A Bayesian Approach for Stochastic Continuum-armed Bandit with Long-term Constraints

Zai Shi [1] Atilla Eryilmaz [1]

[1]ECE, The Ohio State University

Problem Formulation

We consider a problem called stochastic continuum-armed bandit with long-term constraints.

Setup: In each iteration, the decision-maker chooses an action x_t from a compact set $\mathcal{X}$, and then observes a random reward value $f^t(x_t)$ and m constraint values $\{g_j^t(x_t)\}_{j=1}^m$. We assume that $f^t(x_t) = f(x_t) + \varepsilon^t$, where ε^t is a zero-mean random variable independent across t.

Two cases: We consider two cases of the above setup in this paper.

- $g_j^t(x_t) = g_j(x_t), \forall j$ is deterministic;
- $g_j^t(x_t) = g_j(x_t) + \varepsilon_j^t, \forall j$, where ε_j^t is a zero-mean random variable independent across t.

Blackbox: f and g_j unknown. Exact distributions of random variables unknown

Target: To maximize our expected reward $f(x)$ with long-term constraints satisfied in expectation, i.e., $\frac{1}{T}\sum_{t=1}^T g_j(x_t) \leq 0$.

Metric: *Regret*:

$$R_T = \sum_{t=1}^{T}[f(x^*) - f(x_t)] = o(T)$$

Constraint Violation (CV):

$$V_T = ||[\sum_{t=1}^{T}\mathbf{g}(x_t)]^+|| = o(T)$$

where $\mathbf{g}(x)$ is $[g_1(x), .., g_m(x)]$ and $[\mathbf{a}]^+$ means element-wise $\max(0, a)$ in vector $\mathbf{a}$. Here x^* is defined as any global optimum of

$$\max_{x\in\mathcal{X}} f(x)$$
$$s.t.\, g_j(x) \leq 0, j = 1, ..., m.$$

We hope that an algorithm can make both metrics sublinear.

Applications

Since our setup does not require f and g_j to be known or convex, it has wide applications.

- Hyperparameter Tuning
- Data Rate Allocation.

Design Methodology

We use two techniques to design our algorithms.

Bayesian Optimization

Classical BO methods are used for unconstrained optimization with a blackbox objective function f.

- Put a Gaussian process (GP) prior on f and get its posterior distribution after inquiries of f.
- Choose next inquiries based on some **strategy** and update the posterior distribution of f.
- Repeat the above step until the optimal point can be inferred from the posterior distribution.

One strategy used in our paper: IGP-UCB

Algorithm 1 IGP-UCB$(f(x), k, B, R, \lambda, \delta, S)$

Input: Prior $GP(0, k)$, parameters B, R, λ, δ, S.
for $s = 1, ..., S$ **do**
  Set $\beta_s = B + R\sqrt{2(\gamma_{s-1} + 1 + \log(1/\delta))}$.
  Choose $x_s = \arg\max_{x\in\mathcal{X}}\{\mu_{s-1}(x) + \beta_s\sigma_{s-1}(x)\}$.
  Obtain noisy observation of $f(x_s)$.
  Perform update to get μ_s and σ_s.
end for
Output: $x_1, ..., x_S$.

Penalty Approach

It is the strategy of appending a generic penalty function to the objective as follows:

$$f(x) - \sum_{j=1}^{m}\kappa_j\Lambda(g_j(x)), \quad (1)$$

where $\Lambda(\cdot)$ is some penalty function and κ_j is a multiplier for constraint j.

Proposition 1

For any fixed choices of κ_j and $\Lambda(\cdot)$ in (1), and any unconstrained bandit optimization algorithm $\mathcal{M}$ that can produce a sublinear regret for its objective function, applying $\mathcal{M}$ to (1) will fail to yield a sublinear regret and a sublinear CV for all forms of $f(x)$ and $\mathbf{g}(x)$ in our setup.

We need an update rule for multipliers or penalty functions!

Noiseless Constraint Observations

First we assume that the constraint functions are observed without noise. Our algorithm is based on a multiplicative form of performing multiplier-updates.

