Optimal Best Arm Identification with Fixed Confidence

after A. Garivier and E. Kaufmann (COLT 2016)

Remy Degenne

November 11, 2026

Part I

The paper

Chapter 1

Setting

An exponential family $(\nu_\theta)_\theta$ on $\mathbb{R}$, with densities $\exp(\theta x - b(\theta))$ with respect to a measure ξ, is parameterized by its means, and $d(\mu, \mu')$ is the Kullback–Leibler divergence between its members of means μ and μ'. A bandit model is a vector of means $\mu = (\mu_1, \dots, \mu_K)$ of arms in the family, and $\mathcal{S}$ is the set of those with a unique best arm $a^*(\mu)$. For $\mu \in \mathcal{S}$, $\mathrm{Alt}(\mu)$ is the set of models of $\mathcal{S}$ with another best arm, Σ_K the simplex, and the characteristic time is

$$T^*(\mu)^{-1} = \sup_{w \in \Sigma_K} \inf_{\lambda \in \mathrm{Alt}(\mu)} \sum_{a=1}^{K} w_a d(\mu_a, \lambda_a).$$

The maximizers $w^*(\mu)$ are the optimal allocations. In Lean, $\mathcal{S}$ is LMLExplore's best arm identification task restricted to the vectors of means of the family (ρ, id), the arm laws ν being its members. A strategy is an identification algorithm (sampling rule, stopping rule, recommendation); it is δ-PAC if on every $\mu \in \mathcal{S}$ it stops almost surely and recommends $a^*(\mu)$ with probability at least $1 - \delta$.

Chapter 2

The lower bound

Theorem 1 ([GK16, Theorem 1])**.** *Let $\delta \in (0, 1/2]$. For every δ-correct strategy and every $\mu \in \mathcal{S}$, $\mathbb{E}_\mu[\tau_\delta] \geq T^*(\mu)\mathrm{kl}(\delta, 1-\delta)$.*

The paper states it for $\delta \in (0, 1)$, which is false beyond $1/2$ (`notes/paper-errata.md`, E1).

Proof. A case of LMLExplore's lower bound for every identification task, Theorem 2. □

Lemma 2 ([KCG16, Lemma 1])**.** *For every identification task, every δ-correct strategy with $\delta \leq 1/2$ and every admissible model with a single correct answer, $\mathbb{E}[\tau] \geq T^*(\mu)\mathrm{kl}(\delta, 1-\delta)$.*

Proof. In LMLExplore. □

Corollary 3 ([GK16, Remark 2])**.** *For a family of δ-correct strategies and every $\mu \in \mathcal{S}$, $\liminf_{\delta \to 0} \mathbb{E}_\mu[\tau_\delta] / \log(1/\delta) \geq T^*(\mu)$: for every $c < T^*(\mu)$, $\mathbb{E}_\mu[\tau_\delta] \geq c \log(1/\delta)$ for all small enough δ.*

Proof. A case of Theorem 4. □

Lemma 4. *The asymptotic lower bound for every identification task.*

Proof. In LMLExplore. □

Chapter 3

The optimal allocation

Lemma 5 ([GK16, Lemma 3]). *For every $w \in \Sigma_K$, $\inf_{\lambda \in \mathrm{Alt}(\mu)} \sum_a w_a d(\mu_a, \lambda_a) = \min_{b \neq a^*}(w_{a^*} d(\mu_{a^*}, m_b) + w_b d(\mu_b, m_b))$, with $m_b = (w_{a^*}\mu_{a^*} + w_b \mu_b)/(w_{a^*} + w_b)$.*

Proof. In LMLExplore. □

Lemma 6 ([GK16, Lemma 4]). *At the optimal allocation, the costs of all the pairs (a^*, b) are equal.*

Proof. In LMLExplore. □

Theorem 7 ([GK16, Theorem 5]). *Let $K \geq 2$. For $\mu \in \mathcal{S}$ with best arm a, an allocation w is optimal if and only if $w = x/\sum_c x_c$ for weights x with $x_a = 1$ and, for some y and every $b \neq a$, $x_b > 0$ and $g_b(x_b) = d(\mu_a, m_b) + x_b d(\mu_b, m_b) = y$ with $m_b = (\mu_a + x_b\mu_b)/(1 + x_b)$, such that $F(y) = \sum_{b \neq a} d(\mu_a, m_b)/d(\mu_b, m_b) = 1$. The paper writes $x_b = g_b^{-1}(y)$ and computes the unique solution y^* of $F(y) = 1$ by bisection.*

The paper leaves $K \geq 2$ implicit. With a single arm, $F = 0$, while the allocation (1) is optimal (`notes/statement-corrections.md`).

