Kiefer-Wolfowitz algorithm with discontinuities in the parameters

Kinga Tikosi[1] Miklós Rásonyi[2]

[1]Central European University, Budapest

[2]Alfréd Rényi Institute of Mathematics, Budapest

12th European Summer School in Financial Mathematics, Padova, September 2 - 6, 2019

Problem formulation

- Goal: maximize the function $U : \mathbb{R}^d \to \mathbb{R}$
- Suppose we cannot see the fuctuion $U(\theta)$, only take 'noisy measurements', i.e. observe $J(\theta, X)$, where $E[J(\theta, X)] = U(\theta)$.
- Idea: take two measurements to estimate the gradient

Algorithm by J. Kiefer and J. Wolfowitz (1952):

$$\theta_{k+1} = \theta_k + a_k \frac{J(\theta_k + c_k, X) - J(\theta_k - c_k, X)}{2c_k}$$

Kiefer-Wolfowitz Algorithm

Algorithm:

$$\theta_{k+1} = \theta_k + a_k \frac{J(\theta_k + c_k, X) - J(\theta_k - c_k, X)}{2c_k},$$

where (a_k) and (c_k) are positive real sequences such that

$$c_k \to 0$$

$$\sum_{k=0}^{\infty} a_k c_k < \infty$$

$$\sum_{k=0}^{\infty} a_k = \infty$$

$$\sum_{k=0}^{\infty} a_k^2 c_k^{-2} < \infty,$$

e.g. $a_k = k^{-1}$ and $c_k = k^{-\gamma}$ with $\gamma \in (0, \frac{1}{2})$.

Previous results

- For the Robbins-Monroe algorithm the convergence rate in mean is $n^{-1/2}$
- Kiefer-Wolfowitz (1952) convergence in probability
- Convergence in expectation: $n^{-1/3}$ for twice continuously differentiable J, Burkholder (1956)
- a.s. convergence for discontinuous J by S. Laruelle (2011)

Result

Theorem

Under the following assumptions, θ_k converges to the global maximizer θ^* and the best convergence rate is $E|\theta_n - \theta^*| = O(n^{-1/5})$.

Result

Theorem

Under the following assumptions, θ_k converges to the global maximizer θ^* and the best convergence rate is $E|\theta_n - \theta^*| = O(n^{-1/5})$.

1. U is continuously differentiable and its gradient is Lipschitz-continuous; U has a unique maximizer θ^*, where $\nabla U(\theta^*) = 0$;
2. The ODE $\dot{y}_t = \frac{a}{t}\nabla U(y_t)$ fulfills $\left\| \frac{\partial y(t,s,\xi)}{\partial \xi} \right\| \leq C^* \left(\frac{s}{t}\right)^{\alpha}$ for every $\xi, 0 < s < t$, and for some positive reals C^* and α, and the solution trajectories converge to θ^*.
3. The process $J(\theta, X_k)$ is uniformly conditionally L-mixing;
4. $J(\theta, X_k)$ is conditionally locally Lipschitz with polynomially growing Lipschitz-constant,

Assumptions

L-mixing

Let $Y_n(\theta), n \in$ be an L^r-bounded random field for some $r \geq 1$. For all $n \in \mathbb{N}$ define

$$M_r^n(Y) = \operatorname*{ess\,sup}_{\theta} \sup_{k \in \mathbb{N}} \mathbb{E}^{1/r}[|Y_{n+k}(\theta)|^r \, |\mathcal{F}_n],$$

$$\gamma_r^n(\tau, Y) = \operatorname*{ess\,sup}_{\theta} \sup_{k \geq \tau} \mathbb{E}^{1/r} \left[\left| Y_{n+k}(\theta) - \mathbb{E}[Y_{n+k}(\theta) | \mathcal{F}_{n+k-\tau}^{+} \vee \mathcal{F}_n] \right|^r |\mathcal{F}_n \right],$$

$$\Gamma_r^n(Y) = \sum_{\tau=0}^{\infty} \gamma_r^n(\tau, Y).$$

$Y_n(\theta), n \in$ is called uniformly conditionally L-mixing of order (r, s) (short: UCLM-(r, s)) for some $r, s \geq 1$ if the following hold: $Y_n(\theta)$ is L^r-bounded; $Y_n(\theta)$ is adapted to the filtration $\mathcal{F}_n$ for all θ and the sequences $M_r^n(Y)$ and $\Gamma_r^n(Y)$ are L^s-bounded.

Examples

The following specific J fulfills the conditional LLCP condition:

$$J(\theta, x) = \sum_{i=1}^{m_s} \left(\prod_{j=1}^{m_p} \mathbb{1}_{\{g_i^j(x) > \theta\}} \right) l_i(x, \theta),$$

where l_i is LLCP for all i and g_i^j has the form $g_i(x_1, \ldots, x_d) = p_1^i x_1 + \ldots + p_d^i x_d$ for all i and j. Furthermore the conditional density functions $\varphi_n(x, \omega)$ of X_{n+1} w.r.t $\mathcal{F}_n$ is bounded and have exponentially vanishing tails, i.e. $\varphi(x_1, \ldots, x_d) \leq \mathbb{c}_1 e^{\mathbb{c}_2(-|x_1| - \cdots - |x_d|)}$ for all $|x_1|, \ldots, |x_d|$.

Examples

The following specific J fulfills the conditional LLCP condition:

$$J(\theta, x) = \sum_{i=1}^{m_s} \left(\prod_{j=1}^{m_p} \mathbb{1}_{\{g_i^j(x) > \theta\}} \right) l_i(x, \theta),$$

where l_i is LLCP for all i and g_i^j has the form
$g_i(x_1, \ldots, x_d) = p_1^i x_1 + \ldots + p_d^i x_d$ for all i and j. Furthermore the conditional density functions $\varphi_n(x, \omega)$ of X_{n+1} w.r.t $\mathcal{F}_n$ is bounded and have exponentially vanishing tails, i.e. $\varphi(x_1, \ldots, x_d) \leq \mathbb{c}_1 e^{\mathbb{c}_2(-|x_1| - \cdots - |x_d|)}$ for all $|x_1|, \ldots, |x_d|$.
The causal linear process

$$X_k = \sum_{j=0}^{\infty} b_j \varepsilon_{k-j}, \text{ for } k \in \mathbb{Z},$$

where ε_j are i.i.d. for $j \in \mathbb{Z}$ and the coefficients $b_j \in \mathbb{R}$ are such that

$$b_j \leq C_b (j+1)^{-\delta},$$

for some $\delta > \frac{3}{2}$ and $C_b > 0$ satisfies the acquired mixing condition.

THANK YOU!