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 τ_t is stationary: for each $t = 1, \ldots T$, $\tau_t = \tau + \epsilon_t$, where τ is a constant and $(\epsilon_t)_{t=1}^T$ i.i.d with mean 0 and variance σ^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 $t > d$, τ_{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 $t = 1, \ldots T$, $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$. 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.

τ is known

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

According to the model's assumptions, we have access to data that is right and left censored; indeed, δ_t indicates whether bid b_t is above the threshold τ_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 σ^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 $Y_1, \ldots Y_N$ and $\tau_1, \ldots \tau_M$ are iid $\sim \mathcal{N}(\tau, \sigma^2)$ where τ is known. We observe $\delta = (\delta)_{i=1}^M$:

$$\delta_i = \left\{ \begin{array}{ll} 0 & \text{if } \tau_i > b_i \\ 1 & \text{otherwise.} \end{array} \right. , \quad i = 1, \ldots M.$$

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

EM algorithm:

"E-step":

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

where f is the density of the complete data.

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

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

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

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

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

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

where Φ is the CDF of the standard normal distribution.

Then,

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

The PDF of the standard normal distribution is denoted by ϕ, 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:

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

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

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

We deduce:

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

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

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

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

τ is unknown

Now we suppose τ and σ^2 are unknown. In a similar way:

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

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

Now:

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

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

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

$$= \frac{1}{\Phi\left(\tilde{b}_i^{(n)}\right)}\int_{-\infty}^{\tilde{b}_i^{(n)}}(\sigma_n u+\tau_n)^2\exp\left(\frac{-u^2}{2}\right)du$$

$$= \frac{1}{\Phi\left(\tilde{b}_i^{(n)}\right)}\left(\sigma_n^2\int_{-\infty}^{\tilde{b}_i^{(n)}}u^2\phi(u)du + 2\sigma_n\tau_n\int_{-\infty}^{\tilde{b}_i^{(n)}}u\phi(u)du + \tau_n^2\int_{-\infty}^{\tilde{b}_i^{(n)}}\phi(u)du\right).$$

We find:

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

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

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

Therefore,

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

And if $\delta_i = 0$:

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

(ii) $\mathbb{E}\left[\tau_i|\delta_i,\tau_n,\sigma_n\right]$
If $\delta_i = 1$ (similar):

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

If $\delta_i = 0$:

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

(iii) Finally:

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

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

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

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

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

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

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

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

$$\sigma^{*2} = \frac{1}{N+M}\left[\sum_{j=1}^{N}(y_j-\tau^*)^2 + M(\sigma_n^2+(\tau^*-\tau_n)^2)\right.$$

$$\left. +\sigma_n\sum_{i=1}^{M}(2\tau^*-(b_i+\tau_n))\frac{\phi\left(\tilde{b}_i^{(n)}\right)}{\Phi\left(\tilde{b}_i^{(n)}\right)}\delta_i - (2\tau^*-(b_i+\tau_n))\frac{\phi\left(\tilde{b}_i^{(n)}\right)}{1-\Phi\left(\tilde{b}_i^{(n)}\right)}(1-\delta_i)\right]$$

Bandits

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

(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, π, that plays actions $J_1,\ldots,J_T$ 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 μ_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 $\mathcal{N}(\tau, \sigma^2)$ (indendent of arm j). We compute

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

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

We have:

$$\mu_j = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbb{I}_{\{x \leq b\}}(b_{\sup} - b) \frac{\phi(\frac{x-\tau}{\sigma})}{\sigma} \frac{\mathbb{I}_{\{a_0 \leq b \leq a_1\}}}{a_1 - a_0} dx db$$

$$= \frac{1}{a_1 - a_0} \int_{a_0}^{a_1} \int_{-\infty}^{b} (b_{\sup} - b) \frac{\phi\left(\frac{x-\tau}{\sigma}\right)}{\sigma} dx db$$

$$= \frac{1}{a_1 - a_0} \int_{a_0}^{a_1} (b_{\sup} - b) \int_{-\infty}^{b} \frac{\phi\left(\frac{x-\tau}{\sigma}\right)}{\sigma} dx db$$

$$= \left[-\frac{(b_{\sup} - b)^2}{2(a_1 - a_0)} \Phi\left(\frac{b-\tau}{\sigma}\right) \right]_{a_0}^{a_1} + \frac{1}{a_1 - a_0} \int_{a_0}^{a_1} \frac{(b_{\sup} - b)^2}{2\sigma} \phi\left(\frac{b-\tau}{\sigma}\right) db.$$