Proof. The paper-level form of Part II, Theorem 21: for w optimal, $x = w/w_a$ and y the common cost of the pairs divided by w_a; conversely, $w = x/\sum_c x_c$, whose pairs all cost $y/\sum_c x_c$ by homogeneity, with the same weighted means as x. □

Proposition 8 ([GK16, Proposition 6]). *For $\mu \in \mathcal{S}$: (1) $w_a^*(\mu) > 0$ for every arm a; (2) w^* is continuous at μ; (3) if $\mu_{a^*} > \mu_2 \geq \ldots \geq \mu_K$, then $w_2^*(\mu) \geq \ldots \geq w_K^*(\mu)$.*

Proof. Parts (1) and (2) are in LMLExplore (`pos_of_mem_optAlloc_baiTask_restrict`, `continuousAt_of_mem_opt` part (3) is Theorem 23. The paper proves part (3) with the scalar equation of Theorem 7. The proof here uses only Theorem 6 and the monotonicity of the pair costs. □

Chapter 4

Track-and-Stop

Lemma 9 ([GK16, Lemma 7]). *C-tracking ensures* $N_a(t) \geq \sqrt{t + K^2} - 2K$ *and* $\max_a |N_a(t) - \sum_{s<t} w_a^*(\hat{\mu}(s))| \leq K(1 + \sqrt{t})$.

Proof. In LMLExplore. □

Lemma 10 ([GK16, Lemma 8]). *D-tracking ensures* $N_a(t) \geq (\sqrt{t} - K/2)_+ - 1$, *and the proportions of pulls converge to the limit of the targets whenever the targets converge.*

Proof. In LMLExplore. □

Proposition 11 ([GK16, Proposition 9]). *Under C-tracking and D-tracking,* $N_a(t)/t \to w_a^*(\mu)$ *almost surely.*

Proof. In LMLExplore, given the almost sure convergence of the targets $w^*(\hat{\mu}(t))$, which follows from that of the empirical means and the continuity of w^*. □

Theorem 12 ([GK16, Theorem 10]). *For Bernoulli arms, whatever the sampling rule, Chernoff's stopping rule with the threshold* $\beta(t, \delta) = \log(2t(K-1)/\delta)$ *ensures* $\mathbb{P}_\mu(\tau_\delta < \infty, \hat{a}_{\tau_\delta} \neq a^*) \leq \delta$ *for every* $\mu \in \mathcal{S}$.

Proof. Theorem 28, which holds for every $\delta > 0$ and all thresholds at least $\log(2t(K-1)/\delta)$. The paper's prior density in Appendix C.1 has a typo (`notes/paper-errata.md`, E3). □

Proposition 13 ([GK16, Proposition 12]). *For every* $\alpha > 1$ *and every* $\delta \in (0, 1)$ *there is* $C = C(\alpha, K, \delta)$ *such that, whatever the sampling rule, Chernoff's stopping rule with the threshold* $\beta(t, \delta) = \log(Ct^\alpha/\delta)$ *ensures* $\mathbb{P}_\mu(\tau_\delta < \infty, \hat{a}_{\tau_\delta} \neq a^*) \leq \delta$ *for every* $\mu \in \mathcal{S}$.

The paper claims a constant $C(\alpha, K)$ that does not depend on δ, which is false (`notes/paper-errata.md`, E2).

Proof. Theorem 33, for best arm identification on the vectors of means of the family, whose arms are dominated by the members of the family with their means, with $C = \delta e^{c_{\alpha,K} + 2x_\delta}$. It holds for every $\alpha > 0$. □

Chapter 5

Sample complexity

Proposition 14 ([GK16, Proposition 13]). *Let $\alpha \in [1, e/2]$, $r(t) = O(t^\alpha)$ and $\beta(t, \delta) = \log(r(t)/\delta)$. For every sampling rule with $N_a(t)/t \to w_a^*$ almost surely, $\tau_\delta < \infty$ almost surely for every $\delta \in (0, 1)$, and almost surely $\limsup_{\delta \to 0} \tau_\delta / \log(1/\delta) \leq \alpha T^*(\mu)$.*

