

# UCBアルゴリズム

Step1 それぞれのラウンド $n$ において、広告 $i$ に対して以下の数字を求める

$N_i(n)$  - ラウンド $n$ までに、広告 $i$ が選ばれた回数

$R_i(n)$  - ラウンド $n$ までの、広告 $i$ の累積報酬

Step2 上記の二つの数字において、以下を計算

ラウンド $n$ までの、広告 $i$ からの平均報酬

$$\bar{r}_i(n) = \frac{R_i(n)}{N_i(n)}$$

ラウンド $n$ における信頼区間 $(\bar{r}_i(n) - \Delta_i(n), \bar{r}_i(n) + \Delta_i(n))$

$$\Delta_i(n) = \sqrt{\frac{3 \log(n)}{2 N_i(n)}}$$

Step3 UCBが最大となる広告 $i$ を選ぶ UCB :  $\bar{r}_i(n) + \Delta_i(n)$