A Theorem 2

The proof of this section and the lemmas of the next section follow from the proofs of Gabillon et al. [2012]. The modifications we have made to this proof correspond to the introduction of the function g_k which bounds the uncertainty s_k in order to make it simpler to introduce other models. We also introduce a sufficient condition on this bound, i.e. that it is monotonically decreasing in N in order to bound the arm pulls with respect to g_k^{-1}. Ultimately, this form of the theorem reduces the problem of of proving a regret bound to that of checking a few properties of the uncertainty model.

Theorem 2. *Consider a bandit problem with horizon T and K arms. Let $U_k(t)$ and $L_k(t)$ be upper and lower bounds that hold for all times $t \leq T$ and all arms $k \leq K$ with probability $1 - \delta$. Finally, let g_k be a monotonically decreasing function such that $s_k(t) \leq g_k(N_k(t-1))$ and $\sum_k g_k^{-1}(H_{k\epsilon}) \leq T - K$. We can then bound the simple regret as*

$$\Pr(R_{\Omega_T} \leq \epsilon) \geq 1 - KT\delta. \tag{10}$$

Proof. We will first define the event $\mathcal{E}$ such that on this event every mean is bounded by its associated bounds for all times t. More precisely we can write this as

$$\mathcal{E} = \{\forall k \leq K, \forall t \leq T, L_k(t) \leq \mu_k \leq U_k(t)\}.$$

By definition, these bounds are given such that the probability of deviating from a single bound is δ. Using a union bound we can then bound the probability of remaining within all bounds as $\Pr(\mathcal{E}) \geq 1 - KT\delta$.

We will next condition on the event $\mathcal{E}$ and assume regret of the form $R_{\Omega_T} > \epsilon$ in order to reach a contradiction. Upon reaching said contradiction we can then see that the simple regret must be bounded by ϵ with probability given by the probability of event $\mathcal{E}$, as stated above. As a result we need only show that a contradiction occurs.

We will now define $\tau = \arg\min_{t \leq T} B_{J(t)}(t)$ as the time at which the recommended arm attains the minimum bound, i.e. $\Omega_T = J(\tau)$ as defined in (8). Let $t_k \leq T$ be the last time at which arm k is pulled. Note that each arm must be pulled at least once due to the initialization phase. We can then show the following sequence of inequalities:

$$\begin{aligned}
\min(0, s_k(t_k) - \Delta_k) + s_k(t_k) &\geq B_{J(t_k)}(t_k) && \text{(a)}\\
&\geq B_{\Omega_T}(\tau) && \text{(b)}\\
&\geq R_{\Omega_T} && \text{(c)}\\
&> \epsilon. && \text{(d)}
\end{aligned}$$

Of these inequalities, (a) holds by Lemma B3, (c) holds by Lemma B1, and (d) holds by our assumption on the simple regret. The inequality (b) holds due to the definition Ω_T and time τ. Note, that we can also write the preceding inequality as two cases

$$\begin{aligned}
s_k(t_k) > 2s_k(t_k) - \Delta_k > \epsilon, &\quad \text{if } \Delta_k > s_k(t_k);\\
2s_k(t_k) - \Delta_k \geq s_k(t_k) > \epsilon, &\quad \text{if } \Delta_k \leq s_k(t_k)
\end{aligned}$$

This leads to the following bound on the confidence diameter,

$$s_k(t_k) > \max(\tfrac{1}{2}(\Delta_k + \epsilon), \epsilon) = H_{k\epsilon}$$

which can be obtained by a simple manipulation of the above equations. More precisely we can notice that in each case, $s_k(t_k)$ upper bounds both ϵ and $\frac{1}{2}(\Delta_k + \epsilon)$, and thus it obviously bounds their maximum.