Proof. Theorem 37 with $W = \{w^*(\mu)\}$. Every arm gets a positive proportion $w_a^*(\mu) > 0$ of the pulls (Theorem 8), so it is pulled infinitely often and its empirical mean converges almost surely to its mean (Theorem 38). The cost $D(\mu', w', \neg a^*)$ is continuous at $(\mu, w^*(\mu))$, the recommendation is a^* near μ (Theorem 34), and $D(\mu, w^*(\mu), \neg a^*) = T^*(\mu)^{-1} > 0$. The thresholds are $\log(r(t)/\delta) \leq \log(1/\delta) + \log r(t)$ with $\log r(t) = o(t)$. This gives $\limsup_{\delta \to 0} \tau_\delta / \log(1/\delta) \leq T^*(\mu) \leq \alpha T^*(\mu)$, for every $\alpha \geq 1$. □

Theorem 15 ([GK16, Theorem 14]). *Let $\alpha \in [1, e/2]$, $r(t) = O(t^\alpha)$ and $\beta(t, \delta) = \log(r(t)/\delta)$. Track-and-Stop with C-tracking or D-tracking has $\limsup_{\delta \to 0} \mathbb{E}_\mu[\tau_\delta] / \log(1/\delta) \leq \alpha T^*(\mu)$.*

Proof. LMLExplore proves the stronger bound with $T^*(\mu)$ for every polynomial r, Theorem 16, and $T^*(\mu) \leq \alpha T^*(\mu)$ since $\alpha \geq 1$. Near μ, the vectors of means have the same unique best arm (Theorem 34), which the recommendation returns, and the targets are their optimal allocations; the continuity of the targets that D-tracking assumes is Theorem 8 (2), and the continuity of the costs that both assume is Theorem 34. □

Lemma 16. *Track-and-Stop with D-tracking or C-tracking of optimal allocations has $\mathbb{E}_\mu[\tau_\delta] \leq c \log(1/\delta)$ for all small δ, for every $c > T^*(\mu)$.*

Proof. In LMLExplore, through the concentration of the empirical means under tracking [GK16, Lemma 19]. □

Part II

Prerequisites

Chapter 6

Prerequisites

The library-shaped results that the paper needs and that LMLExplore, LML, Tau Ceti and Mathlib do not provide yet. Their Lean code goes to the library's `ForLMLExplore/` folder, and `PAPER_ROADMAP.md` says where each goes in LMLExplore.

Lemma 17. *For an exponential family with divergence d, means μ_1, μ_2 of the family and nonnegative weights w_1, w_2, the cost $w_1 d(\mu_1, m) + w_2 d(\mu_2, m)$ of moving μ_1 and μ_2 to their weighted mean m is at most $w_1 d(\mu_1, z) + w_2 d(\mu_2, z)$ for every mean z of the family. So the cost of a pair is a minimum of functions linear in the weights, and equal to the one with $z = m$ at the weights (w_1, w_2).*

Proof. If $w_1 = w_2 = 0$, both sides vanish. Otherwise m is a mean of the family, and it minimizes $z \mapsto w_1 d(\mu_1, z) + w_2 d(\mu_2, z)$ by the compensation identity $w_1 d(\mu_1, z) + w_2 d(\mu_2, z) = w_1 d(\mu_1, m) + w_2 d(\mu_2, m) + (w_1 + w_2) d(m, z)$. □

Lemma 18. *For means μ, μ_1, μ_2 of the family, $(w_1, w_2) \mapsto d(\mu, (w_1\mu_1 + w_2\mu_2)/(w_1 + w_2))$ is continuous at every pair of nonnegative weights with $w_1 + w_2 > 0$.*

Proof. The weighted mean is continuous there and is a mean of the family, which is open, and d is continuous on pairs of means. □

Lemma 19. *For arms in an exponential family and a vector of means μ with unique best arm a, let $w \in \Sigma_K$ with $w_a > 0$ equalize the costs $C_b(w) = w_a d(\mu_a, m_b) + w_b d(\mu_b, m_b)$ of the pairs (a, b), $b \neq a$, where $m_b = (w_a\mu_a + w_b\mu_b)/(w_a + w_b)$, and satisfy $F = \sum_{b \neq a} d(\mu_a, m_b)/d(\mu_b, m_b) = 1$. Then w is optimal.*

Proof. Write $A_b = d(\mu_a, m_b)$ and $B_b = d(\mu_b, m_b)$, which is positive since $w_a > 0$ and $\mu_b < \mu_a$ give $m_b \neq \mu_b$, and D for the common value of the costs. For every $x \in \Sigma_K$, $\sum_{b \neq a} (x_a A_b + x_b B_b)/B_b = x_a F + \sum_{b \neq a} x_b = 1$. At $x = w$ this is $\sum_{b \neq a} D/B_b = 1$. Let $v \in \Sigma_K$. By Theorem 17 with $z = m_b$, the cost of the pair (a, b) at v is at most $v_a A_b + v_b B_b$. If all of them exceeded D, then $1 = \sum_{b \neq a} D/B_b < \sum_{b \neq a} (v_a A_b + v_b B_b)/B_b = 1$. So some pair costs at most D at v, and by Theorem 5 the cost of the alternatives at v is at most its value D at w. □

