Dealing with Unknown Variances in Best-Arm Identification

Marc Jourdan, Rémy Degenne and Emilie Kaufmann

February 21, 2023

Outline

Goal: Identify the item having the highest average return.

Common assumption: Gaussian with known variance.

⚠ Too restrictive !

This paper:

Unknown variance !

Two approaches to deal with unknown variances:

- ☞ **Plug in** the empirical variance,
- ☞ **Adapt** the transportation costs.

Outline

Goal: Identify the item having the highest average return.

Common assumption: Gaussian with known variance.

⚠ Too restrictive !

This paper:

Unknown variance !

Two approaches to deal with unknown variances:

- ☞ **Plug in** the empirical variance,
- ☞ **Adapt** the transportation costs.

Outline

Goal: Identify the item having the highest average return.

Common assumption: Gaussian with known variance.

⚠ Too restrictive !

This paper:

Unknown variance !

Two approaches to deal with unknown variances:

☞ **Plug in** the empirical variance,

☞ **Adapt** the transportation costs.

Best-arm identification (BAI)

K arms, $\nu_a \in \mathcal{D}$ distribution of arm $a \in [K]$

☞ $\nu_a = \mathcal{N}(\mu_a, \sigma_a^2)$ where (μ_a, σ_a^2) are unknown.

Goal: identify unique $a^\star = \arg\max_a \mu_a$ with confidence $1 - \delta$.

Algorithm: at time t,

- **Sequential test**: if the stopping time τ_δ is reached, then return the candidate answer $\hat{a}_t$.
- **Sampling rule**: pull arm a_t and observe $X_t \sim \nu_{a_t}$.

Objective: Minimize $\mathbb{E}_{\boldsymbol{\nu}}[\tau_\delta]$ for δ-correct algorithms

$$\mathbb{P}_{\boldsymbol{\nu}}\left[\tau_\delta < +\infty,\ \hat{a}_{\tau_\delta} \neq a^\star\right] \leq \delta\,.$$

Best-arm identification (BAI)

K arms, $\nu_a \in \mathcal{D}$ distribution of arm $a \in [K]$

☞ $\nu_a = \mathcal{N}(\mu_a, \sigma_a^2)$ where (μ_a, σ_a^2) are unknown.

Goal: identify unique $a^\star = \arg\max_a \mu_a$ with confidence $1 - \delta$.

Algorithm: at time t,

- **Sequential test**: if the stopping time τ_δ is reached, then return the candidate answer $\hat{a}_t$.
- **Sampling rule**: pull arm a_t and observe $X_t \sim \nu_{a_t}$.

Objective: Minimize $\mathbb{E}_{\boldsymbol{\nu}}[\tau_\delta]$ for δ-correct algorithms

$$\mathbb{P}_{\boldsymbol{\nu}}\left[\tau_\delta < +\infty,\ \hat{a}_{\tau_\delta} \neq a^\star\right] \leq \delta \,.$$

Best-arm identification (BAI)

K arms, $\nu_a \in \mathcal{D}$ distribution of arm $a \in [K]$

☞ $\nu_a = \mathcal{N}(\mu_a, \sigma_a^2)$ where (μ_a, σ_a^2) are unknown.

Goal: identify unique $a^\star = \arg\max_a \mu_a$ with confidence $1 - \delta$.

Algorithm: at time t,

- **Sequential test**: if the stopping time τ_δ is reached, then return the candidate answer $\hat{a}_t$.
- **Sampling rule**: pull arm a_t and observe $X_t \sim \nu_{a_t}$.

Objective: Minimize $\mathbb{E}_{\boldsymbol{\nu}}[\tau_\delta]$ for δ-correct algorithms

$$\mathbb{P}_{\boldsymbol{\nu}} \left[\tau_\delta < +\infty,\ \hat{a}_{\tau_\delta} \neq a^\star\right] \leq \delta \,.$$