Now, for any arm k we can consider the final number of arm pulls, which we can write as

$$\begin{aligned}
N_k(T) = N_k(t_k - 1) + 1 &\leq g^{-1}(s_k(t_k)) + 1\\
&< g^{-1}(H_{k\epsilon}) + 1.
\end{aligned}$$

This holds due to the definition of g as a monotonic decreasing function, and the fact that we pull each arm at least once during the initialization stage. Finally, by summing both sides with respect to k we can see that $\sum_k g^{-1}(H_{k\epsilon}) + K > T$, which contradicts our definition of g in the Theorem statement. □

B Lemmas

In order to simplify notation in this section, we will first introduce $B(t) = \min_k B_k(t)$ as the minimizer over all gap indices for any time t. We will also note that this term can be rewritten as

$$B(t) = B_{J(t)}(t) = U_{j(t)}(t) - L_{J(t)}(t),$$

which holds due to the definitions of $j(t)$ and $J(t)$.

Lemma B1. *For any sub-optimal arm $k \neq k^*$, any time $t \in \{1, \ldots, T\}$, and on event $\mathcal{E}$, the immediate regret of pulling that arm is upper bounded by the index quantity, i.e. $B_k(t) \geq R_k$.*

Proof. We can start from the definition of the bound and expand this term as

$$\begin{aligned}
B_k(t) &= \max_{i \neq k} U_i(t) - L_k(t)\\
&\geq \max_{i \neq k} \mu_i - \mu_k = \mu^* - \mu_k = R_k.
\end{aligned}$$

The first inequality holds due to the assumption of event $\mathcal{E}$, whereas the following equality holds since we are only considering sub-optimal arms, for which the best alternative arm is obviously the optimal arm. □

Lemma B2. *For any time t let $k = a_t$ be the arm pulled, for which the following statements hold:*

$$\text{if } k = j(t), \text{ then } L_{j(t)}(t) \le L_{J(t)}(t),$$
$$\text{if } k = J(t), \text{ then } U_{j(t)}(t) \le U_{J(t)}(t).$$

Proof. We can divide this proof into two cases based on which of the two arms is selected.

Case 1: let $k = j(t)$ be the arm selected. We will then assume that $L_{j(t)}(t) > L_{J(t)}(t)$ and show that this is a contradiction. By definition of the arm selection rule we know that $s_{j(t)}(t) \ge s_{J(t)}(t)$, from which we can easily deduce that $U_{j(t)}(t) > U_{J(t)}(t)$ by way of our first assumption. As a result we can see that

$$\begin{aligned} B_{j(t)}(t) &= \max_{j \ne j(t)} U_j(t) - L_{j(t)}(t) \\ &< \max_{j \ne J(t)} U_j(t) - L_{J(t)}(t) = B_{J(t)}(t). \end{aligned}$$

This inequality holds due to the fact that arm $j(t)$ must necessarily have the highest upper bound over all arms. However, this contradicts the definition of $J(t)$ and as a result it must hold that $L_{j(t)}(t) \le L_{J(t)}(t)$.

Case 2: let $k = J(t)$ be the arm selected. The proof follows the same format as that used for $k = j(t)$. ☐

Corollary B2. *If arm $k = a_t$ is pulled at time t, then the minimum index is bounded above by the uncertainty of arm k, or more precisely*

$$B(t) \le s_k(t).$$

Proof. We know that k must be restricted to the set $\{j(t), J(t)\}$ by definition. We can then consider the case that $k = j(t)$, and by Lemma B2 we know that this imposes an order on the lower bounds of each possible arm, allowing us to write

$$B(t) \le U_{j(t)}(t) - L_{j(t)}(t) = s_{j(t)}(t)$$

from which our corollary holds. We can then easily see that a similar argument holds for $k = J(t)$ by ordering the upper bounds, again via Lemma B2. ☐

Lemma B3. *On event $\mathcal{E}$, for any time $t \in \{1, \ldots, T\}$, and for arm $k = a_t$ the following bound holds on the minimal gap,*

$$B(t) \le \min(0, s_k(t) - \Delta_k) + s_k(t).$$

Proof. In order to prove this lemma we will consider a number of cases based on which of $k \in \{j(t), J(t)\}$ is selected and whether or not one or neither of these arms corresponds to the optimal arm k^*. Ultimately, this results in six cases, the first three of which we will present are based on selecting arm $k = j(t)$.

