Conditional quantile sequantial estimation for stochastic codes

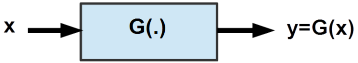
T. Labopin-Richard F. Gamboa A. Garivier

Institut de mathématiques de Toulouse

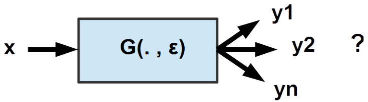
September 2, 2015

Table of contents

What is a stochastic code ?

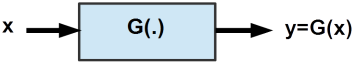


Numerical code : Y=G(X)
G(x) is a real number

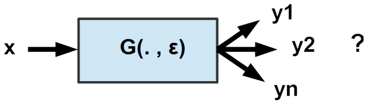


Stochastic code : Y=G(X, ε)
G(x, ε) is a random variable

What is a stochastic code ?



Numerical code : Y=G(X)
G(x) is a real number



Stochastic code : Y=G(X, ε)
G(x, ε) is a random variable

Goal : Estimate the quantile of the law $\mathcal{L}\left(G(X,\epsilon)|X=x\right)$ using as few as possible calls to the code.

- Stochastic algorithm for the quantile of level α for the law X :

$$\begin{cases} \theta_0 & \in \mathbb{R} \\ \theta_{n+1} & = \theta_n - \frac{1}{n^\gamma}\left(\mathbf{1}_{X_{n+1} \leq \theta_n} - \alpha\right) \end{cases}$$

- Stochastic algorithm for the quantile of level α for the law X :

$$\begin{cases} \theta_0(x) \in \mathbb{R} \\ \theta_{n+1}(x) = \theta_n(x) - \dfrac{1}{n^\gamma}\left(\mathbf{1}_{X_{n+1} \leq \theta_n(x)} - \alpha\right) \end{cases}$$

- Stochastic algorithm for the quantile of level α for the law X :

$$\begin{cases} \theta_0(x) \in \mathbb{R} \\ \theta_{n+1}(x) = \theta_n(x) - \frac{1}{n^\gamma}\left(\mathbf{1}_{X_{n+1} \leq \theta_n(x)} - \alpha\right) \end{cases}$$

- Problem : each call to the code is very expensive. We want a method which gives a good approximation of every conditional quantile.

- Stochastic algorithm for the quantile of level α for the law X :

$$\begin{cases} \theta_0(x) \in \mathbb{R} \\ \theta_{n+1}(x) = \theta_n(x) - \frac{1}{n^\gamma}\left(\mathbf{1}_{X_{n+1} \leq \theta_n(x)} - \alpha\right) \end{cases}$$

- Problem : each call to the code is very expensive. We want a method which gives a good approximation of every conditional quantile.
- Idea :

  1) Fix a budget N.
  2) Take an inputs sample $(X_1, \ldots, X_N)$.
  3) Observe the corresponding outputs $(Y_1, \ldots, Y_N)$.
  4) Develop an algorithm we can apply for each x and using only the previous information.

- Stochastic algorithm for the quantile of level α for the law X :

$$\begin{cases} \theta_0(x) \in \mathbb{R} \\ \theta_{n+1}(x) = \theta_n(x) - \dfrac{1}{n^\gamma}\left(\mathbf{1}_{X_{n+1} \leq \theta_n} - \alpha\right) \end{cases}$$

- Stochastic algorithm for the quantile of level α for the law X :

$$\begin{cases} \theta_0(x) \in \mathbb{R} \\ \theta_{n+1}(x) = \theta_n(x) - \frac{1}{n^\gamma}\left(\mathbf{1}_{X_{n+1} \leq \theta_n} - \alpha\right) \end{cases}$$

- New algorithm using stochastic algorithm and k-nearest neighborhoods theory :

$$\begin{cases} \theta_0(x) \in \mathbb{R} \\ \theta_{n+1}(x) = \theta_n(x) - \frac{1}{n^\gamma}\left(\mathbf{1}_{Y_{n+1} \leq \theta_n(x)} - \alpha\right)\mathbf{1}_{X_{n+1} \in kNN_n(x)} \end{cases}$$

where $k_n = \lfloor n^\beta \rfloor$.

