How to estimate a bound for the probability of having reached a global optimum

6-601-09 Simulation Monte Carlo

Geneviève Gauthier

HEC Montréal

Reference

Based on Robbins, H. E. (1968). Estimating the Total Probability of the Unobserved Outcomes of an Experiment, *The Annals of Mathematical Statistics*, **39**, 256-257.

Context

- Let $E_1, E_2, \ldots$ be the possible outcomes of a random variable with unknown probabilities $p_1, p_2, \ldots$respectively. (We may think of $E_1, E_2, \ldots$ as the local optimums)
- Let $X_i^{(n)}$ be the number of times E_i occurs in n independent trials.
  - Note that $X_i^{(n)}$ in binomial(n, p_i).
- We are interested in estimating the probability

$$U^{(n)} = \sum_{i=1}^{\infty} p_i \mathbf{1}_{X_i^{(n)}=0}$$

  of the unobserved outcomes after n independent trials.
  - It cannot be computed directly since the p_i's are unknown.
  - Moreover, this is not a standard parameter estimation since $U^{(n)}$ is stochastic.

Application

1. We want to optimize numerically the fonction $g(\boldsymbol{\theta})$ that has many local optimum labelled $E_1, E_2, \ldots$
   1.1 The global optimum is one of them.
   1.2 $\boldsymbol{\theta}$ could be a vector of parameters.
2. We initialize a standard optimization algorithm using an initial value $\boldsymbol{\theta}^*$.
   2.1 After the optimization stage, one of the local optimum will be reached.
3. To improve the chances to reach the global optimum, we repeat the previous step n times, choosing the n initial points independly and uniformly distributed over a $d-$dimensionnal hyperbox of reasonnable values for $\boldsymbol{\theta}$.
4. Note that the probability of not having reach the global optimum is $\leq U^{(n)}$.

Ponctual estimation I

1. Let
$$nV^{(n)} = \sum_{i=1}^{\infty} \mathbf{1}_{X_i^{(n)}=1}$$
be the number of singletons observed in the n independent trials.
2. Define $W^{(n+1)} = U^{(n)} - V^{(n+1)}$.
3. **Result**. $\mathrm{E}\left[W^{(n+1)}\right] = 0$.
   3.1 This could be interpreted as $V^{(n+1)}$ is an unbias estimator of $U^{(n)}$.

Ponctual estimation II

4. **Proof.**

$$\begin{aligned}
\mathrm{E}\left[W^{(n+1)}\right] &= \mathrm{E}\left[\sum_{i=1}^{\infty} p_i \mathbf{1}_{X_i^{(n)}=0} - \frac{1}{n+1}\sum_{i=1}^{\infty}\mathbf{1}_{X_i^{(n+1)}=1}\right] \\
&= \sum_{i=1}^{\infty}\left(p_i \mathrm{E}\left[\mathbf{1}_{X_i^{(n)}=0}\right] - \frac{1}{n+1}\mathrm{E}\left(\mathbf{1}_{X_i^{(n+1)}=1}\right)\right) \\
&= \sum_{i=1}^{\infty}\left(p_i\left(1-p_i\right)^n - \frac{1}{n+1}\left(n+1\right)p_i\left(1-p_i\right)^n\right) \\
&= 0.
\end{aligned}$$

5. Since $0 = \mathrm{E}\left[W^{(n+1)}\right] = \mathrm{E}\left[U^{(n)} - V^{(n+1)}\right]$, the objective is to show that the proportion

$$V^{(n+1)} = n^{-1}\sum_{i=1}^{\infty}\mathbf{1}_{X_i^{(n)}=1}$$

of singletons is not a too bad approximation for $U^{(n)}$.

Precision

Using properties of the binomial distribution, one can show that

$$\mathrm{E}\left[\left(W^{(n+1)}\right)^2\right] = \frac{1}{n+1}\sum_{i=1}^{\infty} p_i\left(1-p_i\right)^n\left(1+\left(n-1\right)p_i\right) - \frac{1}{n+1}\sum_{i=1}^{\infty}\sum_{j\neq i} p_i p_j\left(1-p_i-p_j\right)^n .$$

Precision

Since $1 - x \leq \exp(-x)$ for all $x \geq 0$,

$$\begin{aligned} & \sum_{i=1}^{\infty} p_i (1 - p_i)^n (1 + (n-1) p_i) \\ \leq \; & \sum_{i=1}^{\infty} p_i e^{-np_i} e^{-(1-n)p_i} \\ = \; & \sum_{i=1}^{\infty} p_i e^{-p_i} \\ \leq \; & \sum_{i=1}^{\infty} p_i = 1. \end{aligned}$$

