

# Summer project

## Learning to bid above a threshold

02344391

May 2024

### Model 1

- Probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  with filtration  $\mathbb{F} = (\mathcal{F}_t)_{t=1}^T$ ;
- Threshold  $\tau_t$  is stationary: for each  $t = 1, \dots, T$ ,  $\tau_t = \tau + \epsilon_t$ , where  $\tau$  is a constant and  $(\epsilon_t)_{t=1}^T$  i.i.d with mean 0 and variance  $\sigma^2$ ;
- At each round  $t$ , we propose a bid:  $b_t$ ;
- At each round  $t$ , we have the information:  $\delta_t = \mathbb{I}_{\{\tau_t \leq b_t\}}$ ;
- Extension: after some delay  $d$ , we observe the true value. Thus for  $t > d$ ,  $\tau_{t-d}$  is  $\mathcal{F}_t$ -measurable.

The goal is to find a strategy that allows for bids above the threshold as quickly as possible while minimising the difference between the threshold and the bid. If we consider the model with the true threshold values provided with a delay, this information should be used to model the distribution of the thresholds, particularly to find the variance and provide a confidence interval that the bid is above the threshold.

We suppose that for any  $t = 1, \dots, T$ ,  $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$ . In this initial model, we aim to write an algorithm that proposes bids to achieve a desired acceptance rate. Indeed, depending on the desired objectives, the strategies can vary: a high acceptance rate might correspond to high bids. In an auction example, this could lead to significant financial loss, and we might prefer a lower acceptance rate.

### $\tau$ is known

Initially, we assume that  $\tau$  is known; the bids will then aim to find the value of  $\sigma^2$ . We can, for example, consider the case of an auction where all items have the same price estimate  $\tau$ , but the price is noisy.

According to the model's assumptions, we have access to data that is right and left censored; indeed,  $\delta_t$  indicates whether bid  $b_t$  is above the threshold  $\tau_t$  or not. In the case where we consider the threshold values known after a certain delay, we also have observed data. We use the Expectation Maximisation (EM) Algorithm to estimate the value of  $\sigma^2$ . ("Expectation-Maximization (EM) was described by Arthur Dempster, Nan Laird, and Donald Rubin in a classic 1977 paper in the Journal of the Royal Statistical Society.")

Observed data are denoted by  $y = (y_i)_{i=1}^N$  generated by  $(Y_i)_{i=1}^N$ , and unobserved data are denoted by  $\mathcal{T} = (\tau_i)_{i=1}^M$ . We know that  $Y_1, \dots, Y_N$  and  $\tau_1, \dots, \tau_M$  are iid  $\sim \mathcal{N}(\tau, \sigma^2)$  where  $\tau$  is known. We observe  $\delta = (\delta_i)_{i=1}^M$ :

$$\delta_i = \begin{cases} 0 & \text{if } \tau_i > b_i \\ 1 & \text{otherwise.} \end{cases}, \quad i = 1, \dots, M.$$

Here, we have reindexed the bids and thresholds to separate the observed data from the unobserved data.

EM algorithm:

”E-step”:

$$Q(\sigma, \sigma_n) = \mathbb{E} [\log(f(\mathcal{T}, Y | \sigma)) | Y = y, \delta, \sigma_n]$$

where  $f$  is the density of the complete data.

We can derive  $Q(\sigma, \sigma_n)$ :

$$\begin{aligned} Q(\sigma, \sigma_n) &= \mathbb{E} [\log(f(\mathcal{T}, Y | \sigma)) | Y = y, \delta, \sigma_n] \\ &= -\frac{N}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^N (y_i - \tau)^2 - \frac{M}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^M \mathbb{E} [(\tau_i - \tau)^2 | \delta_i, \sigma_n] \\ &= -\frac{N+M}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^N (y_i - \tau)^2 - \frac{1}{2\sigma^2} \sum_{i=1}^M \mathbb{E} [(\tau_i - \tau)^2 | \delta_i, \sigma_n] \end{aligned}$$

We need to compute  $\mathbb{E} [(\tau_i - \tau)^2 | \delta_i, \sigma_n]$  for  $i = 1, \dots, M$ . Let  $i \in \{1, \dots, M\}$ , we suppose  $\delta_i = 1$ , thus  $\tau_i \leq b_i$ . Then:

$$\mathbb{E} [(\tau_i - \tau)^2 | \delta_i, \sigma_n] = \frac{1}{\mathbb{P}[\tau_i \leq b_i | \sigma_n]} \mathbb{E} [(\tau_i - \tau)^2 \mathbb{I}_{\{\tau_i \leq b_i\}} | \sigma_n]$$

