Improved bandit algorithms: Go Bayesian!

Emilie Kaufmann, Telecom ParisTech & CNRS

Bayesian vs. Frequentist Model for MAB

K independent arms. Arm a depends on parameter θa and has expectation μa: optimal arm is a∗ = argmax μa and μ∗ = μa∗, is the highest expectation of reward associated.

Two probabilistic modelings

Frequentist:

• θ1,...,θK unknown parameters
• (Y,a,t) is i.i.d. with distribution νa at time t, arm A is chosen and reward X = YA,t is observed

Bayesian:

• θa ∼ πa
• (Y,a,t) is i.i.d. conditionally to θa with distribution νaθ

At time t, arm A is chosen and reward X = YA,t is observed

Two measures of performance

• Minimize (classic) regret #
\[ R_a(θ) = E_θ (θ^* - θ_A) \]
• Minimize “Bayesian” regret
\[ R_a = R_a(θ) dπ(θ) \]

Optimal algorithms

• Asymptotically optimal algorithms satisfy, for a : μa < μ∗,
\[ \lim_{n→∞} E_θ[N_a(n)] ≤ \frac{1}{KL(ν_0, ν_0')} \]

They are optimal in the sense of Lai and Robbins’ lower bound (1985) on the number of draw of a sub-optimal arm

⇒ Our goal: Design algorithms inspired by the Bayesian modeling that are asymptotically optimal in the frequentist setting.

Bayesian versus frequentist algorithms

Some quantities that naturally arise in the Bayesian modeling are

• \[ Π_t = (π_1^t,...,π_K^t) \] the current posterior over (θ1,...,θK)
• \[ Λ_t = (λ_1^t,...,λ_K^t) \] the current posterior over the means (μ1^t,...,μK^t)

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 Πt to determine action Λt.

Bayes-UCB and Thompson Sampling

Bayes-UCB algorithm chooses \[ A_t = \text{argmax}_{a=1...K} q_a(t) \], with
\[ q_a(t) = Q - \frac{1}{(t log(n))^\epsilon} λ_a^t \]

Thompson Sampling is a randomized algorithm:
\[ \forall a ∈ \{1...K\}, \quad θ_a(t) ∼ λ_a^t \]

\[ A_t = \text{argmax}_{a=1...K} θ_a(t) \]

Parameters: c (in practice, take c = 0), initial prior Π0

Thompson Sampling: Theoretical elements

• TS is asymptotically optimal for Bernoulli bandits

Theorem 2 Let ε > 0. With b defined below, for every sub-optimal arm a, there exists a constant N(b, ε, θa∗) such that
\[ E[N_a(n)] ≤ \frac{1 + ε}{d(θ_a, θ_a^*)} n + N(b, ε, θ_a∗) + 5 + 2C_b \]

Proof Bottleneck: For some constants b = b(μ) ∈ (0,1) and C_b < ∞, \[ \sum_{t=1}^{n} N_i(t) ≤ b^t ≤ C_b \]

Bayes-UCB: Theoretical elements

\[ q_a(0) = 0 \]
\[ q_a(n) = \frac{\log(n) + \log(\log(n))}{KL(ν_0, ν_0')} \]

Bayes-UCB is asymptotically optimal for Bernoulli bandits

Theorem 1 Let ε > 0; for the Bayes-UCB algorithm with parameter b ≥ 5, the number of draw of a sub-optimal arm a is such that:
\[ E_θ[N_a(n)] ≤ \frac{1}{KL(ν_0, ν_0')} \log(n) + o_ε(\log(n)) \]

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): \[ q_a(t) = \frac{1}{α(t)} \log(\log(n)) \]

Numerical experiments and beyond

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