Algorithm 2 GP-UCB with Noiseless Constraints

Initialize c and $\kappa_j^1 = 1$ for all j.
for $l = 1, ..., L$ **do**
  Run IGP-UCB$(f(x) - \sum_{j=1}^m \kappa_j^l(\psi(g_j(x)) - 1), k_l, B, R, \lambda, \delta/L, S)$ for S iterations to produce $\{x_l^1, ..., x_l^S\}$, while obtaining S observations $\{f(x_l^s) + \varepsilon_l^s\}_{s=1}^S$ and $\{g_j(x_l^s)\}_{s=1}^S, \forall j$ sequentially with the above outputs, where ε_l^s is the observation noise of f.
  Set $\kappa_j^{l+1} = \kappa_j^l\psi\left(\frac{1}{S}\sum_{s=1}^S g_j(x_l^s)\right), \forall j$
end for
Output: $\{\{x_l^s\}_{l=1}^L\}_{s=1}^S$.

Theorem 1. (Performance of Alg. 2)

For a class of penalty functions $\psi(\cdot)$, with certain assumptions and appropriate parameters, Alg. 2 can achieve

$$R_{LS} = O(BL\sqrt{S\tilde{\gamma}_S} + L\sqrt{S\tilde{\gamma}_S(\tilde{\gamma}_S + \log(L/\delta))})$$
$$V_{LS} = O(S\psi_+^{-1}(L + (BL\sqrt{\frac{\tilde{\gamma}_S}{S}}$$
$$+ L\sqrt{\frac{\tilde{\gamma}_S^2 + \tilde{\gamma}_S\log(L/\delta)}{S}})))$$

with probability at least $1 - \delta$.

Remark

- How to choose $\psi(\cdot)$ needs to be taken care of.
- There exists a tradeoff between regret and CV based on selection of L and S. For most common kernels we can make them both sublinear.

Noisy Constraint Observations

All constraint functions are observed with noise.

Why not use Alg. 2: The multiplicative form of multiplier-updates will amplify the noise in the constraint observations.

Therefore, our second algorithm is based on additive multiplier-updates.

Algorithm 3 GP-UCB with Noisy Constraints

Initialize $\kappa_j^1 = 0$ for all j.
for $l = 1, ..., L$ **do**
  Make the following S decisions sequentially:

$$\{x_l^1, ..., x_l^S\} = \text{IGP-UCB}(f(x) - \sum_{j=1}^m \kappa_j^l g_j(x), k_l,$$
$$B, \sqrt{(1 + \sum_{j=1}^m(\kappa_j^l)^2)R}, \lambda, \delta/(2L), S),$$

  while obtaining S observations $\{f(x_l^s) + \varepsilon_l^s\}_{s=1}^S$ and $\{g_j(x_l^s) + \varepsilon_{l,j}^s\}_{s=1}^S, \forall j$ sequentially with the above outputs, where ε_l^s is the observation noise of f and $\varepsilon_{l,j}^s$ is the observation noise of g_j.
  Set $\kappa_j^{l+1} = [\kappa_j^l + \mu\sum_{s=1}^S(g_j(x_l^s) + \varepsilon_{l,j}^s)/S]^+, \forall j$
end for
Output: $\{\{x_l^s\}_{l=1}^L\}_{s=1}^S$.

Theorem 2. (Performance of Alg. 3)

With certain assumptions and appropriate parameters, Alg. 3 can achieve

$$R_{LS} = O\left(L(B + \sqrt{\tilde{\gamma}_S + \log\frac{2L}{\delta}})\sqrt{S\tilde{\gamma}_S} + S\sqrt{L}\right)$$
$$V_{LS} = O(L\sqrt{S\log(L/\delta)} + S\sqrt{L})$$

with a probability at least $1 - \delta$.

Remark: Worse bounds than Alg. 2 due to the noise in the constraint observations. Both metrics can be sublinear for most common kernels.

Numerical Results

- Synthetic functions
- Data rate allocation problems

Acknowledgements

We thank the NSF grants: IIS-2112471, CNS-NeTS-2106679, CNS-NeTS-2007231, CNS-SpecEES-1824337, CNS-NeTS-1717045; and the ONR Grant N00014-19-1-2621 for their support of this work. We also thank the suggestions of all the reviewers and the meta-reviewer for the improvement of this paper.