First, since  $\tau_i | \sigma_n \sim \mathcal{N}(\tau, \sigma_n^2)$ ,

$$\mathbb{P}[\tau_i \leq b_i | \sigma_n] = \mathbb{P} \left[ \frac{\tau_i - \tau}{\sigma_n} \leq \frac{b_i - \tau}{\sigma_n} \middle| \sigma_n \right] = \Phi \left( \frac{b_i - \tau}{\sigma_n} \right)$$

where  $\Phi$  is the CDF of the standard normal distribution.

Then,

$$\begin{aligned} \mathbb{E} [(\tau_i - \tau)^2 \mathbb{I}_{\{\tau_i \leq b_i\}} | \sigma_n] &= \int_{-\infty}^{b_i} \frac{(t - \tau)^2}{\sqrt{2\pi\sigma_n^2}} \exp \left( -\frac{1}{2\sigma_n^2}(t - \tau)^2 \right) dt \\ &= \sigma_n^2 \int_{-\infty}^{b_i} \frac{(t - \tau)^2}{\sigma_n^2} \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{1}{2\sigma_n^2}(t - \tau)^2 \right) \frac{dt}{\sigma_n} \\ &= \sigma_n^2 \int_{-\infty}^{\frac{b_i - \tau}{\sigma_n}} u^2 \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{u^2}{2} \right) du \end{aligned}$$

The PDF of the standard normal distribution is denoted by  $\phi$ , and by integration by parts:

$$\begin{aligned} \int_{-\infty}^{\frac{b_i - \tau}{\sigma_n}} u^2 \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{u^2}{2} \right) du &= [-u\phi(u)]_{-\infty}^{\frac{b_i - \tau}{\sigma_n}} + \int_{-\infty}^{\frac{b_i - \tau}{\sigma_n}} \phi(u) du \\ &= -\frac{b_i - \tau}{\sigma_n} \phi \left( \frac{b_i - \tau}{\sigma_n} \right) + \Phi \left( \frac{b_i - \tau}{\sigma_n} \right) \end{aligned}$$

Finally:

$$\mathbb{E} [(\tau_i - \tau)^2 | \delta_i, \sigma_n] = \sigma_n^2 \left( 1 - \frac{b_i - \tau}{\sigma_n} \frac{\phi \left( \frac{b_i - \tau}{\sigma_n} \right)}{\Phi \left( \frac{b_i - \tau}{\sigma_n} \right)} \right).$$

In a similar way, if  $\delta_i = 0$ :

$$\mathbb{E} [(\tau_i - \tau)^2 | \delta_i, \sigma_n] = \sigma_n^2 \left( 1 + \frac{b_i - \tau}{\sigma_n} \frac{\phi\left(\frac{b_i - \tau}{\sigma_n}\right)}{1 - \Phi\left(\frac{b_i - \tau}{\sigma_n}\right)} \right).$$

We deduce:

$$\begin{aligned} Q(\sigma, \sigma_n) &= -\frac{N+M}{2} \log(2\pi\sigma^2) \\ &\quad - \frac{1}{2\sigma^2} \sum_{i=1}^N (y_i - \tau)^2 \\ &\quad - \frac{\sigma_n^2}{2\sigma^2} \sum_{i=1}^M \left[ \left( 1 - \frac{b_i - \tau}{\sigma_n} \frac{\phi\left(\frac{b_i - \tau}{\sigma_n}\right)}{\Phi\left(\frac{b_i - \tau}{\sigma_n}\right)} \right) \delta_i + \left( 1 + \frac{b_i - \tau}{\sigma_n} \frac{\phi\left(\frac{b_i - \tau}{\sigma_n}\right)}{1 - \Phi\left(\frac{b_i - \tau}{\sigma_n}\right)} \right) (1 - \delta_i) \right] \end{aligned}$$

