V_n := (X_n - x_*)^2$$

wts $V_n \xrightarrow{a.s.} 0$.

$$\mathbb{E}[V_{n+1} | X_0^n] = \mathbb{E}\{(X_n - a_n Z_{n+1} - x_*)^2 | X_0^n\}$$

$$= \mathbb{E}\{V_n - 2a_n Z_{n+1}(X_n - x_*) + a_n^2 Z_{n+1}^2 | X_0^n\}$$

$$= V_n - 2a_n (X_n - x_*) F(X_n) + a_n^2 \mathbb{E}(Z_{n+1}^2 | X_0^n)$$

$$\leq V_n - 2a_n |X_n - x_*| |F(X_n)| + a_n^2 C^2$$

want to get a sup MG.

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

$$\boxed{\mathbb{E}(Y_{n+1} | X_0^n) \leq Y_n}$$

Assume $\boxed{\sum_{k=0}^{\infty} a_k^2 < \infty}$

$$Y_n \geq -C^2 \sum_{k=}^{\infty} a_k^2 =: -M$$

MG convergence

$$Y_n \xrightarrow{a.s.} Y_\infty$$

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

$$V_n \xrightarrow{a.s.} V_\infty \qquad \text{wts } V_\infty = 0 \text{ a.s.}$$

Also note that

$$V_{n+1} - V_n = -2a_n \left(X_n - \cancel{a}.\, x_*\right) Z_{n+1} + a_n^2 Z_{n+1}^2$$

$$2 \sum_{k=n}^{\infty} a_k (X_k - x_*) Z_{k+1} = V_n - V_\infty + \sum_{k=n}^{\infty} a_k Z_{k+1}^2$$

$$\lim_{n \to \infty} \sum_{k=n}^{\infty} a_k (X_k - x_*) Z_{k+1} = 0 \qquad (\#)$$

Suppose I can take expectation

$$\lim_{n\to\infty} \sum_{k=n}^{\infty} a_k \mathbb{E}(X_k - x_*) F(X_k) = 0$$

$$\Rightarrow \text{ if } \boxed{\sum_{k=1}^{\infty} a_k = \infty}$$

$$\Rightarrow \sum_k \lim_{k\to\infty} \mathbb{E}(X_k - x_*) F(X_k) = 0$$

$$X_k \to x_* \quad \boxed{X_k \xrightarrow{P} x_*}$$

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

Where is this type of alg used?

ML.

minimize $\quad U(x) = \frac{1}{n}\sum_{i=1}^{n} U_i(x)$

example i

$X_0 = x$

Draw

equivalently solve

$$F(x) = 0$$

$$F(x) = \nabla U(x) = \frac{1}{n}\sum_{i=1}^{n} F_i(x)$$

$$F_i(x) = \nabla U_i(x)$$

<u>Algorithm</u>

$$X_0 = x_0$$

for $k \geq 0$, draw $I_k \sim \mathrm{Unif}(\{1, \ldots n\})$

$$X_{k+1} = X_k - a_k \underbrace{F_{I_k}(X_k)}_{Z_{k+1}}$$