- Stochastic algorithm for the quantile of level α for the law X :

$$\begin{cases} \theta_0(x) \in \mathbb{R} \\ \theta_{n+1}(x) = \theta_n(x) - \frac{1}{n^\gamma}\left(\mathbf{1}_{X_{n+1}\leq\theta_n} - \alpha\right) \end{cases}$$

- New algorithm using stochastic algorithm and k-nearest neighborhoods theory :

$$\begin{cases} \theta_0(x) \in \mathbb{R} \\ \theta_{n+1}(x) = \theta_n(x) - \frac{1}{n^\gamma}\left(\mathbf{1}_{Y_{n+1}\leq\theta_n(x)} - \alpha\right)\mathbf{1}_{X_{n+1}\in kNN_n(x)} \end{cases}$$

where $k_n = \lfloor n^\beta \rfloor$.

- What are the optimal parameters γ and β ? Under which assumptions is the algorithm convergent ? What about non-asymptotic results ?

Hypothesis

Let us introduce the map

$$\mathcal{C}: \begin{cases}(E, d_H) \longrightarrow (\mathcal{M}^1(\mu), d_{VT}) \\ A \longmapsto \mathcal{L}(Y|X \in A)\end{cases}$$

- (E, d_H) : set of the sets on the metric space $\left(\mathbb{R}^d, ||.||\right)$ with d_H the Hausdorff distance.
- $(\mathcal{M}^1(\mu), d_{VT})$: set of the probability measures with d_{VT} is the total variation distance.

Hypothesis

Let us introduce the map

$$\mathcal{C}: \begin{cases} (E, d_H) \longrightarrow (\mathcal{M}^1(\mu), d_{VT}) \\ A \longmapsto \mathcal{L}(Y|X \in A) \end{cases}$$

- (E, d_H) : set of the sets on the metric space $\left(\mathbb{R}^d, ||.||\right)$ with d_H the Hausdorff distance.
- $(\mathcal{M}^1(\mu), d_{VT})$: set of the probability measures with d_{VT} is the total variation distance.

Assumption A1 The function $\mathcal{C}$ is M- Lipschitz.

Hypothesis

Assumption A2 The law of inputs has a density function and this density is lower-bounded by a constant $C_{inputs} > 0$ on its support.

$$\Rightarrow \; \forall \theta_1 \in Supp(X), \; (F(\theta_1) - F(\theta^*))\,(\theta_1 - \theta^*) \geq C_{inputs}(\theta_1 - \theta^*)^2.$$

Hypothesis

Assumption A2 The law of inputs has a density function and this density is lower-bounded by a constant $C_{inputs} > 0$ on its support.

$$\Rightarrow \forall \theta_1 \in Supp(X), \ (F(\theta_1) - F(\theta^*))(\theta_1 - \theta^*) \geq C_{inputs}(\theta_1 - \theta^*)^2.$$

Assumption A3 The code function g is at values in a compact $[a, b]$.

$\Rightarrow \forall x$, $\theta_n(x)$ is bounded a.s uniformly in ω. Let us denote R the such uniform bound of $(\theta_n - \theta^*)^2$.

Hypothesis

Assumption A2 The law of inputs has a density function and this density is lower-bounded by a constant $C_{inputs} > 0$ on its support.

$$\Rightarrow \ \forall \theta_1 \in Supp(X), \ (F(\theta_1) - F(\theta^*))(\theta_1 - \theta^*) \geq C_{inputs}(\theta_1 - \theta^*)^2.$$

Assumption A3 The code function g is at values in a compact $[a, b]$.

$\Rightarrow$ $\forall x$, $\theta_n(x)$ is bounded a.s uniformly in ω. Let us denote R the such uniform bound of $(\theta_n - \theta^*)^2$.

Assumption A4 For each x, the law $g(X, \epsilon)|X = x$ has a density which is lower-bounded by a constant $D_{code}(x) > 0$.

$\Rightarrow$ avoids technical developments to deal with $\mathbb{E}(||X - x||_{(k_n,n)})$ of $\mathbb{P}(X \in kPPV_n(x))$.

Theorem : almost sure convergence

Let x be a fixed input. Under assumptions **A1** and **A2**, the algorithm at x is a.s convergent if and only if $\frac{1}{2} < \gamma < \beta < 1$.