**"M-step":** We compute  $\frac{\partial Q(\sigma, \sigma_n)}{\partial \sigma^2}$ :

$$\begin{aligned} \frac{\partial Q(\sigma, \sigma_n)}{\partial \sigma^2} &= -\frac{N+M}{2\sigma^2} \\ &\quad + \frac{1}{2\sigma^4} \sum_{i=1}^N (y_i - \tau)^2 \\ &\quad + \frac{\sigma_n^2}{2\sigma^4} \sum_{i=1}^M \left[ \left( 1 - \frac{b_i - \tau}{\sigma_n} \frac{\phi\left(\frac{b_i - \tau}{\sigma_n}\right)}{\Phi\left(\frac{b_i - \tau}{\sigma_n}\right)} \right) \delta_i + \left( 1 + \frac{b_i - \tau}{\sigma_n} \frac{\phi\left(\frac{b_i - \tau}{\sigma_n}\right)}{1 - \Phi\left(\frac{b_i - \tau}{\sigma_n}\right)} \right) (1 - \delta_i) \right] \end{aligned}$$

And  $Q(\sigma, \sigma_n)$  is maximised for:

$$\sigma^{*2} = \frac{1}{N+M} \left[ \sum_{j=1}^N (y_j - \tau)^2 + \sum_{i=1}^M \left( 1 - \tilde{b}_i^{(n)} \frac{\phi\left(\tilde{b}_i^{(n)}\right)}{\Phi\left(\tilde{b}_i^{(n)}\right)} \right) \delta_i + \left( 1 + \tilde{b}_i^{(n)} \frac{\phi\left(\tilde{b}_i^{(n)}\right)}{1 - \Phi\left(\tilde{b}_i^{(n)}\right)} \right) (1 - \delta_i) \right]$$

where  $\tilde{b}_i^{(n)} = \frac{b_i - \tau}{\sigma_n}$ .

**$\tau$  is unknown**

Now we suppose  $\tau$  and  $\sigma^2$  are unknown. In a similar way:

**"E-step":** we compute  $Q((\tau, \sigma), (\tau_n, \sigma_n))$ , thus, we consider:

$$\mathbb{E} [(\tau_i - \tau)^2 | \delta_i, \tau_n, \sigma_n] = \mathbb{E} [\tau_i^2 | \delta_i, \tau_n, \sigma_n] - 2\tau \mathbb{E} [\tau_i | \delta_i, \tau_n, \sigma_n] + \tau^2.$$

Now:

$$\tilde{b}_i^{(n)} = \frac{b_i - \tau_n}{\sigma_n}.$$

**(i)**  $\mathbb{E} [\tau_i^2 | \delta_i, \tau_n, \sigma_n]$   
If  $\delta_i = 1$ :

$$\begin{aligned}
\mathbb{E} [\tau_i^2 | \delta_i, \tau_n, \sigma_n] &= \frac{1}{\Phi(\tilde{b}_i^{(n)})} \int_{-\infty}^{b_i} \frac{x^2}{\sqrt{2\pi}\sigma_n} \exp\left(-\frac{(x - \tau_n)^2}{2\sigma_n^2}\right) dx \\
&= \frac{1}{\Phi(\tilde{b}_i^{(n)})} \int_{-\infty}^{\tilde{b}_i^{(n)}} (\sigma_n u + \tau_n)^2 \exp\left(-\frac{-u^2}{2}\right) du \\
&= \frac{1}{\Phi(\tilde{b}_i^{(n)})} \left( \sigma_n^2 \int_{-\infty}^{\tilde{b}_i^{(n)}} u^2 \phi(u) du + 2\sigma_n \tau_n \int_{-\infty}^{\tilde{b}_i^{(n)}} u \phi(u) du + \tau_n^2 \int_{-\infty}^{\tilde{b}_i^{(n)}} \phi(u) du \right).
\end{aligned}$$

We find:

$$\begin{aligned}
\sigma_n^2 \int_{-\infty}^{\tilde{b}_i^{(n)}} u^2 \phi(u) du &= \sigma_n^2 \left( \Phi(\tilde{b}_i^{(n)}) - \tilde{b}_i^{(n)} \phi(\tilde{b}_i^{(n)}) \right) \\
2\sigma_n \tau_n \int_{-\infty}^{\tilde{b}_i^{(n)}} u \phi(u) du &= -2\sigma_n \tau_n \phi(\tilde{b}_i^{(n)}) \\
\tau_n^2 \int_{-\infty}^{\tilde{b}_i^{(n)}} \phi(u) du &= \tau_n^2 \Phi(\tilde{b}_i^{(n)})
\end{aligned}$$