Case 1: consider $k^* = k = j(t)$. We can then see that the following sequence of inequalities holds,

$$\mu_{(2)} \overset{(a)}{\ge} \mu_{J(t)}(t) \overset{(b)}{\ge} L_{J(t)}(t) \overset{(c)}{\ge} L_{j(t)}(t) \overset{(d)}{\ge} \mu_k - s_k(t).$$

Here (b) and (d) follow directly from event $\mathcal{E}$ and (c) follows from Lemma B2. Inequality (a) follows trivially from our assumption that $k = k^*$, as a result $J(t)$ can only be as good as the 2nd-best arm. Using the definition of Δ_k and the fact that $k = k^*$, the above inequality yields

$$s_k(t) - (\mu_k - \mu_{(2)}) = s_k(t) - \Delta_k \ge 0$$

Therefore the min in the result of Lemma B3 vanishes and the result follows from Corollary B2.

Case 2: consider $k = j(t)$ and $k^* = J(t)$. We can then write

$$\begin{aligned} B(t) &= U_{j(t)}(t) - L_{J(t)}(t) \\ &\le \mu_{j(t)}(t) + s_{j(t)}(t) - \mu_{J(t)}(t) + s_{J(t)}(t) \\ &\le \mu_k - \mu^* + 2s_k(t) \end{aligned}$$

where the first inequality holds from event $\mathcal{E}$, and the second holds because by definition the selected arm must have higher uncertainty. We can then simplify this as

$$\begin{aligned} &= 2s_k(t) - \Delta_k \\ &\le \min(0, s_k(t) - \Delta_k) + s_k(t), \end{aligned}$$

where the last step evokes Corollary B2.

Case 3: consider $k = j(t) \ne k^*$ and $J(t) \ne k^*$. We can then write the following sequence of inequalities,

$$\mu_{j(t)}(t) + s_{j(t)}(t) \overset{(a)}{\ge} U_{j(t)}(t) \overset{(b)}{\ge} U_{k^*}(t) \overset{(c)}{\ge} \mu^*.$$

Here (a) and (c) hold due to event $\mathcal{E}$ and (b) holds since by definition $j(t)$ has the highest upper bound other than $J(t)$, which in turn is not the optimal arm by assumption in this case. By simplifying this expression we obtain $s_k(t) - \Delta_k \ge 0$, and hence the result follows from Corollary B2 as in Case 1.

Cases 4–6: consider $k = J(t)$. The proofs for these three cases follow the same general form as the above cases and is omitted. Cases 1 through 6 cover all possible scenarios and prove Lemma B3. ☐

Lemma B4. *Consider a normally distributed random variable $X \sim \mathcal{N}(\mu, \sigma^2)$ and $\beta \ge 0$. The probability that X is within a radius of $\beta\sigma$ from its mean can then be written as*

$$\Pr\big(|X - \mu| \le \beta\sigma\big) \ge 1 - e^{-\beta^2/2}.$$

Proof. Consider $Z \sim \mathcal{N}(0,1)$. The probability that Z exceeds some positive bound $c > 0$ can be written

$$\begin{aligned}\Pr(Z > c) &= \frac{e^{-c^2/2}}{\sqrt{2\pi}} \int_c^\infty e^{(c^2-z^2)/2}\, dz \\ &= \frac{e^{-c^2/2}}{\sqrt{2\pi}} \int_c^\infty e^{-(z-c)^2/2 - c(z-c)}\, dz \\ &\leq \frac{e^{-c^2/2}}{\sqrt{2\pi}} \int_c^\infty e^{-(z-c)^2/2}\, dz = \tfrac{1}{2} e^{-c^2/2}.\end{aligned}$$

The inequality holds due to the fact that $e^{-c(z-c)} \leq 1$ for $z \geq c$. Using a union bound we can then bound both sides as $\Pr(|Z| > c) \leq e^{-c^2/2}$. Finally, by setting $Z = (X - \mu)/\sigma$ and $c = \beta$ we obtain the bound stated above. $\square$