Lemma 20. *For arms in an exponential family, a vector of means μ with unique best arm a and at least one other arm, the optimal allocation satisfies $F = 1$, with the notation of Theorem 19.*

Proof. The optimal allocation w has positive entries (in LMLExplore). Suppose $F \neq 1$, and let $S = \sum_{b \neq a} 1/B_b > 0$. The direction u with $u_a = S/(F-1)$ and $u_b = (1 - A_b u_a)/B_b$ for $b \neq a$ has $\sum_c u_c = u_a + S - F u_a = 0$ and $u_a A_b + u_b B_b = 1$ for every $b \neq a$. Let $v = w + tu$, with m'_b the weighted mean of μ_a and μ_b at v. The cost of the pair (a, b) at v is $v_a d(\mu_a, m'_b) + v_b d(\mu_b, m'_b)$, and its cost at w is at most $w_a d(\mu_a, m'_b) + w_b d(\mu_b, m'_b)$ (Theorem 17), so the difference is at least $t\,(u_a d(\mu_a, m'_b) + u_b d(\mu_b, m'_b))$. As $t \to 0$, the second factor tends to $u_a A_b + u_b B_b = 1$ (Theorem 18). So for some small $t > 0$, v is in the simplex and every pair costs more at v than at w: by Theorem 5, the cost of the alternatives is larger at v, which contradicts the optimality of w. The proof needs neither a derivative of d nor the inverses of the costs of the pairs, which the paper uses. □

Theorem 21. *For arms in an exponential family, a vector of means μ with unique best arm a and at least one other arm, an allocation w is optimal if and only if $w \in \Sigma_K$, $w_a > 0$, the costs of the pairs (a, b), $b \neq a$, are equal, and $\sum_{b \neq a} d(\mu_a, m_b)/d(\mu_b, m_b) = 1$, with the notation of Theorem 19 [GK16, Theorem 5]. LMLRoadmaps: Identification 3.4.*

The paper writes the costs of the pairs as $w_a g_b(w_b/w_a)$, with the increasing function $g_b(x) = d(\mu_a, m_b(x)) + x\, d(\mu_b, m_b(x))$, and reduces the optimization to the scalar equation $F(y) = 1$ in the common value y of the g_b, through their inverses. The properties of g_b^{-1} and of F as a function of y (monotonicity, limits, the bisection) are not formalized: the characterization above does not need them.

Proof. The optimal allocation has positive entries (in LMLExplore), equalizes the costs of the pairs (Theorem 6) and satisfies $F = 1$ (Theorem 20). Conversely, Theorem 19. □

Lemma 22. *For an exponential family with divergence d, a mean μ_1 of the family and nonnegative weights w_1, w_2, the cost $w_1 d(\mu_1, m) + w_2 d(\mu_2, m)$ of moving μ_1 and μ_2 to their weighted mean $m = (w_1\mu_1 + w_2\mu_2)/(w_1 + w_2)$ is nonincreasing in the mean $\mu_2 \leq \mu_1$ of the family. It is decreasing when $w_1, w_2 > 0$.*

Proof. Let $\mu'_2 < \mu_2 \leq \mu_1$ and let m' be the weighted mean of μ_1 and μ'_2, so that $\mu'_2 < m' \leq \mu_1$. The pair cost of (μ_1, μ_2) is the minimum of $w_1 d(\mu_1, x) + w_2 d(\mu_2, y)$ over the means $x \leq y$. If $\mu_2 \leq m'$, take $x = y = m'$: then $d(\mu_2, m') < d(\mu'_2, m')$, since $d(\cdot, m')$ decreases below m'. If $m' < \mu_2$, take $x = m'$ and $y = \mu_2$: then $d(\mu_2, \mu_2) = 0 < d(\mu'_2, m')$. In both cases the result is at most $w_1 d(\mu_1, m') + w_2 d(\mu'_2, m')$, the pair cost of (μ_1, μ'_2), and strictly less when $w_2 > 0$. If a weight vanishes, both costs are zero. □