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

$$\int_{a_0}^{a_1} \frac{(b_{\sup} - b)^2}{2\sigma} \phi\left(\frac{b-\tau}{\sigma}\right) db = \int_{\tilde{a}_0}^{\tilde{a}_1} \frac{(b_{\sup} - \sigma b - \tau)^2}{2} \phi(b) db$$

$$= \frac{1}{2} \left(\sigma^2 \int_{\tilde{a}_0}^{\tilde{a}_1} b^2 \phi(b) db \right.$$

$$+ 2\sigma(\tau - b_{\sup}) \int_{\tilde{a}_0}^{\tilde{a}_1} b\phi(b) db$$

$$\left. + (\tau - b_{\sup})^2 \int_{\tilde{a}_0}^{\tilde{a}_1} \phi(b) db \right).$$

We have:

$$\sigma^2 \int_{\tilde{a}_0}^{\tilde{a}_1} b^2 \phi(b) db = \sigma^2 \left(\Phi(\tilde{a}_1) - \Phi(\tilde{a}_0) - \tilde{a}_1 \phi(\tilde{a}_1) + \tilde{a}_0 \phi(\tilde{a}_0) \right)$$

$$2\sigma(\tau - b_{\sup}) \int_{\tilde{a}_0}^{\tilde{a}_1} b\phi(b) db = 2\sigma(b_{\sup} - \tau) \left(\phi(\tilde{a}_1) - \phi(\tilde{a}_0) \right)$$

$$(\tau - b_{\sup})^2 \int_{\tilde{a}_0}^{\tilde{a}_1} \phi(b) db = (\tau - b_{\sup})^2 \left(\Phi(\tilde{a}_1) - \Phi(\tilde{a}_0) \right).$$

Finally:

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

$$\left. + \phi(\tilde{a}_1) \left(2(b_{\sup} - \tau) - \sigma \tilde{a}_1 \right) - \phi(\tilde{a}_0) \left(2(b_{\sup} - \tau) - \sigma \tilde{a}_0 \right) \right)$$

Now, we compute the second moment:

$$\mathbb{E}[X_{j,t}^2] = \mathbb{E}[\mathbb{I}_{\{\tau_t \le b^{(j)}\}}(b_{\sup} - b^{(j)})^2]$$

$$= \frac{1}{a_1 - a_0} \int_{a_0}^{a_1} (b_{\sup} - b)^2 \int_{-\infty}^{b} \frac{\phi\left(\frac{x-\tau}{\sigma}\right)}{\sigma} dx \, db$$

$$= \left[-\frac{(b_{\sup} - b)^3}{3(a_1 - a_0)} \Phi\left(\frac{b - \tau}{\sigma}\right)\right]_{a_0}^{a_1} + \frac{1}{a_1 - a_0} \int_{a_0}^{a_1} \frac{(b_{\sup} - b)^3}{3\sigma} \phi\left(\frac{b - \tau}{\sigma}\right) db.$$

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

$$\int_{a_0}^{a_1} \frac{(b_{\sup} - b)^3}{3\sigma} \phi\left(\frac{b - \tau}{\sigma}\right) db = \int_{\tilde{a}_0}^{\tilde{a}_1} \frac{(b_{\sup} - \sigma b - \tau)^3}{3} \phi(b) \, db$$

$$= \frac{1}{3}\left(-\sigma^3 \int_{\tilde{a}_0}^{\tilde{a}_1} b^3 \phi(b) \, db \right.$$

$$+ 3\sigma^2(b_{\sup} - \tau) \int_{\tilde{a}_0}^{\tilde{a}_1} b^2 \phi(b) \, db$$

$$- 3\sigma(b_{\sup} - \tau)^2 \int_{\tilde{a}_0}^{\tilde{a}_1} b \phi(b) \, db$$

$$\left. + (b_{\sup} - \tau)^3 \int_{\tilde{a}_0}^{\tilde{a}_1} \phi(b) \, db\right).$$