Therefore,

$$\mathbb{E} [\tau_i^2 | \delta_i, \tau_n, \sigma_n] = \sigma_n^2 + \tau_n^2 - \sigma_n(b_i + \tau_n) \frac{\phi(\tilde{b}_i^{(n)})}{\Phi(\tilde{b}_i^{(n)})}.$$

And if  $\delta_i = 0$ :

$$\mathbb{E} [\tau_i^2 | \delta_i, \tau_n, \sigma_n] = \sigma_n^2 + \tau_n^2 + \sigma_n(b_i + \tau_n) \frac{\phi(\tilde{b}_i^{(n)})}{1 - \Phi(\tilde{b}_i^{(n)})}.$$

(ii)  $\mathbb{E} [\tau_i | \delta_i, \tau_n, \sigma_n]$

If  $\delta_i = 1$  (similar):

$$\mathbb{E} [\tau_i | \delta_i, \tau_n, \sigma_n] = \tau_n - \sigma_n \frac{\phi(\tilde{b}_i^{(n)})}{\Phi(\tilde{b}_i^{(n)})}.$$

If  $\delta_i = 0$ :

$$\mathbb{E} [\tau_i | \delta_i, \tau_n, \sigma_n] = \tau_n + \sigma_n \frac{\phi(\tilde{b}_i^{(n)})}{1 - \Phi(\tilde{b}_i^{(n)})}$$

(iii) Finally:

$$\mathbb{E} [(\tau_i - \tau)^2 | \delta_i, \tau_n, \sigma_n] = \begin{cases} \sigma_n^2 + (\tau - \tau_n)^2 + \sigma_n(2\tau - (b_i + \tau_n)) \frac{\phi(\tilde{b}_i^{(n)})}{\Phi(\tilde{b}_i^{(n)})} & \text{if } \delta_i = 1 \\ \sigma_n^2 + (\tau - \tau_n)^2 - \sigma_n(2\tau - (b_i + \tau_n)) \frac{\phi(\tilde{b}_i^{(n)})}{1 - \Phi(\tilde{b}_i^{(n)})} & \text{otherwise.} \end{cases}.$$

**"M-step":** We compute  $\frac{\partial Q((\tau, \sigma), (\tau_n, \sigma_n))}{\partial \tau}$ :

$$\frac{\partial Q((\tau, \sigma), (\tau_n, \sigma_n))}{\partial \tau} = \frac{1}{\sigma^2} \left( \sum_{i=1}^N y_i - (M+N)\tau + \sum_{i=1}^M \mathbb{E} [\tau_i | \delta_i, \tau_n, \sigma_n] \right)$$

And  $\frac{\partial Q((\tau, \sigma), (\tau_n, \sigma_n))}{\partial \tau} = 0$  for:

$$\tau^* = \frac{1}{M+N} \left( \sum_{j=1}^N y_j + M\tau_n - \sigma_n \sum_{i=1}^M \left[ \frac{\phi(\tilde{b}_i^{(n)})}{\Phi(\tilde{b}_i^{(n)})} \delta_i - \frac{\phi(\tilde{b}_i^{(n)})}{1-\Phi(\tilde{b}_i^{(n)})} (1-\delta_i) \right] \right).$$

Now we compute  $\frac{\partial Q((\tau, \sigma), (\tau_n, \sigma_n))}{\partial \sigma^2}$ :

$$\frac{\partial Q((\tau, \sigma), (\tau_n, \sigma_n))}{\partial \sigma^2} = -\frac{N+M}{2} \frac{1}{\sigma^2} + \frac{1}{2\sigma^4} \left( \sum_{i=1}^N (y_i - \tau)^2 + \sum_{i=1}^M \mathbb{E} [(\tau_i - \tau)^2 | \delta_i, \sigma_n] \right)$$