Theorem 23. *For arms in an exponential family and a vector of means with unique best arm a, the optimal allocation is ordered like the means on the arms other than a: for $b, c \neq a$, $w^*_c(\mu) \leq w^*_b(\mu)$ if and only if $\mu_c \leq \mu_b$, and $w^*_c(\mu) < w^*_b(\mu)$ if and only if $\mu_c < \mu_b$ [GK16, Proposition 6]. The same holds for the β-optimal allocations, the maximizers of $D(\mu, w, \neg a)$ among the allocations with $w_a = \beta$: for $\beta > 0$, and in the strict form for $\beta \in (0, 1)$. No LMLRoadmaps target covers it yet.*

Proof. Let w be β-optimal, so that the costs $C_b(w_b)$ of the pairs (a, b), with $C_b(x) = \beta d(\mu_a, m_b) + x d(\mu_b, m_b)$, are all equal (Theorem 6 and its β form). Let $b, c \neq a$ with $\mu_c \leq \mu_b$. If $w_b < w_c$, then C_b increases, so $C_b(w_b) < C_b(w_c)$. By Theorem 22, $C_b(w_c) \leq C_c(w_c)$. This contradicts $C_b(w_b) = C_c(w_c)$. If $\mu_c < \mu_b$ and $w_b \leq w_c$, then $C_b(w_b) \leq C_b(w_c) < C_c(w_c)$, using $w_c > 0$, a contradiction again. The optimal allocation is β-optimal for $\beta = w^*_a \in (0, 1)$; the two implications give the equivalences. □

Lemma 24. *For a sequence of $n = k + m \geq 1$ observations with k ones and m zeros, let $\mathrm{kt}(k,m) = \int_0^1 u^k(1-u)^m \frac{du}{\pi\sqrt{u(1-u)}} = B(\frac{1}{2}+k, \frac{1}{2}+m)/B(\frac{1}{2},\frac{1}{2})$ be its Krichevsky–Trofimov probability, the mixture of the Bernoulli likelihoods over the arcsine law $\mathrm{Beta}(\frac{1}{2},\frac{1}{2})$. Then $u^k(1-u)^m \leq (k/n)^k(m/n)^m \leq 2\sqrt{n}\,\mathrm{kt}(k,m)$ for every $u \in [0,1]$ [GK16, Lemma 11], and for $p \in (0,1)$, $n\,d(k/n,p) \leq \log(2\sqrt{n}\,\mathrm{kt}(k,m)/(p^k(1-p)^m))$. That* kt *is a probability on $\{0,1\}^n$ is not formalized: the proof of Theorem 26 does not need it.*

Proof. The moments of the beta laws give the closed form, and the recursion of the beta function $\mathrm{kt}(k,m+1) = \mathrm{kt}(k,m)\,(m+\frac{1}{2})/(n+1)$. By induction on m (from $\mathrm{kt}(0,1) = \frac{1}{2}$ and the symmetry in k and m), the claim $k^k m^m \leq 2\sqrt{n}\,n^n\,\mathrm{kt}(k,m)$ propagates from m to $m+1$ when $(m+1)^{m+1}\sqrt{n}\,n^n \leq \sqrt{n+1}\,(m+\frac{1}{2})\,m^m(n+1)^n$, which in logarithms is $(m+1)\log(m+1) - m\log m - \log(m+\frac{1}{2}) \leq 1 \leq (n+\frac{1}{2})\log(1+1/n)$. The right inequality is the first term of the series of $\log(1+1/n)$ in powers of $1/(2n+1)$; the left one, for $m \geq 1$, says that $t \mapsto (1+t)\log(1+t) - t - t\log(1+t/2)$, whose derivative is $\log\frac{1+t}{1+t/2} - \frac{t}{2+t} \leq 0$, is nonpositive at $t = 1/m$ [WST95, Appendix]. The maximum likelihood bound is $n\,d(k/n,u) \geq 0$, and $n\,d(k/n,p) = \log((k/n)^k(m/n)^m/(p^k(1-p)^m))$. □

Lemma 25. *Along a run of any algorithm against a stochastic bandit ν, if $\int e^{z(a,y)}\,d\nu_a(y) \leq 1$ for every arm a, then $n \mapsto \exp(\sum_{k<n} z(A_k,Y_k))$ is a supermartingale for the filtration of the history and the next action. If $z = z_\ell$ depends measurably on a parameter ℓ drawn from a probability measure ρ, then for every $c > 0$, $\mathbb{P}(\exists n,\ c \leq \int \exp(\sum_{k<n} z_\ell(A_k,Y_k))\,d\rho(\ell)) \leq 1/c$.*

