Review: Multi-armed Bandit Models for the Optimal Design of Clinical Trials: Benefits and Challenges

Ken Tanaka Hernández

International Centre for Theoretical Physics

Project for Reinforcement Learning:
Prof. Antonio Celani

July 9, 2023

Motivation

Goals

- Identify the "Best" treatment : *Exploration or Learning*
- Treat patients as "Effectively" as possible during the trial: *Exploitation or Earnings*

Introduction

Bayesian Bernoulli K-Armed Bandit Problem

Let $y_{k,t} \in \mathbf{Y}_{k,t} \sim \text{Ber}(\rho_k)$, where:

Treatment $\equiv k \in \{1, \ldots, K\} \leftarrow$ Arm

Patient $\equiv t \in \{1, \ldots, N\} \leftarrow$ Time

The *Bayesian* feature: $\rho_k \in \mathbf{P}_k \sim \text{Beta}(\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} x^{\beta-1}$;

Conjugate Prior distribution

$$(\alpha, \beta) = \begin{cases} (S_{k,0}, F_{k,0}) \in \mathbb{N}_+^2 & \\ (S_{k,0} + S_{k,t}, F_{k,0} + F_{k,t}) & : t \geq 1 \end{cases}$$

(Succesful, Failure) $\equiv (S_{k,t}, F_{k,t}) \in \mathbb{N}_0^2$

Action space:

$$\mathbb{A}_k \ni a_{k,t} = \{0, 1\}$$

Markovian *transition probability rule*:

$$P_k\{\mathbf{s}_{k,t+1} | \mathbf{s}_{k,t}, a_{k,t}\} \sim$$

$$\mathbf{s}_{k,t+1} = \begin{cases} \begin{cases} (S_{k,0} + S_{k,t} + 1, F_{k,0} + F_{k,t}) : (S_{k,0} + S_{k,t})/c_t \\ (S_{k,0} + S_{k,t}, F_{k,0} + F_{k,t} + 1) : (F_{k,0} + F_{k,t})/c_t \end{cases} & : a_{k,t} = 1 \\ \mathbf{s}_{k,t} & : a_{k,t} = 0 \end{cases}$$

$$c_t = S_{k,0} + S_{k,t} + F_{k,0} + F_{k,t}$$

$$R(\mathbf{s}_{k,t}, a_{k,t}) = \frac{S_{k,0} + S_{k,t}}{c_t} a_{k,t}$$

Introduction

Objective

$$V_\pi^*(\mathbf{s}) = \max_\pi \mathbb{E}_\pi \left[\sum_{t=1}^{N} \gamma^t \sum_{k=1}^{K} R(\mathbf{s}_{k,t}, a_{k,t}) | \mathbf{s}_0 = \mathbf{s} \right]$$

Bayesian regret

$$R_{-1} = N \max_k (\rho_k) - \mathbb{E}_\pi \left[\sum_{k=0}^{K} \sum_{t=1}^{N} a_{k,t} \right]$$