Theorem : rate of convergence

Let x be an fixed input. Under hypothesis **A1**, **A2**, **A3** and **A4**, for all $0 < \gamma < 1$, $0 < \beta < 1$ and $1 > \epsilon > 1 - \beta$, for $n \geq 2^{\frac{1}{\epsilon-(1-\beta)}} := N_0$,

$$\mathbb{E}\left[(\theta_n(x) - \theta^*(x))^2\right] \leq R \exp\left(-\frac{3n^{1-\epsilon}}{8}\right) + a_0(x) \exp\left(-2D_{code}(x) \sum_{k=1}^{n} \frac{1}{k^{\gamma+\epsilon}}\right) + \sum_{k=1}^{n} \exp\left(-2D_{code}(x) \sum_{i=k}^{n} \frac{1}{i^{\gamma+\epsilon}}\right) \beta_k$$

where

$$\beta_n = R \exp\left(-\frac{3n^{1-\epsilon}}{8}\right) + 2\sqrt{R} M D(d) \gamma_{n+1} \left(\frac{k_n}{n+1}\right)^{\frac{1}{d}+1} + \gamma_{n+1}^2 \frac{k_n}{n+1},$$

$$D(d) = \sqrt[d]{2}\left(1 + \frac{8}{3d} + \frac{1}{\sqrt[d]{C_{input} H(d)}}\right) \text{ and } H(d) = \frac{\pi^{\frac{5}{2}}}{\Gamma\left(\frac{d}{2}\right)+1}.$$

Theorem : best parameters

Under the same hypothesis, the mean square error decreases more rapidly when parameters are $\gamma = \frac{1}{1+d}$ and $\beta = \gamma + \eta$ where $\eta > 0$ is as small as possible. We indeed obtain with these parameters for $n \geq \max(N_0, N_1)$

$$\mathbb{E}\left[(\theta_n(x) - \theta^*(x))^2\right] \leq \frac{C_1}{n^{\frac{1}{1+d} - \eta'}}$$

where constants are explicitly known.

We tested two models for $X \sim \mathcal{U}([-1,1])$, $\epsilon \sim \mathcal{U}([-0.5,0.5])$ and $x = 0$: $g(X,\epsilon) = X^2 + \epsilon$ and $g(X,\epsilon) = |X| + \epsilon$:

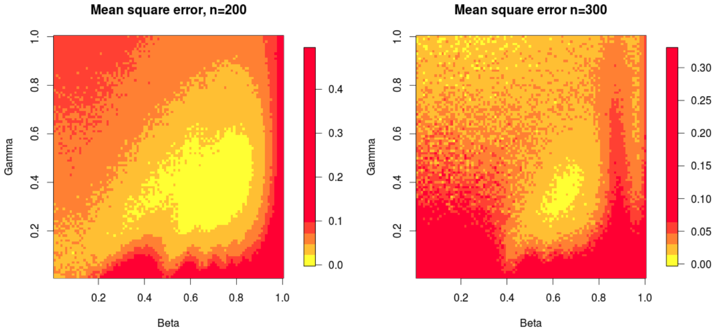


FIGURE: Convergence of the mean square error in function of β and γ.

Dimension 2 and 3

We tested the model $g(X, \epsilon) = ||X||^2 + \epsilon$ for $X \sim \mathcal{U}([-1, 1]^d)$, $\epsilon \sim \mathcal{U}([-0.5, 0.5])$ and $x = 0_{\mathbb{R}^d}$.

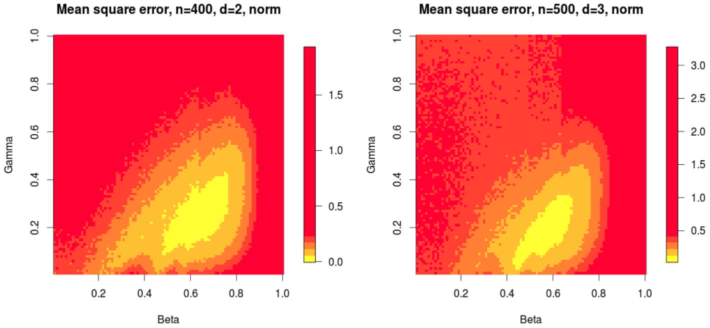


FIGURE: Convergence of the mean square error in function of β and γ.

Conclusion and perspectives

Conclusion :

- We have introduced an algorithm to estimate the conditional quantile of the outpout of a stochastic code.
- We know how to tune the parameters to obtain the best rate of convergence.
- Numerical simulations show that results are good in practice.

Perspectives :

- Adapt the proof for less restrictive assumptions.
- Improve the rank N_0.
- Apply this algorithm on real data (EDF and renewable energies).

Thank you for your attention