Precision

Therefore

$$
\begin{aligned}
& \mathrm{E}\left[\left(W^{(n+1)}\right)^2\right] \\
= \quad & \frac{1}{n+1}\sum_{i=1}^{\infty} p_i\left(1-p_i\right)^n\left(1+\left(n-1\right)p_i\right) \\
& \underbrace{-\frac{1}{n+1}\sum_{i=1}^{\infty}\sum_{j\neq i} p_i p_j\left(1-p_i-p_j\right)^n}_{\geq 0} \\
\leq \quad & \frac{1}{n+1}.
\end{aligned}
$$

A loose bound for the precision

Tchebychev's inequality implies that for any positive ε,

$$P\left[\left|U^{(n)} - V^{(n+1)}\right| > \varepsilon\right] \leq \varepsilon^{-2}\mathrm{E}\left[\left(W^{(n+1)}\right)^2\right].$$

We conclude that

$$\begin{aligned} & P\left[V^{(n+1)} - \varepsilon \leq U^{(n)} \leq V^{(n+1)} + \varepsilon\right] \\ \geq & \quad 1 - \frac{1}{\varepsilon^2}\mathrm{E}\left[\left(W^{(n+1)}\right)^2\right] \\ \geq & \quad 1 - \frac{1}{\varepsilon^2\left(n+1\right)} \\ \rightarrow & \quad 1 \text{ as } n \rightarrow \infty \end{aligned}$$

Second moment I

Proof

Result.

$$\mathrm{E}\left[\left(W^{(n+1)}\right)^2\right] = \frac{1}{n+1}\sum_{i=1}^{\infty} p_i\left(1-p_i\right)^n\left(1+\left(n-1\right)p_i\right) - \frac{1}{n+1}\sum_{i=1}^{\infty}\sum_{j\neq i} p_i p_j\left(1-p_i-p_j\right)^n .$$

Second moment II

Proof

Proof.

$$\mathrm{E}\left[\left(W^{(n+1)}\right)^2\right] = \mathrm{E}\left[\left(\sum_{i=1}^{\infty}\left(p_i \mathbf{1}_{X_i^{(n)}=0} - \frac{1}{n+1}\mathbf{1}_{X_i^{(n+1)}=1}\right)\right)^2\right]$$

$$= \sum_{i=1}^{\infty}\left(p_i^2 \mathrm{E}\left[\mathbf{1}_{X_i^{(n)}=0}\right] - \frac{2p_i}{n+1}\mathrm{E}\left[\mathbf{1}_{X_i^{(n)}=0}\mathbf{1}_{X_i^{(n+1)}=1}\right] + \frac{1}{(n+1)^2}\mathrm{E}\left[\mathbf{1}_{X_i^{(n+1)}=1}\right]\right) +$$

$$\sum_{i=1}^{\infty}\sum_{j\neq i}\left(\begin{array}{c} p_i p_j \mathrm{E}\left[\mathbf{1}_{X_i^{(n)}=0}\mathbf{1}_{X_j^{(n)}=0}\right] - \frac{p_i}{n+1}\mathrm{E}\left[\mathbf{1}_{X_i^{(n)}=0}\mathbf{1}_{X_j^{(n+1)}=1}\right] \\ -\frac{p_j}{n+1}\mathrm{E}\left[\mathbf{1}_{X_i^{(n+1)}=1}\mathbf{1}_{X_j^{(n)}=0}\right] + \frac{1}{(n+1)^2}\mathrm{E}\left[\mathbf{1}_{X_i^{(n+1)}=1}\mathbf{1}_{X_j^{(n+1)}=1}\right] \end{array}\right)$$

$$= \sum_{i=1}^{\infty}\left(p_i^2(1-p_i)^n - \frac{2p_i}{n+1}(1-p_i)^n p_i + \frac{1}{(n+1)^2}(n+1)p_i(1-p_i)^n\right) +$$

$$\sum_{i=1}^{\infty}\sum_{j\neq i}\left(\begin{array}{c} p_i p_j (1-p_i-p_j)^n \\ -\frac{p_i}{n+1}\left[(1-p_i-p_j)^n p_j + np_j(1-p_i-p_j)^{n-1}(1-p_j)\right] \\ -\frac{p_j}{n+1}\left[(1-p_i-p_j)^n p_i + np_i(1-p_i-p_j)^{n-1}(1-p_i)\right] \\ +\frac{1}{(n+1)^2}(n+1)p_i np_j(1-p_i-p_j)^{n-1} \end{array}\right)$$

$$= \frac{1}{n+1}\sum_{i=1}^{\infty} p_i(1-p_i)^n(1+(n-1)p_i) - \frac{1}{n+1}\sum_{i=1}^{\infty}\sum_{j\neq i} p_i p_j (1-p_i-p_j)^n$$