First,

$$\sigma^3 \int_{\tilde{a}_0}^{\tilde{a}_1} b^3 \phi(b) \, db = \sigma^3\left(\tilde{a}_0^2 \phi(\tilde{a}_0) - \tilde{a}_1^2 \phi(\tilde{a}_1) + 2\int_{\tilde{a}_0}^{\tilde{a}_1} b\phi(b)db\right)$$

Then we have:

$$\mathbb{E}[X_{j,t}^2] = \frac{1}{3(a_1 - a_0)}\left((b_{\sup} - a_0)^3 \Phi(\tilde{a}_0) - (b_{\sup} - a_1)^3 \Phi(\tilde{a}_1)\right.$$

$$+ \sigma^3\left(\tilde{a}_1^2 \phi(\tilde{a}_1) - \tilde{a}_0^2 \phi(\tilde{a}_0) + 2\left(\phi(\tilde{a}_1) - \phi(\tilde{a}_0)\right)\right)$$

$$+ 3\sigma^2(b_{\sup} - \tau)\left(\Phi(\tilde{a}_1) - \Phi(\tilde{a}_0) - \tilde{a}_1 \phi(\tilde{a}_1) + \tilde{a}_0 \phi(\tilde{a}_0)\right)$$

$$+ 3\sigma(b_{\sup} - \tau)^2\left(\phi(\tilde{a}_1) - \phi(\tilde{a}_0)\right)$$

$$\left. + (b_{\sup} - \tau)^3\left(\Phi(\tilde{a}_1) - \Phi(\tilde{a}_0)\right)\right)$$

$$= \frac{1}{3(a_1 - a_0)}\left(\Phi(\tilde{a}_1)\left((b_{\sup} - \tau)^3 + 3\sigma^2(b_{\sup} - \tau) - (b_{\sup} - a_1)^3\right)\right.$$

$$- \Phi(\tilde{a}_0)\left((b_{\sup} - \tau)^3 + 3\sigma^2(b_{\sup} - \tau) - (b_{\sup} - a_0)^3\right)$$

$$+ \phi(\tilde{a}_1)\left(\sigma^3\left(\tilde{a}_1^2 + 2\right) + 3\sigma(b_{\sup} - \tau)^2 - 3\sigma^2(b_{\sup} - \tau)\tilde{a}_1\right)$$

$$\left. - \phi(\tilde{a}_0)\left(\sigma^3\left(\tilde{a}_0^2 + 2\right) + 3\sigma(b_{\sup} - \tau)^2 - 3\sigma^2(b_{\sup} - \tau)\tilde{a}_0\right)\right)$$

From Bernstein's inequality for $\alpha > 0$, since $|\hat{\mu}_{j,t} - \mu_j| \le b_{sup}$, here $b_{sup} = 1$:

$$\mathbb{P}\left[|\hat{\mu}_{j,t} - \mu_j| > \alpha\right] \le 2\exp\left(-\frac{N_j(t)\alpha^2}{2\mathbb{V}[X_j] + \frac{2}{3}\alpha}\right)$$

Corollary: since $|\hat{\mu}_{j,t} - \mu_j| \le 1$, with probability $1 - \delta$:

7

$$|\hat{\mu}_{j,t} - \mu_j| \le \sqrt{\frac{4\mathbb{V}[X_j]\log\frac{2}{\delta}}{N_j(t)}} + \frac{4\log\frac{2}{\delta}}{3N_j(t)}$$

Now, we want to find $n_\delta \in \mathbb{N}$ such that $|\hat{\mu}_{j,t} - \mu_j| \le \alpha$ with probability $1 - \delta$. We need to solve, for $z \in \mathbb{R}$:

$$-\alpha z^2 + \sqrt{4\mathbb{V}[X_j]\kappa}z + \frac{4}{3}\kappa = 0$$

where $\kappa = \log\frac{2}{\delta}$.

