In a Nutshell

What is the performance of Bayesian bandit algorithms from a frequentist point of view? Bayes-UCB and Thompson Sampling appear to outperform frequentist algorithms on their own ground, which is supported by optimal regret bound for the Bernoulli case.

Bayesian vs. Frequentist Model for MAB

K independent arms. Arm a depends on parameter \(\theta_a\) and has expectation \(\mu_a\); optimal arm is \(a^\ast = \arg\max_a \mu_a\) and \(\mu^\ast = \mu_{a^\ast}\) is the highest expectation of reward associated.

Two probabilistic modelings

- **Frequentist**: \(\theta_1, \ldots, \theta_K\) unknown parameters
- **Bayesian**: \(\theta_1, \ldots, \theta_K \sim \pi\)

At time \(t\), arm \(A_t\) is chosen and reward \(X_t = Y_{A_t,t}\) is observed

Two measures of performance

- Minimize (classic) regret
- Minimize “Bayesian” regret

\[ R_n(\theta) = \mathbb{E}_\theta \left[ \sum_{t=1}^n \theta^* - \theta_{A_t} \right] \]

\[ R_n = \int R_n(\theta)d\pi(\theta) \]

Optimal algorithms

- **Asymptotically optimal algorithms** satisfy, for \(a: \mu_a < \mu^\ast\),
  \[ \lim_{n \to \infty} \frac{\mathbb{E}_\theta[N_a(n)]}{\log(n)} \leq \frac{1}{\text{KL}(\nu_{a^\ast}, \nu_a)} \]
  They are optimal in the sense of Lai and Robbins' lower bound (1985) on the number of draws of a sub-optimal arm

\[ \Rightarrow \text{Our goal: Design algorithms inspired by the Bayesian modeling that are asymptotically optimal in the frequentist setting.} \]

Bayesian vs. Frequentist Algorithms

Some quantities that naturally arise in the Bayesian modeling are

**Prior**: \(\Pi_1 = \{\pi_1, \ldots, \pi_K\}\) the current posterior over \(\theta_1, \ldots, \theta_K\)

**Posterior**: \(\Lambda_1 = \{\lambda_1, \ldots, \lambda_K\}\) the current posterior over the means \(\mu_1, \ldots, \mu_K\)

Successful algorithms inspired by the frequentist modeling use

- Upper Confidence Bound for the empirical mean... (UCB)
- ... built using KL-divergence (KL-UCB, asymptotically optimal)

Whereas a **Bayesian algorithm** uses \(\Pi_1\) to determine action \(A_t\).

Bayes-UCB and Thompson Sampling

**Bayes-UCB algorithm** chooses \(A_t = \arg\max_a q_a(t)\), with \(a = 1, \ldots, K\)

\[ q_a(t) = Q \left( 1 - \frac{1}{(\log(t))^c} ; \lambda_a^t \right) \]

**Thompson Sampling** is a randomized algorithm:

\[ \forall a \in \{1, K\}, \quad \theta_a(t) \sim \lambda_a^t \]

\[ A_t = \arg\max_a \theta_a(t) \]

Parameters: \(c\) (in practice, take \(c = 0\)), initial prior \(\Pi_0\)

Bayes-UCB: Theoretical elements

\(\nu_a\) is the Bernoulli distribution \(\mathcal{B}(\theta_a)\), \(\pi_a^0\) the (conjugate) prior \(\mathcal{U}(0, 1)\)

**Bayes-UCB is asymptotically optimal for Bernoulli bands**

**Theorem 1** Let \(c > 0\); for the Bayes-UCB algorithm with parameter \(c \geq 5\), the number of draws of a sub-optimal arm \(a\) is such that:

\[ \mathbb{E}_\theta[N_a(n)] \leq \frac{1}{\text{KL}(\nu_{a^\ast}, \nu_a)} \log(n) + o_n \left( \log(n) \right) \]

**Bayes-UCB is very close to a frequentist algorithm**

The Bayes-UCB index \(q_a(t)\) is closely related to the one used by the KL-UCB algorithm (Cappé et al. 2013):

\[ \tilde{u}_a(t) = \arg\max_{t > t_0} \left\{ d \left( \frac{S_a(t)}{N_a(t)} \right) \leq \frac{\log(t) + c\log(\log(t))}{N_a(t)} \right\} \]

where \(d(x) = \text{KL}(\mathcal{B}(x) ; \mathcal{B}(y)) = x \log \frac{x}{y} + (1 - x) \log \frac{1 - x}{1 - y}\)

Bayes-UCB appears to build automatically confidence intervals based on Kullback-Leibler divergence, that are adapted to the geometry of the problem in this specific case.

Thompson Sampling: Theoretical elements

**TS is asymptotically optimal for Bernoulli bands**

**Theorem 2** Let \(c > 0\). With \(b\) defined below, for every sub-optimal arm \(a\), there exists a constant \(N(b, \epsilon, \theta, \theta^\ast)\) such that

\[ \mathbb{E}[N_a(n)] \leq (1 + \epsilon) \log(n) + o_n \left( \frac{\ln(n) + \ln(n)}{d(\theta_a, \theta^\ast)} \right) + N(b, \epsilon, \theta, \theta^\ast) + 5 + 2b \]

**Proof Bottleneck:** For some constants \(b = b(\mu) \in (0, 1)\) and \(C_\theta < \infty\),

\[ \sum_{t=1}^{n} \mathbb{P}(N_t(t) \leq t^b) \leq C_\theta \]

Numerical experiments and beyond

- **Bayesian algorithms are practically as efficient as optimal frequentist algorithms or even better!**

Regret of the various algorithms as a function of time. The red curve shows the lower bound, the black bold curve the mean regret and the dark and light shaded areas the central 99% and the upper 0.05%

- **They are easier to implement:** KL-UCB solves an optimization problem whereas Thompson Sampling only produces one sample!
- **They are easy to generalize:** general models where sampling from a posterior distribution is possible (using MCMC), sparse linear bandit, contextual bandit model...

References