Sample complexity lower bound

Garivier and Kaufmann (2016): For all δ-correct algorithm,

$$\forall \boldsymbol{\nu} \in \mathcal{D}^K, \quad \liminf_{\delta \to 0} \frac{\mathbb{E}_{\boldsymbol{\nu}}[\tau_\delta]}{\log(1/\delta)} \ge T^\star(\mu, \sigma^2) \, ,$$

where $T^\star(\mu, \sigma^2)^{-1} = \max_{w \in \triangle_K} \min_{a \neq a^\star} C(a^\star, a; w)$ and

$$2C(a^\star, a; w) = \inf_{\lambda \in (\mu_a, \mu_{a^\star})} \sum_{b \in \{a^\star, a\}} w_b \log\left(1 + \frac{(\mu_b - \lambda)^2}{\sigma_b^2}\right) .$$

Known variance

$$2C_{\sigma^2}(a^\star, a; w) = \inf_{\lambda \in (\mu_a, \mu_{a^\star})} \sum_{b \in \{a^\star, a\}} w_b \frac{(\mu_b - \lambda)^2}{\sigma_b^2} = \frac{(\mu_{a^\star} - \mu_a)^2}{\sigma_{a^\star}^2 / w_{a^\star} + \sigma_a^2 / w_a} .$$

Sample complexity lower bound

Garivier and Kaufmann (2016): For all δ-correct algorithm,

$$\forall \boldsymbol{\nu} \in \mathcal{D}^K, \quad \liminf_{\delta \to 0} \frac{\mathbb{E}_{\boldsymbol{\nu}}[\tau_\delta]}{\log(1/\delta)} \geq T^\star(\mu, \sigma^2) ,$$

where $T^\star(\mu, \sigma^2)^{-1} = \max_{w \in \triangle_K} \min_{a \neq a^\star} C(a^\star, a; w)$ and

$$2C(a^\star, a; w) = \inf_{\lambda \in (\mu_a, \mu_{a^\star})} \sum_{b \in \{a^\star, a\}} w_b \log\left(1 + \frac{(\mu_b - \lambda)^2}{\sigma_b^2}\right) .$$

Known variance

$$2C_{\sigma^2}(a^\star, a; w) = \inf_{\lambda \in (\mu_a, \mu_{a^\star})} \sum_{b \in \{a^\star, a\}} w_b \frac{(\mu_b - \lambda)^2}{\sigma_b^2} = \frac{(\mu_{a^\star} - \mu_a)^2}{\sigma_{a^\star}^2 / w_{a^\star} + \sigma_a^2 / w_a} .$$

Sample complexity lower bound

Garivier and Kaufmann (2016): For all δ-correct algorithm,

$$\forall \boldsymbol{\nu} \in \mathcal{D}^K, \quad \liminf_{\delta \to 0} \frac{\mathbb{E}_{\boldsymbol{\nu}}[\tau_\delta]}{\log(1/\delta)} \geq T^\star(\mu, \sigma^2) \ ,$$

where $T^\star(\mu, \sigma^2)^{-1} = \max_{w \in \triangle_K} \min_{a \neq a^\star} C(a^\star, a; w)$ and

$$2C(a^\star, a; w) = \inf_{\lambda \in (\mu_a, \mu_{a^\star})} \sum_{b \in \{a^\star, a\}} w_b \log\left(1 + \frac{(\mu_b - \lambda)^2}{\sigma_b^2}\right) \ .$$

Known variance

$$2C_{\sigma^2}(a^\star, a; w) = \inf_{\lambda \in (\mu_a, \mu_{a^\star})} \sum_{b \in \{a^\star, a\}} w_b \frac{(\mu_b - \lambda)^2}{\sigma_b^2} = \frac{(\mu_{a^\star} - \mu_a)^2}{\sigma_{a^\star}^2 / w_{a^\star} + \sigma_a^2 / w_a} \ .$$

How to obtain a δ-correct sequential test ?

☞ recommend the empirical best arm

$$\hat{a}_t = \arg\max_{a \in [K]} \mu_{t,a} \ ,$$

with $N_{t,a} = \sum_{s \in [t]} \mathbb{1}\left(a_s = a\right)$ and MLE (μ_t, σ_t^2) defined as

$$\mu_{t,a} = \frac{1}{N_{t,a}} \sum_{s \in [t]} \mathbb{1}\left(a_s = a\right) X_s \quad \text{and} \quad \sigma_{t,a}^2 = \frac{1}{N_{t,a}} \sum_{s \in [t]} \mathbb{1}\left(a_s = a\right) (X_s - \mu_{t,a})^2 .$$

☞ calibrated GLR and EV-GLR stopping rules.

How to obtain a δ-correct sequential test ?