We deduce  $Q((\tau, \sigma), (\tau_n, \sigma_n))$  is maximised for  $\tau = \tau^*$  and:

$$\begin{aligned} \sigma^{*2} &= \frac{1}{N+M} \left[ \sum_{j=1}^N (y_j - \tau^*)^2 + M(\sigma_n^2 + (\tau^* - \tau_n)^2) \right. \\ &\quad \left. + \sigma_n \sum_{i=1}^M (2\tau^* - (b_i + \tau_n)) \frac{\phi(\tilde{b}_i^{(n)})}{\Phi(\tilde{b}_i^{(n)})} \delta_i - (2\tau^* - (b_i + \tau_n)) \frac{\phi(\tilde{b}_i^{(n)})}{1-\Phi(\tilde{b}_i^{(n)})} (1-\delta_i) \right] \end{aligned}$$

## Bandits

We assume that the threshold values are bounded and that the bounds are known ( $b_{\inf}, b_{\sup}$ ). We partition the domain of the threshold values into different zones where  $J$  arms can propose their bids. At time  $t$ , arm  $J_t \in \{1, \dots, J\}$  proposes a bid.

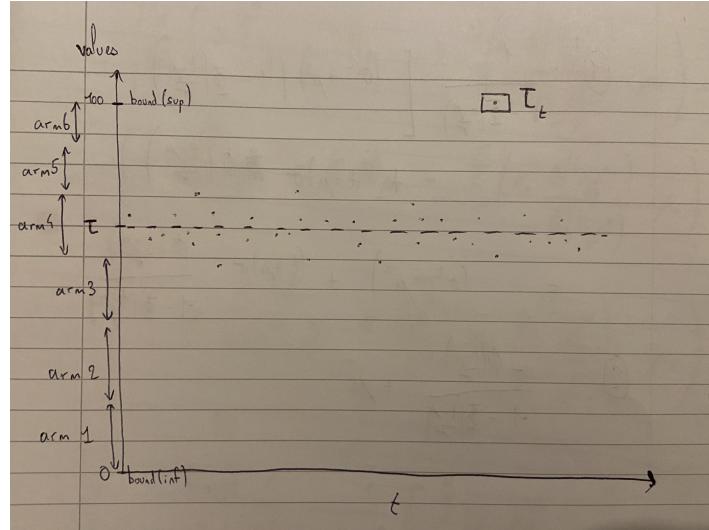


Figure 1: Partitioning of zones for each arm.

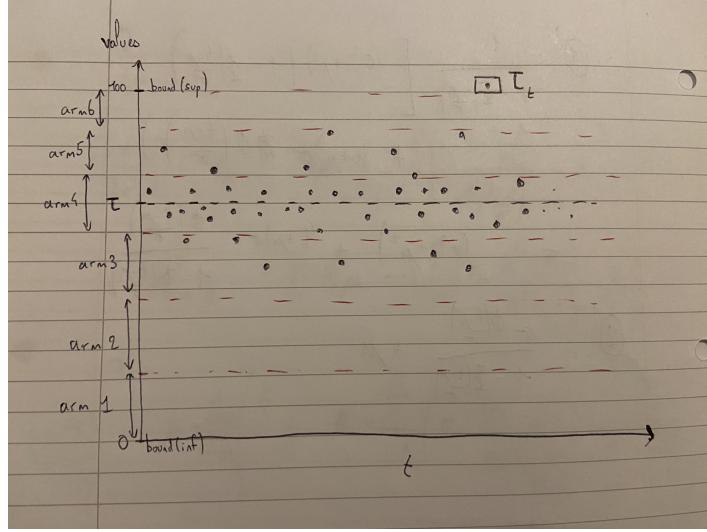


Figure 2: Zones for each arm.

(Lecture notes) "For any arm  $j$ , let

$$N_j(t) = \sum_{s=1}^t \mathbb{I}_{\{J_s=j\}}$$

be the number of times arm  $j$  has been played up to (and including) round  $t$ ."

Now, we need to define the rewards for each proposed bid:

$$X_{j,t} = \delta_t(b_{\text{sup}} - b_t)$$

Here, the reward is higher if the bid is low; we aim to exceed the threshold with a minimised bid. However, if the bid is below the threshold, it is considered a failure, so the reward is 0.

We can define the empirical average reward of each arm as

$$\hat{\mu}_{j,t} = \frac{1}{N_j(t)} \sum_{s=1}^t X_{j,s} \mathbb{I}_{\{J_s=j\}}.$$

(Lecture notes) "Lemma 3.2.1. (Hoeffding's Lemma) If the random variable  $X$  takes values in  $[a, b]$ , then  $X$  is  $\frac{b-a}{2}$  sub-Gaussian.

