Confusing naming in the stochastic methods

Andersen Ang

Mathématique et recherche opérationnelle
UMONS, Belgium

manshun.ang@umons.ac.be Homepage: angms.science

First draft : June 21, 2019
Last update : June 21, 2019

The stochastic cases

Consider the minimization problem

$$\min_{\mathbf{x}\in\mathcal{Q}} F(\mathbf{x}).$$

It is a stochastic programming problem if

$$F(\mathbf{x}) = \mathbb{E}_{\xi}\big[f(\mathbf{x},\xi)\big].$$

where

- $\mathbf{x}$ is the optimization variable
- ξ is the random variable
- the expectation is taken over the random variable ξ as an integral

$$F(\mathbf{x}) = \int_{\xi} p(\xi) f(\mathbf{x},\xi) d\xi$$

- The key point is that F involves randomness inside the function

The non-stochastic cases : average of finite sum

Consider the minimization problem

$$\min_{\mathbf{x} \in \mathcal{Q}} F(\mathbf{x}).$$

It is a randomized programming problem if

$$F(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} f_i(\mathbf{x}).$$

where

- F is the average of a finite sum ($n \neq \infty$)
- $\mathbf{x}$ is the optimization variable
- n is the number of sum
- f_i is a component of F
- There is no random variable, it is a deterministic function

Relationship between stochastic and average of finite sum

The function $F_1(\mathbf{x}) = \mathbb{E}_\xi\big[f(\mathbf{x},\xi)\big]$ becomes $F_2(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^n f_i(\mathbf{x})$ if

$$
\begin{aligned}
F_1(\mathbf{x}) = \mathbb{E}_\xi\big[f(\mathbf{x},\xi)\big] &= \int_\xi p(\xi) f(\mathbf{x},\xi) d\xi \\
\text{(discretization of } \mathbb{E}\text{)} &= \sum_{i=1}^n p_i f(\mathbf{x},\xi_i) \\
\text{(all } p_i \text{ uniformly distributed)} &= \sum_{i=1}^n \frac{1}{n} f(\mathbf{x},\xi_i) \\
\Big(f(\mathbf{x},\xi) = f_i(\mathbf{x})\Big) &= \frac{1}{n}\sum_{i=1}^n f_i(\mathbf{x}) = F_2(\mathbf{x})
\end{aligned}
$$

Note that F_2 is **deterministic**. There is NO randomness in F_2. The randomness in F_1 is expressed by the random variable ξ, while in F_2 there is nothing related to randomness (the i here are deterministic). Hence even if $n \to \infty$ in F_2, F_2 does not "go back" to F_1.

The so called "Stochastic gradient" is not stochastic

Consider solving the problem of average of finite sum

$$\min_{\mathbf{x} \in \mathcal{Q}} F(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} f_i(\mathbf{x})$$

using the so-called "stochastic gradient" :

- Step 1. Take $i_k \in \{1, 2, \ldots, n\}$ by random under a distribution
- Step 2. Perform Update on using the partial gradient $\nabla_{i_k} F$

This is NOT stochastic programming :

- There is no random variable (e.g. ξ) in F
- The randomness is on i_k which is **outside** the function F

Hence it is wrong to call such method "stochastic".
It should be called "randomized partial gradient".

Wrong naming in the "stochastic algorithms"

For the following **deterministic** problem

$$\min_{\mathbf{x}\in\mathcal{Q}} F(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} f_i(\mathbf{x}),$$

algorithms that call themselves "stochastic" are misusing the term "stochastic". For examples :

- Stochastic gradients (`SG`)
- Stochastic averaged gradients (`SAG`)
- Stochastic variance reduced gradient (`SVRG`)

These method should be called "randomized method".

However, because these misused names already got popular, people keep using these "misused names" now as a convention. Thus it is important to make sure the difference between these "stochastic methods" and those real "stochastic methods".

-END-