☞ recommend the empirical best arm

$$\hat{a}_t = \arg\max_{a \in [K]} \mu_{t,a} \ ,$$

with $N_{t,a} = \sum_{s \in [t]} \mathbb{1}(a_s = a)$ and MLE (μ_t, σ_t^2) defined as

$$\mu_{t,a} = \frac{1}{N_{t,a}} \sum_{s \in [t]} \mathbb{1}(a_s = a) X_s \quad \text{and} \quad \sigma_{t,a}^2 = \frac{1}{N_{t,a}} \sum_{s \in [t]} \mathbb{1}(a_s = a) (X_s - \mu_{t,a})^2 .$$

☞ calibrated GLR and EV-GLR stopping rules.

GLR stopping rule **[Adapt]**

$$\tau_\delta = \inf\{t \in \mathbb{N} \mid \forall a \neq \hat{a}_t,\ Z_a(t) > c_{\hat{a}_t, a}(N_t, \delta)\} \ ,$$

$$2Z_a(t) = \inf_{\lambda \in [\mu_{t,a}, \mu_{t,\hat{a}_t}]} \sum_{b \in \{\hat{a}_t, a\}} N_{t,b} \log\left(1 + \frac{(\mu_{t,b} - \lambda)^2}{\sigma_{t,b}^2}\right) \ ,$$

where $(c_{a,b})_{a \neq b}$ is a family of thresholds.

EV-GLR stopping rule **[Plug in]**

$$\tau_\delta^{\mathsf{EV}} = \inf\{t \in \mathbb{N} \mid \forall a \neq \hat{a}_t,\ Z_a^{\mathsf{EV}}(t) > c_{\hat{a}_t, a}^{\mathsf{EV}}(N_t, \delta)\} \ ,$$

$$2Z_a^{EV}(t) = \frac{(\mu_{t,\hat{a}_t} - \mu_{t,a})^2}{\sigma_{t,\hat{a}_t}^2 / N_{t,\hat{a}_t} + \sigma_{t,a}^2 / N_{t,a}} \ ,$$

where $(c_{a,b}^{\mathsf{EV}})_{a \neq b}$ is a family of thresholds.

Stopping rules

GLR stopping rule **[Adapt]**

$$\tau_\delta = \inf\{t \in \mathbb{N} \mid \forall a \neq \hat{a}_t,\ Z_a(t) > c_{\hat{a}_t,a}(N_t, \delta)\} \,,$$

$$2 Z_a(t) = \inf_{\lambda \in [\mu_{t,a}, \mu_{t,\hat{a}_t}]} \sum_{b \in \{\hat{a}_t, a\}} N_{t,b} \log\left(1 + \frac{(\mu_{t,b} - \lambda)^2}{\sigma_{t,b}^2}\right) \,,$$

where $(c_{a,b})_{a \neq b}$ is a family of thresholds.

EV-GLR stopping rule **[Plug in]**

$$\tau_\delta^{\mathrm{EV}} = \inf\{t \in \mathbb{N} \mid \forall a \neq \hat{a}_t,\ Z_a^{\mathrm{EV}}(t) > c^{\mathrm{EV}}_{\hat{a}_t,a}(N_t, \delta)\} \,,$$

$$2 Z_a^{EV}(t) = \frac{(\mu_{t,\hat{a}_t} - \mu_{t,a})^2}{\sigma^2_{t,\hat{a}_t}/N_{t,\hat{a}_t} + \sigma^2_{t,a}/N_{t,a}} \,,$$

where $(c^{\mathrm{EV}}_{a,b})_{a \neq b}$ is a family of thresholds.

Example: **GLR** stopping rule **[Adapt]**

☞ Calibration by time-uniform concentration: with probability $1 - \delta$,

$$\forall t \in \mathbb{N}, \forall a \neq a^\star, \quad \sum_{b \in \{a, a^\star\}} N_{t,b} \log \left(1 + \frac{(\mu_{t,b} - \mu_b)^2}{\sigma_{t,b}^2} \right) \leq 2 c_{a,a^\star}(N_t, \delta) \, .$$

Per-arm concentration:

☞ **Student thresholds**, quantiles-based as $(\mu_{t,b} - \mu_b)/\sigma_{t,b} \sim \mathcal{T}_{N_{t,a}-1}$.