Proof. Given the history and the action of round n, the observation Y_n has law ν_{A_n}, so the conditional expectation of $e^{z(A_n,Y_n)}$ is at most 1. Mixtures of nonnegative supermartingales are supermartingales (in LMLExplore), with initial value 1 here, and Ville's inequality (in LMLExplore) concludes. □

Lemma 26. *Along a run of any algorithm against a Bernoulli bandit with means $p \in (0,1)^K$, for two arms $a \neq b$ and every $x > 0$, $\mathbb{P}(\exists n,\ \log(2nx) < N_a(n)\,d(\hat{\mu}_a(n),p_a) + N_b(n)\,d(\hat{\mu}_b(n),p_b)) \leq 1/x$ [GK16, proof of Theorem 10].*

Proof. Take $z_{(u,v)}(c,y)$ the log-likelihood ratio of the observation y of the arm c under the bandit whose means on a and b are u and v against the true one, and $\rho = \mathrm{Beta}(\frac{1}{2},\frac{1}{2})^{\otimes 2}$. Its sum over the rounds is $S_a\log(u/p_a) + (N_a - S_a)\log((1-u)/(1-p_a))$ plus the same for b, so the mixture is the product $\Lambda_a\Lambda_b$ of the Krichevsky–Trofimov likelihood ratios $\Lambda_c = \mathrm{kt}(S_c, N_c - S_c)/(p_c^{S_c}(1-p_c)^{N_c-S_c})$, the observations being in $\{0,1\}$ almost surely. By Theorem 24, $N_c\,d(\hat{\mu}_c,p_c) \leq \log(w_c\Lambda_c)$ with $w_c = 2\sqrt{N_c}$, or $w_c = 1$ when $N_c = 0$. On the event, one of the two arms has an observation (the right side is positive), and then $w_a w_b \leq 2n$, so $\log(2nx) < \log(2n\Lambda_a\Lambda_b)$ and $x < \Lambda_a\Lambda_b$: Theorem 25 with $c = x$ concludes. The paper applies Lemma 11 also to an arm with no observation, where it fails (`notes/paper-errata.md`, E4). □

Lemma 27. *For best arm identification with Bernoulli arms on the vectors of means in $(0,1)$, the GLR statistic of the plug-in statistic of the empirical means for the arm i is at most $N_i\,d(\hat{\mu}_i,\lambda_i) + N_j\,d(\hat{\mu}_j,\lambda_j)$ for every arm $j \neq i$ and all $\lambda_i \leq \lambda_j$ in $(0,1)$.*

Proof. The cost of the alternatives to i is the infimum of the transportation cost from the empirical means over the vectors λ' of $(0,1)^K$ with $\lambda'_i \leq \lambda'_b$ for some $b \neq i$, since the cost is continuous there (in LMLExplore). Take λ' with $\lambda'_i = \lambda_i$, $\lambda'_j = \lambda_j$, and $\lambda'_c = (1-t)\hat{p}_c + t/2$ on the other arms, where $\hat{p}_c$ is $\hat{\mu}_c$ clamped to $[0,1]$: by the χ^2 bound, $d(\hat{p}_c,\lambda'_c) \leq t$, and $t \to 0$. □

Lemma 28. *For best arm identification with Bernoulli arms on the vectors of means in* $(0,1)$ *and every* $\delta > 0$*, whatever the sampling and recommendation rules, the GLRT with the plug-in statistic of the empirical means and thresholds at least* $\log(2n(K-1)/\delta)$ *stops with an arm that is not the best with probability at most* δ *[GK16, Theorem 10]. LMLRoadmaps: Identification 4.2.*

Proof. If the GLRT stops at round n with an arm $i \neq a^*$, then by Theorem 27 with $\lambda = \mu$ on the pair (i, a^*), $\log(2n(K-1)/\delta) < N_i\, d(\hat{\mu}_i, \mu_i) + N_{a^*}\, d(\hat{\mu}_{a^*}, \mu_{a^*})$. By Theorem 26 with $x = (K-1)/\delta$, each of the $K-1$ wrong arms contributes at most $\delta/(K-1)$. □

Lemma 29. *Let* $Z_0, Z_1, \ldots$ *be i.i.d. with the law of an integrable* Y*, with rate function* I*, and let* $\hat{\mu}_n$ *be the mean of* $Z_0, \ldots, Z_{n-1}$*. For* $a > 0$ *and* $r \geq 0$*,* $\mathbb{P}(\exists n \geq a,\ \hat{\mu}_n > \mathbb{E}[Y],\ I(\hat{\mu}_n) > r) \leq e^{-ar}$*. For* $a > 0$ *and* $z > 0$*,* $\mathbb{P}(\exists n,\ a \leq n < b,\ nI(\hat{\mu}_n) \geq z) \leq 2e^{-az/b}$*.*

