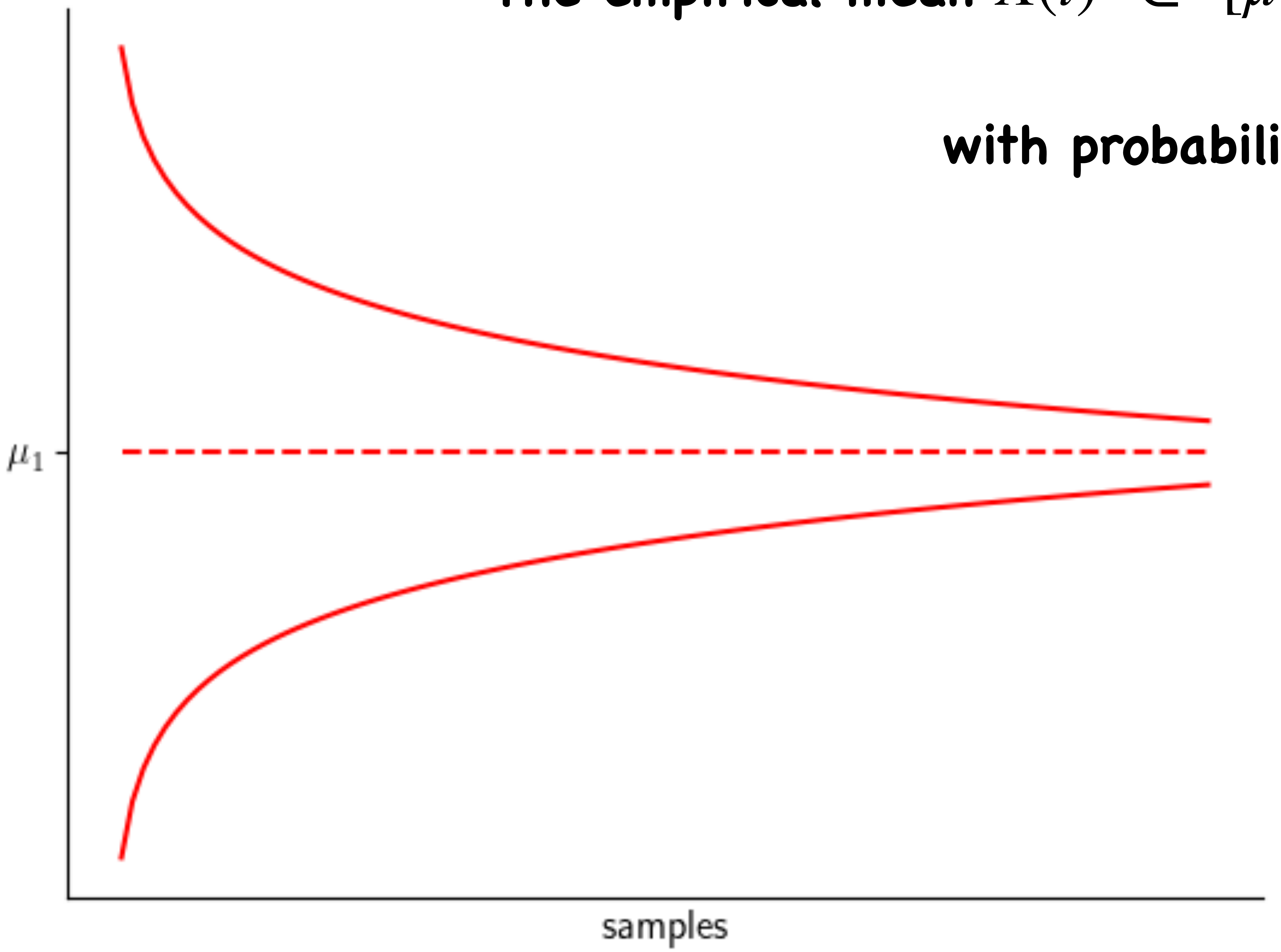
Recall from Hoeffding's inequality

The empirical mean $\bar{X}(t) \in [\mu_1 \pm \alpha_t]$ where $\alpha_t = \sqrt{\frac{4\log(Kt/\delta)}{t}}$

with probability atleast $1-\delta$



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.