- Need to trade of observing vs explicitly playing low probability arms.
- Spend half our exploration time h doing each.
- Assume $q_i \in [0, \frac{1}{2}]$ and order variables such that $q_1 \leq q_2 \leq ... \leq q_N$
- Let $m \in [2, N] = \min i : q_i \geq \frac{1}{i}$
- Divide the arms into low probability $\{(i, 1) : q_i < m\}$ and frequent $\{(i, 0) \forall i \cup (i, 1) : q_i > m\}$
- Divide the $h/2$ explicit play budget between the m low probability arms, giving $\frac{h}{2m}$ samples each.
- For the frequent arms, we expect $\sim q_i \frac{h}{2} \geq \frac{h}{2m}$ samples from the observe phase.