☞ **Box thresholds**, combining confidence regions on $\mu_{t,a}$ and $\sigma_{t,a}^2$.

Calibration of the stopping thresholds

Example: **GLR** stopping rule **[Adapt]**

- Calibration by time-uniform concentration: with probability $1-\delta$,

$$\forall t \in \mathbb{N}, \forall a \neq a^\star, \quad \sum_{b \in \{a, a^\star\}} N_{t,b} \log \left(1 + \frac{(\mu_{t,b} - \mu_b)^2}{\sigma_{t,b}^2} \right) \leq 2 c_{a,a^\star}(N_t, \delta) \, .$$

Per-arm concentration:

- **Student thresholds**, quantiles-based as $(\mu_{t,b} - \mu_b)/\sigma_{t,b} \sim \mathcal{T}_{N_{t,a}-1}$.
- **Box thresholds**, combining confidence regions on $\mu_{t,a}$ and $\sigma_{t,a}^2$.

Calibration of the stopping thresholds

Example: **GLR** stopping rule **[Adapt]**

☞ Calibration by time-uniform concentration: with probability $1-\delta$,

$$\forall t \in \mathbb{N}, \forall a \neq a^\star, \quad \sum_{b \in \{a, a^\star\}} N_{t,b} \log \left(1 + \frac{(\mu_{t,b} - \mu_b)^2}{\sigma_{t,b}^2} \right) \leq 2 c_{a,a^\star}(N_t, \delta) \, .$$

Per-arm concentration:

☞ **Student thresholds**, quantiles-based as $(\mu_{t,b} - \mu_b)/\sigma_{t,b} \sim \mathcal{T}_{N_{t,a}-1}$.

☞ **Box thresholds**, combining confidence regions on $\mu_{t,a}$ and $\sigma_{t,a}^2$.

Concentration of the empirical variances

Theorem

With probability $1-\delta$,

$$\forall t \in \mathbb{N}, \quad \sigma_{t+1}^2/\sigma^2 - 1 \lesssim 2\left(\log(1/\delta) + \log\log t\right)/t \,,$$

$$\forall t \geq \frac{2\log(1/\delta)}{\log\log(1/\delta)}, \quad \sigma_{t+1}^2/\sigma^2 - 1 \gtrsim -2\left(\log(1/\delta) + \log\log t\right)/t \,.$$

Proof idea: "peeling" method on sub-Exp processes (Howard et al., 2020).

Beyond box: pairwise concentration

☞ **KL thresholds**

Theorem

With probability $1-\delta$,

$$\forall t \in \mathbb{N}, \forall a \neq a^\star, \quad \sum_{b \in \{a, a^\star\}} N_{t,b} \, \mathrm{KL}((\mu_{t,b}, \sigma_{t,b}^2), (\mu_b, \sigma_b^2)) \leq c_{a,a^\star}(N_t, \delta) \,,$$

where $c_{a,b}(N, \delta) = +\infty$ *if* $\min\{N_a, N_b\} \lesssim \frac{2\log(1/\delta)}{\log\log(1/\delta)}$, *else*

$$c_{a,b}(N, \delta) \approx \log(1/\delta) + \sum_{c \in \{a,b\}} \log\log N_c \,.$$

Proof idea: "peeling" with a crude per-arm concentration to do a quadratic approximation of KL, hence obtaining concentration on the sum of KL.

Best of Both (BoB) thresholds

Theorem

*The family of **BoB thresholds** is δ-correct for the **GLR** stopping rule. It is defined as $c_{a,b}(N, \delta) = +\infty$ if $\min\{N_a, N_b\} \lesssim \frac{2\log(1/\delta)}{\log\log(1/\delta)}$, else solution of*

$$\text{maximize} \quad \frac{1}{2} \sum_{c \in \{a,b\}} N_c \log(1 + y_c) \quad \text{under the constraints}$$

$$\forall c \in \{a, b\}, \quad y_c \geq 0, \quad \max\{x_c y_c, 1 - x_c\} \lesssim \frac{2}{t}(\log(1/\delta) + \log\log N_c),$$

$$\frac{1}{2} \sum_{c \in \{a,b\}} N_c \left((1 + y_c)x_c - 1 - \log x_c\right) \lesssim \log(1/\delta) + \sum_{c \in \{a,b\}} \log\log N_c \, .$$

Proof idea: combine per-arm and pairwise concentration (Box and KL).