Proof. Along the filtration of the past variables, Z_n is independent of the past, so its conditional moment generating function is that of Y, and the first bound is LMLExplore's time-uniform Chernoff bound with the rate function, with every variable selected. For the second, $nI(\hat{\mu}_n) \geq z$ with $n < b$ forces $I(\hat{\mu}_n) > z/b$, which vanishes at the mean; above the mean this is the first bound with $r = z/b$, and below the mean the first bound for $-Z$ and $-Y$. □

Lemma 30. *In the reward stream model, for an arm with integrable rewards and rate function* I_a*, every* $\eta > 0$ *and* $y \geq 0$*, with* $\kappa_\eta = \log(2/(1-e^{-\eta}))$*,* $\mathbb{P}(\exists s,\ sI_a(\hat{\mu}_s) \geq 2(\eta\lfloor \log_2 s \rfloor + \kappa_\eta + y)) \leq e^{-y}$*.*

Proof. A deviation with $s \geq 1$ rewards lies on the dyadic slice $2^k \leq s < 2^{k+1}$ with $k = \lfloor \log_2 s \rfloor$, and at $s = 0$ there is none. By Theorem 29 with $a = 2^k$ and $b = 2^{k+1}$, the slice k has probability at most $2e^{-(\eta k + \kappa_\eta + y)}$, and these sum to $2e^{-\kappa_\eta - y}/(1 - e^{-\eta}) = e^{-y}$. □

Lemma 31. *Along a run of any algorithm, if the rewards of each arm are integrable with rate function* I_a*, then for* $\eta > 0$ *and* $x \geq K$*,* $\mathbb{P}(\exists t,\ \sum_a N_a(t) I_a(\hat{\mu}_a(t)) \geq 2(K(\eta\lfloor \log_2 t \rfloor + \kappa_\eta) + x)) \leq (x/K)^K e^{K-x}$*. For arms dominated by members of exponential families, the same holds with the extended divergences* $\bar{d}_a(\hat{\mu}_a(t), \mu_a)$ *in place of* $I_a(\hat{\mu}_a(t))$*.*

Proof. In the stream model the arms are independent, and Theorem 30 gives each half statistic $sI_a(\hat{\mu}_s)/2$ the exponential tail above the monotone function $\eta\lfloor \log_2 s \rfloor + \kappa_\eta$ that [LS20, Proposition 33.9] (in LMLExplore) combines over the arms. LMLExplore's joint event transfer turns the stream bound into a bound along the run, where $\sum_a N_a(t) = t$. The rate function of a dominated arm is at least the extended divergence of its family. □

Lemma 32. *For every identification task on vectors of means, arms dominated by exponential families, every sampling rule,* $p > 0$ *and* $x \geq K$*, the GLRT with the generalized likelihood ratio evidence and thresholds at least* $p \log n + c_{p,K} + 2x$ *at every round* $n \geq 1$*, with* $c_{p,K} = 2K \log(2/(1 - e^{-p \log 2/(2K)}))$*, stops with an incorrect answer with probability at most* $(x/K)^K e^{K-x}$*.*

Proof. The GLRT errs only if the evidence against the true model exceeds the threshold at some round, which does not happen at round 0, where the evidence vanishes. With $\eta = p \log 2/(2K)$, $2K\eta\lfloor \log_2 n \rfloor \leq p \log n$, so the thresholds dominate those of Theorem 31. □

Lemma 33. *For every identification task on vectors of means, arms dominated by exponential families, every sampling rule,* $p > 0$*,* $\delta \in (0, 1]$ *and* $C \geq \delta e^{c_{p,K} + 2x_\delta}$*, where* $x_\delta = \log(1/\delta) + 2K(1 + \log(1 + \log(1/\delta)))$ *is LMLExplore's level, the GLRT with the threshold* $\log(Ct^p/\delta)$ *is* δ*-correct [GK16, Proposition 12], and it is* δ*-PAC when it stops almost surely. LMLRoadmaps: Identification 4.2.*

Proof. For $n \geq 1$, $\log(Cn^p/\delta) \geq p \log n + c_{p,K} + 2x_\delta$, and $(x_\delta/K)^K e^{K-x_\delta} \leq \delta$ with $x_\delta \geq K$ (LMLExplore). □