The regret of an algorithm,  $\pi$ , that plays actions  $J_1, \dots, J_T$  over  $T$  rounds is defined as

$$\mathfrak{R}_T(\pi) = \sum_{t=1}^T (\mu^* - \mu_{J_t})$$

where  $\mu^* = \max_{1 \leq j \leq J} \mu_j$  and  $\mu_j$  is the expected reward of arm  $j$ .  
"

We suppose that arm  $j$  proposes a bid in  $[a_0, a_1]$  ( $0 \leq a_0 < a_1 \leq 1$ ) uniformly and the threshold distribution is  $\mathcal{N}(\tau, \sigma^2)$  (independent of arm  $j$ ). We compute

$$\mu_j = \mathbb{E}[X_{j,t}] = \mathbb{E}[\mathbb{I}_{\{\tau_t \leq b^{(j)}\}} (b_{\text{sup}} - b^{(j)})]$$

where  $b^{(j)} \sim U[a_0, a_1]$  and  $\tau_t \sim \mathcal{N}(\tau, \sigma^2)$ .

We have:

$$\begin{aligned}
\mu_j &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbb{I}_{\{x \leq b\}} (b_{\sup} - b) \frac{\phi(\frac{x-\tau}{\sigma})}{\sigma} \frac{\mathbb{I}_{\{a_0 \leq b \leq a_1\}}}{a_1 - a_0} dx db \\
&= \frac{1}{a_1 - a_0} \int_{a_0}^{a_1} \int_{-\infty}^b (b_{\sup} - b) \frac{\phi(\frac{x-\tau}{\sigma})}{\sigma} dx db \\
&= \frac{1}{a_1 - a_0} \int_{a_0}^{a_1} (b_{\sup} - b) \int_{-\infty}^b \frac{\phi(\frac{x-\tau}{\sigma})}{\sigma} dx db \\
&= \left[ -\frac{(b_{\sup} - b)^2}{2(a_1 - a_0)} \Phi\left(\frac{b - \tau}{\sigma}\right) \right]_{a_0}^{a_1} + \frac{1}{a_1 - a_0} \int_{a_0}^{a_1} \frac{(b_{\sup} - b)^2}{2\sigma} \phi\left(\frac{b - \tau}{\sigma}\right) db.
\end{aligned}$$

And (with  $\tilde{a}_0 = \frac{a_0 - \tau}{\sigma}$  and  $\tilde{a}_1 = \frac{a_1 - \tau}{\sigma}$ ):

$$\begin{aligned}
\int_{a_0}^{a_1} \frac{(b_{\sup} - b)^2}{2\sigma} \phi\left(\frac{b - \tau}{\sigma}\right) db &= \int_{\tilde{a}_0}^{\tilde{a}_1} \frac{(b_{\sup} - \sigma b - \tau)^2}{2} \phi(b) db \\
&= \frac{1}{2} \left( \sigma^2 \int_{\tilde{a}_0}^{\tilde{a}_1} b^2 \phi(b) db \right. \\
&\quad + 2\sigma(\tau - b_{\sup}) \int_{\tilde{a}_0}^{\tilde{a}_1} b \phi(b) db \\
&\quad \left. + (\tau - b_{\sup})^2 \int_{\tilde{a}_0}^{\tilde{a}_1} \phi(b) db \right).
\end{aligned}$$

We have:

$$\begin{aligned}
\sigma^2 \int_{\tilde{a}_0}^{\tilde{a}_1} b^2 \phi(b) db &= \sigma^2 (\Phi(\tilde{a}_1) - \Phi(\tilde{a}_0) - \tilde{a}_1 \phi(\tilde{a}_1) + \tilde{a}_0 \phi(\tilde{a}_0)) \\
2\sigma(\tau - b_{\sup}) \int_{\tilde{a}_0}^{\tilde{a}_1} b \phi(b) db &= 2\sigma(b_{\sup} - \tau) (\phi(\tilde{a}_1) - \phi(\tilde{a}_0)) \\
(\tau - b_{\sup})^2 \int_{\tilde{a}_0}^{\tilde{a}_1} \phi(b) db &= (\tau - b_{\sup})^2 (\Phi(\tilde{a}_1) - \Phi(\tilde{a}_0)).
\end{aligned}$$

