Bayesian vs. Frequentist Model for MAB

K independent arms depending on a parameter \( \theta \) (Bernoulli distribution for the sake of simplicity); optimal arm is \( j^* = \arg \max \theta_j \) and \( \theta^* = \theta_{j^*} \), is the highest expectation of reward associated

Two probabilistic modellings

- **Frequentist**: 
  - \( \theta_1, \ldots, \theta_K \) unknown parameters
  - \( (Y_{t,j})_i \) i.i.d. with Bernoulli distribution \( B(\theta_j) \)
  - \( (Y_{t,j})_i \) are i.i.d. conditionally to \( \theta_j \) with distribution \( B(\theta_j) \)

At time \( t \) + 1, arm \( I_t \) is chosen and reward \( X_{t+1} = Y_{t+1,j_t} \) is observed

- **Bayesian**: 
  - Minimize (classic) regret \( R_n(\theta) = \mathbb{E}_{\theta}[X^t_{t+1}] \) 
  - Minimize “bayesian” regret \( R_n = R_n(\theta)d\mathbb{Q}(\theta) \)

Bayesian : Two measures of performance

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

Bayes-UCB : a simple bayesian strategy

Some ideas for using the posterior : sampling from the posterior (Thompson Sampling) using quantiles : fixed or adaptive (Bayes-UCB) adapt the bayesian exact solution from Gittins (FHG-algorithm)

Bayes-UCB algorithm is the index policy associated to

\[
q_t(j) = \frac{1}{t} \left(1 - \frac{1}{\log(n)^a}\right) - \text{quantile of the posterior distribution}
\]

\[
\text{Beta}(\theta,\beta) \text{ the prior over each arm}
\]

\[
\mathbb{S}_t = \frac{S_{t,j}}{ \mathbb{S}_t,j} + 1, \mathbb{S}_t,j = \text{number of ones observed from arm } j \text{ until time } t
\]

\[
\mathbb{E}[X_t] = \mathbb{E}[X_t|S_{t,j} = I_t, L_t = j] = \frac{S_{t,j} + a}{S_{t,j} + 1 + S_{t,j} + a + b}
\]

Finite-Horizon-Gittins algorithm is the index policy associated to \( \psi(t,S_{t,j}) \)

The computation of the index is more involved

- Theoretical guarantee: Bayesian optimal

On the efficiency of Bayesian bandit algorithms from a frequentist point of view

Emilie Kaufmann, Olivier Cappé and Aurélien Garivier (name@telecom-paristech.fr)