Lemma 34. *For best arm identification restricted to an open set of vectors of means, such as the vectors of means of a nondegenerate exponential family, the models with unique best arm a form an open set. For arms in a nondegenerate exponential family and best arm identification restricted to the vectors of means of the family, at every model μ with unique best arm a, $(\mu', w) \mapsto D(\mu', w, \neg a)$ is continuous within $\mathbb{R}^K \times \Sigma_K$ at (μ, w) for every $w \in \Sigma_K$. These are the continuity hypotheses of LMLExplore's sample complexity bounds for Track-and-Stop. LMLExplore proves the continuity as a private lemma; LMLExplore pull request #242 makes it public and adds the openness.*

Proof. The models with unique best arm a are an intersection of finitely many strict inequalities between means, intersected with the open set of the restriction. The closed form of Theorem 5 is continuous on the vectors of means of the family with best arm a and the nonnegative allocations, a neighbourhood of (μ, w) within $\mathbb{R}^K \times \Sigma_K$, since the set of means of the family is open. □

Lemma 35. *Along a sequence of histories on which the GLR statistic of the recommended answer after t rounds is at least κt for all large t, consider the GLRTs with thresholds $\beta_t(h, \delta) \leq \log(1/\delta) + f(t)$, $f(t) = o(t)$. If $\kappa > 0$, the GLRT stops at every δ at which the thresholds have this form, and $\tau_\delta \leq c \log(1/\delta)$ for all small δ, for every $c > 1/\kappa$.*

Proof. Take $1/c < \eta < \kappa$ and $L = \log(1/\delta)$. From the round $N = \max(M, \lceil L/\eta \rceil)$ on, for a fixed M, the threshold is at most $L + (\kappa - \eta)N/2 < \kappa N$, so that $\tau_\delta \leq N \leq M + L/\eta + 1$, which is at most cL when L is large. □

Lemma 36. *Let W be a compact set of allocations such that $(\theta', w) \mapsto D(\theta', w, \neg i)$ is lower semicontinuous at (θ, w) on the simplex for every $w \in W$, and let the recommendation be i near θ. Along a sequence of histories on which the estimates converge to θ and the proportions of pulls $N(t)/t$ approach W, the GLR statistic of the plug-in statistic, $tD(\hat{\theta}_t, N(t)/t, \neg i)$, is at least κt for all large t, for every $\kappa < \inf_{w \in W} D(\theta, w, \neg i)$.*

Proof. By compactness of W and lower semicontinuity, $D(\theta', w', \neg i) > \kappa$ for θ' near θ and w' in the simplex near W, which $\hat{\theta}_t$ and $N(t)/t$ are for all large t. □

Lemma 37. *For the GLRT with thresholds $\beta_t(h, \delta) \leq \log(1/\delta) + f(t)$, $f(t) = o(t)$, if almost surely the estimates converge to the model and the proportions of pulls to a set W of allocations, then almost surely $\tau_\delta \leq c \log(1/\delta)$ for all small δ, for every $c > (\inf_{w \in W} D(\mu, w, \neg a^*))^{-1}$ [GK16, Proposition 13]. If the infimum is positive, the GLRT stops almost surely at every δ at which the thresholds have this form. LMLRoadmaps: Identification 4.4.*

Proof. On the almost sure event of the convergences, Theorem 36 gives the linear growth of the statistic, and Theorem 35 the bounds on the stopping time. □

Lemma 38. *In a run of any algorithm against a stochastic bandit with integrable rewards, almost surely the empirical mean of every arm pulled infinitely often converges to the mean of the arm.*

Proof. The rewards of the successive pulls of an arm are i.i.d. with the law of the arm (in LML), and after $N_a(t)$ pulls the empirical mean is the mean of the first $N_a(t)$ of them: the strong law of large numbers applies. □

Bibliography

[GK16] Aurélien Garivier and Emilie Kaufmann, *Optimal best arm identification with fixed confidence*, Proceedings of the 29th Conference on Learning Theory (COLT), 2016, arXiv:1602.04589v2.

[KCG16] Emilie Kaufmann, Olivier Cappé, and Aurélien Garivier, *On the complexity of best-arm identification in multi-armed bandit models*, Journal of Machine Learning Research **17** (2016), 1–42.

[LS20] Tor Lattimore and Csaba Szepesvári, *Bandit algorithms*, Cambridge University Press, 2020.

[WST95] Frans M. J. Willems, Yuri M. Shtarkov, and Tjalling J. Tjalkens, *The context-tree weighting method: Basic properties*, IEEE Transactions on Information Theory **41** (1995), no. 3, 653–664.