Finally:

$$\begin{aligned}
\mu_j &= \frac{\sigma}{2(a_1 - a_0)} \left( \Phi(\tilde{a}_1) (\sigma - \tilde{a}_1 (\tau - 2b_{\sup} + a_1)) - \Phi(\tilde{a}_0) (\sigma - \tilde{a}_0 (\tau - 2b_{\sup} + a_0)) \right. \\
&\quad \left. + \phi(\tilde{a}_1) (2(b_{\sup} - \tau) - \sigma \tilde{a}_1) - \phi(\tilde{a}_0) (2(b_{\sup} - \tau) - \sigma \tilde{a}_0) \right)
\end{aligned}$$

## 1 Zooming algorithm

refs:

- ” Multi-Armed Bandits in Metric Spaces Kleinberg Slivkins Upfal ”
- ” Contextual Bandits with Similarity Information Aleksandrs Slivkins ”

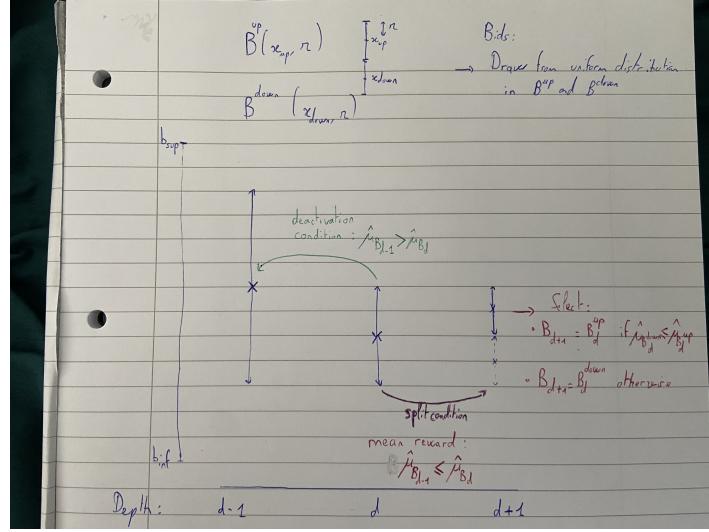


Figure 3: Zooming algorithm.

## 2 Algorithms used so far

In the simulations,  $b_{\text{inf}} = 0$  and  $b_{\text{sup}} = 1$

---

### Algorithm 1 Naive epsilon-greedy algorithm

---

**Require:** Number of zones  $J$  (partition of  $[b_{\text{inf}}, b_{\text{sup}}]$ ), vector of rewards  $\text{rew} = \mathbf{0}_J$ , counter  $\text{count} = \mathbf{0}_J$ , exploration rate  $\epsilon$ , vector of threshold  $\tau$  of length  $T$ .

```

for  $j = 1, \dots, J$  do ▷ Initialisation
    Draw bid  $b$  from a uniform distribution in zone  $j$ .
    Update  $\text{rew}+ = \text{reward}(b, \tau_j)$ ;  $\text{count}_j+ = 1$ 
end for
for  $t = j + 1, \dots, T$  do
    if  $\text{Uniform}([0, 1]) < \epsilon$  then ▷ Exploration
        Choose randomly  $j^* \in \{1, \dots, J\}$ 
    else ▷ Exploitation
         $j^* = \text{argmax}_{j \in \{1, \dots, J\}} (\text{rew}_j / \max(1, \text{count}_j))_{j=1}^J$ 
    end if
    Draw bid  $b$  from a uniform distribution in zone  $j^*$ .
    Update  $\text{rew}_{j^*}+ = \text{reward}(b, \tau_t)$ ;  $\text{count}_{j^*}+ = 1$ 
end for
return  $\text{rew}, \text{count}$ 

```

---



---

### Algorithm 2 Zooming algorithm

---

**Require:** Number of rounds by episode  $T_{ep}$  (even number).

$r = b_{\text{sup}} - b_{\text{inf}}/4$ ;  $B_{\text{inf}} = B(r, r)$ ;  $B_{\text{sup}} = B(3r, r)$ ;  $\text{listB} = [B_{\text{inf}}, B_{\text{sup}} = B(3r, r)]$  ▷ Initialisation

---

## 3 Time series

refs:

” Adapting to a Changing Environment: the Brownian Restless Bandits Aleksandrs Slivkins and Eli Upfal ”

$$\tau_t = \tau(t) + \epsilon_t$$