The roots are:

$$z_{\pm} = \frac{\sqrt{\mathbb{V}[X_j]\kappa} \pm \sqrt{\mathbb{V}[X_j]\kappa + \frac{4}{3}\alpha\kappa}}{\alpha}$$

Finally, we find:

$$n_\delta \ge \log\frac{2}{\delta}\left(\frac{\sqrt{\mathbb{V}[X_j]} + \sqrt{\mathbb{V}[X_j] + \frac{4}{3}\alpha}}{\alpha}\right)^2$$

We need to find a way to choose α. Given different σ, the expected rewards are plot in figure (1), for a partition of 50 increments.
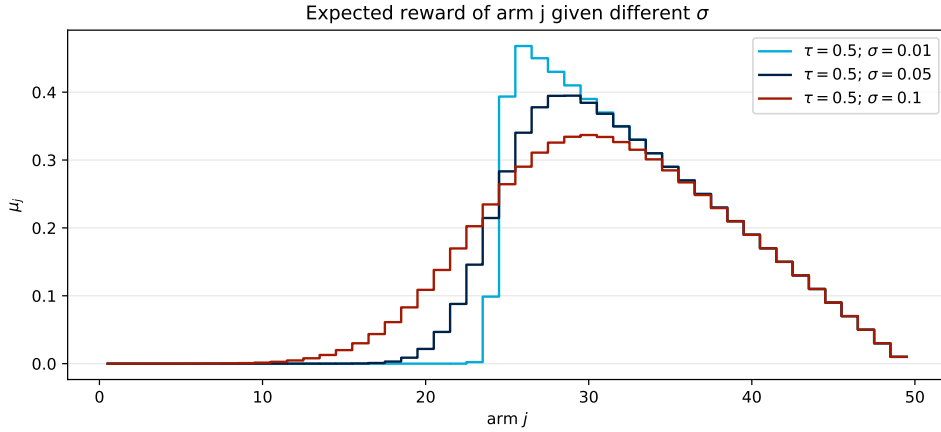

Figure 1: Expected rewards for each arm (partition $[i/50, (i+1)/50], i = 0, \ldots 49$)

When three arms are aligned, a rule is needed to definitively select the middle arm. In the algorithm used, the rule is simple: for a period of n_{ep}, the algorithm selects the middle arm with a probability p and the other arms equally with a probability of $1 - p$. As long as the best arm is the middle one, the algorithm increases the probability p until it tends towards 1.

1   Last update

Sequential halving Algorithm with EM

- Based on a predefined partition, run the sequential halving algorithm until 3 arms are aligned;

- Place bids on the three selected arms. If the empirical mean of the middle arm is not the largest, return to the sequential halving algorithm step;

- Otherwise, run the EM algorithm on these bids to obtain an estimate of the standard deviation ($\hat{\sigma}$) and the mean ($\hat{\tau}$) of the threshold distribution;

- Calculate the variance estimate for the middle arm using $\hat{\sigma}$ and $\hat{\tau}$, and calculate the estimate of half the difference between the mean of the best arm and that of the second best;

- Calculate the number of iterations ($n_{\text{Bernstein}}$) to construct a confidence interval according to Bernstein's inequality;

- Play the three selected arms $n_{\text{Bernstein}}$ times. If the middle arm is still the best, it is chosen; otherwise, return to the sequential halving algorithm step.

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 $rew = \mathbf{0}_J$, counter $count = \mathbf{0}_J$, exploration rate ϵ, vector of threshold τ of length T.
    for $j = 1, \ldots, J$ **do**   ▷ Initialisation
        Draw bid b from a uniform distribution in zone j.
        Update $rew+ = reward(b, \tau_j)$; $count_j+ = 1$
    end for
    for $t = j + 1, \ldots, T$ **do**
        if Uniform($[0, 1]$) $< \epsilon$ **then**   ▷ Exploration
            Choose randomly $j^* \in \{1, \ldots J\}$
        else   ▷ Exploitation
            $j^* = \text{argmax}_{j \in \{1, \ldots J\}} (rew_j / \max(1, count_j))_{j=1}^{J}$
        end if
        Draw bid b from a uniform distribution in zone j^*
        Update $rew_{j^*}+ = reward(b, \tau_t)$; $count_{j^*}+ = 1$
    end for
    return rew, $count$

---

3 Time series

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

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