Simulations

$\mu = (0, -0.2)$, $\sigma^2 = (1, 0.5)$, uniform sampling.

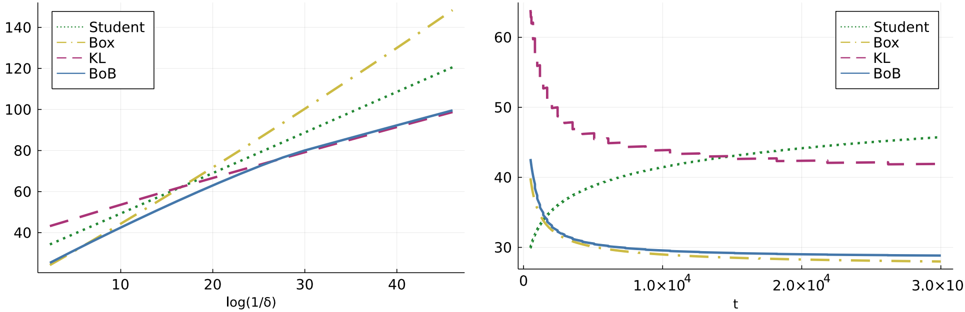

Figure: Thresholds for the GLR stopping rule as a function of (a) $\log(1/\delta)$ for $t = 5000$ and (b) t for $\delta = 0.01$.

Sampling rule wrappers

Example: **EB-TCI** (Jourdan et al., 2022)

☞ sample leader $B^{\text{EB}}_{t+1} = \hat{a}_t$ with probability $1/2$, else sample challenger

$$\textbf{[Adapt]} \quad C^{\text{TCI}}_{t+1} = \arg\min_{a \neq \hat{a}_t} \{ Z_a(t) + \log N_{t,a} \} \ ,$$

$$\textbf{[Plug in]} \quad C^{\text{EVTCI}}_{t+1} = \arg\min_{a \neq \hat{a}_t} \{ Z^{\text{EV}}_a(t) + \log N_{t,a} \} \ ,$$

Other BAI algorithms studied with the [Adapt]/[Plug in] wrappers:

- Track-and-Stop (Garivier and Kaufmann, 2016),
- DKM (Degenne et al., 2019) *[empirically]*,
- FWS (Wang et al., 2021) *[empirically]*.

Sampling rule wrappers

Example: **EB-TCI** (Jourdan et al., 2022)

☞ sample leader $B_{t+1}^{\text{EB}} = \hat{a}_t$ with probability $1/2$, else sample challenger

$$\textbf{[Adapt]} \quad C_{t+1}^{\text{TCI}} = \arg\min_{a \neq \hat{a}_t} \{ Z_a(t) + \log N_{t,a} \} ,$$

$$\textbf{[Plug in]} \quad C_{t+1}^{\text{EVTCI}} = \arg\min_{a \neq \hat{a}_t} \{ Z_a^{\text{EV}}(t) + \log N_{t,a} \} ,$$

Other BAI algorithms studied with the **[Adapt]**/**[Plug in]** wrappers:

- Track-and-Stop (Garivier and Kaufmann, 2016),
- DKM (Degenne et al., 2019) *[empirically]*,
- FWS (Wang et al., 2021) *[empirically]*.

Sample complexity upper bound

Theorem ([Adapt])

Using the GLR stopping with an asymptotically tight family of thresholds, EB-TCI satisfies that, for instances $\nu \in \mathcal{D}^K$ having distinct means,

$$\limsup_{\delta \to 0} \frac{\mathbb{E}_\nu [\tau_\delta]}{\log(1/\delta)} \le T^\star_{1/2}(\nu) \, .$$

Asymptotically tight threshold, i.e. $c(\cdot, \delta) \sim_{\delta \to 0} \log(1/\delta)$.

☞ KL and BoB thresholds are asymptotically tight (not Student and Box).

Theorem ([Plug in])

For all asymptotically tight family of thresholds $(c_{a,b})_{a \neq b}$ and problem independent constant $\alpha > 0$, combining EB-EVTCI with the EV-GLR stopping rule using $(\alpha c_{a,b})_{a \neq b}$ yields an algorithm which is not δ-correct.

