Successive Rejection Algorithm

- Assume that the rewards are bounded in [0,1]
- The algorithm is as follows

Sample each arm once,

If at sample t,

$\overline{X}_{\max} - \overline{X}_j \geq 2\alpha_t$ then remove arm j from consideration where $\alpha_t = \sqrt{\frac{4\log(Kt/\delta)}{t}}$

Repeat till one arm is left, and announce it as the best arm.

Successive Rejection Algorithm

- Intuitive idea: Best arm is never rejected as
- $\overline{X}_t(t) \geq \mu_1 - \alpha_t$
- $\overline{X}_a(t) \leq \mu_a + \alpha_t$
- So $\overline{X}_a(t) - \overline{X}_t(t) \leq 2\alpha_t - (\mu_1 - \mu_a)$