Sample complexity upper bound

Theorem ([Adapt])

Using the GLR stopping with an asymptotically tight family of thresholds, EB-TCI satisfies that, for instances $\nu \in \mathcal{D}^K$ having distinct means,

$$\limsup_{\delta \to 0} \frac{\mathbb{E}_\nu [\tau_\delta]}{\log(1/\delta)} \leq T^\star_{1/2}(\nu) \, .$$

Asymptotically tight threshold, i.e. $c(\cdot, \delta) \sim_{\delta \to 0} \log(1/\delta)$.

☞ KL and BoB thresholds are asymptotically tight (not Student and Box).

Theorem ([Plug in])

For all asymptotically tight family of thresholds $(c_{a,b})_{a \neq b}$ and problem independent constant $\alpha > 0$, combining EB-EVTCI with the EV-GLR stopping rule using $(\alpha c_{a,b})_{a \neq b}$ yields an algorithm which is not δ-correct.

Empirical results ($\delta = 0.01$)

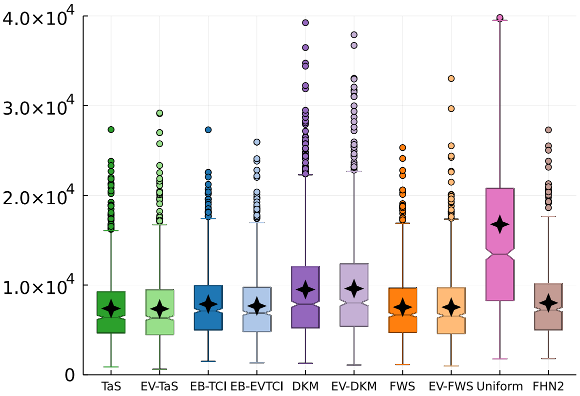

Figure: Empirical stopping time on random Gaussian instances ($K = 10$): $(\mu_1, \sigma_1^2) = (0, 1)$ and $-\mu_a \sim \mathcal{U}([0.2, 1.0])$ and $\sigma_a^2 \sim \mathcal{U}([0.1, 10])$ for all $a \neq 1$.

Conclusion

Two approaches to deal with unknown variances:

☞ **Plug in** the empirical variance,

☞ **Adapt** the transportation costs.

Two stopping rules, **GLR** and **EV-GLR**,

☞ calibrated with time-uniform concentration.

Two sampling rule **wrappers**, e.g. EB-TCI.

The impact of not knowning the variance is rather small !

References

Degenne, R., Koolen, W. M., and Ménard, P. (2019). Non-Asymptotic Pure Exploration by Solving Games. In *Advances in Neural Information Processing Systems*.

Garivier, A. and Kaufmann, E. (2016). Optimal best arm identification with fixed confidence. In *Conference on Learning Theory*.

Howard, S. R., Ramdas, A., McAuliffe, J., and Sekhon, J. (2020). Time-uniform chernoff bounds via nonnegative supermartingales. *Probability Surveys*, 17:257–317.

Jourdan, M., Degenne, R., Baudry, D., De Heide, R., and Kaufmann, E. (2022). Top two algorithms revisited. *Advances in Neural Information Processing Systems*.

Wang, P.-A., Tzeng, R.-C., and Proutiere, A. (2021). Fast pure exploration via frank-wolfe. *Advances in Neural Information Processing Systems*.

Questions ?

Appendix

Empirical variance: time-uniform concentration

Concentration on σ_{t+1}^2 after $t+1$ i.i.d. samples

With probability $1-\delta$,

$$\forall t \in \mathbb{N},\ \sigma_{t+1}^2/\sigma^2 \leq \overline{W}_{-1}(1+2g(t,\delta)/t) - 1/t \quad \text{with} \quad \overline{W}_{-1}(x) \approx x + \log x\ ,$$
$$\forall t \geq t_0(\delta),\ \sigma_{t+1}^2/\sigma^2 \geq \overline{W}_{0}(1+2g(t,\delta)/t) - 1/t \quad \text{with} \quad \overline{W}_{0}(x) \approx e^{-x+e^{-x}}\ ,$$

where $g(t,\delta) \approx \log(1/\delta) + \log\log t$ and $t_0(\delta) \approx 2\log(1/\delta)/\log\log(1/\delta)$.