Supplement to: "Risk and optimal policies in bandit experiments"
Karun Adusumilli (University of Pennsylvania)

Appendix B. Rates of convergence to the PDE solution

The results of Barles and Jakobsen (2007, Theorem 3.1) provide a bound on the rate of convergence of $V_n^*(\cdot)$ to $V^*(\cdot)$. The technical requirements to obtain this are described in their Assumptions A2 and S1-S3. Assumptions A2 and S1-S2 are straightforward to verify using the regularity conditions given for Theorem 2 with the additional requirement $\sup_s |\mu^+(s)| < \infty$.

Assumption S3 of Barles and Jakobsen (2007) is a strengthening of the consistency requirement in (A.3) and (A.4). Suppose that the test function $\phi \in \mathcal{C}^\infty(\mathcal{S})$ is such that $\left|\partial_t^{\beta_0} D_{(x,q)}^{\beta}\phi(x,q,t)\right| \le K\varepsilon^{1-2\beta_0-\|\beta\|}$ for all $\beta_0 \in \mathbb{N}, \beta \in \mathbb{N}\times\mathbb{N}$. Then by a third order Taylor expansion as in the proof of Theorem 2 and some tedious but straightforward algebra,

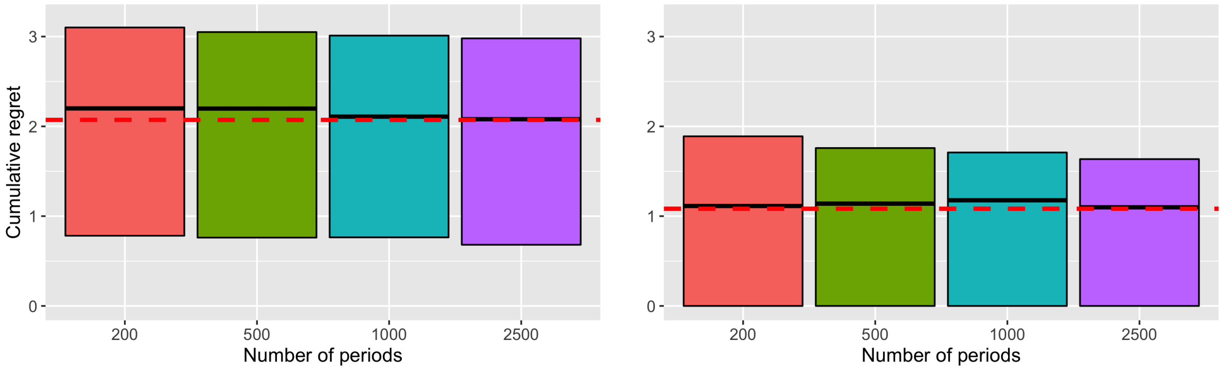
$$\left|nS_n(z,\phi(z)+\rho,[\phi+\rho]) - F(D^2\phi(s), D\phi(s), s)\right| \le E(n,\varepsilon) \equiv \frac{\bar{K}}{n^{1/2}\varepsilon^2},$$

where $\bar{K}$ depends only on K, defined above, and the upper bounds on $\mu^+(\cdot), \mu(\cdot)$. The above suffices to verify the Assumption S3 of Barles and Jakobsen (2007); note that the definition of $S(\cdot)$ in that paper is equivalent to $nS_n(\cdot)$ here.

Under the above conditions, Barles and Jakobsen (2007, Theorem 3.1) implies

$$\begin{aligned} V^* - V_n^* &\lesssim \sup_{\varepsilon}\left(\varepsilon + E(n,\varepsilon)\right) \lesssim n^{-1/6} \text{ and} \\ V_n^* - V^* &\lesssim \sup_{\varepsilon}\left(\varepsilon^{1/3} + E(n,\varepsilon)\right) \lesssim n^{-1/14}. \end{aligned} \tag{B.1}$$

The asymmetry of the rates is an artifact of the techniques of Barles and Jakobsen (2007). The rates are also far from optimal. The results of Barles and Jakobsen (2007), while being relatively easy to apply, do not exploit any regularity properties of the approximation scheme. There do exist approximation schemes for PDE (2.8) that converge at the faster $n^{-1/2}$ rates. While it is unknown whether (3.1) is one of them, we do find that in practice the quality of approximation of V^* with V_n^* is far better than what (B.1) appears to suggest; the Monte-Carlo simulation in Figure B.1 attests to this (the simulation employs a normal prior $\mu \sim \mathcal{N}(0, 50^2)$ with $\sigma = 5$).



A: Thompson sampling B: Optimal Bayes policy

Note: The parameter values are $\mu_0 = 0$, $\nu = 50$ and $\sigma = 5$. The dashed red lines denote the values of asymptotic Bayes risk. Black lines within the bars denote the Bayes risk in finite samples. The bars describe the interquartile range of regret.

FIGURE B.1. Monte-Carlo simulations

APPENDIX C. LOWER BOUNDS ON MINIMAX RISK

Recall the definition of $V_{n,\pi}(0;h)$ from Section 5.4 as the frequentist risk under some $\pi \in \Pi$. We also make the dependence of $V_n^*(0), V^*(0)$ on the priors m_0 explicit by writing them as $V_n^*(0;m_0), V^*(0;m_0)$. Clearly, $\inf_{\pi\in\Pi}\sup_{|h|\le\Gamma} V_{n,\pi}(0;h) \ge V_n^*(0;m_0)$ for any prior m_0 supported on $|h| \le \Gamma$. So, Theorem 5 implies

$$\lim_{n\to\infty}\inf_{\pi\in\Pi}\sup_{|h|\le\Gamma} V_{n,\pi}(0;h) \ge \sup_{m_0\in\mathcal{P}} V^*(0;m_0)$$

where $\mathcal{P}$ is the set of all compactly supported distributions. We now claim that

$$\sup_{m_0\in\mathcal{P}} V^*(0;m_0) = \bar{V}^*, \tag{C.1}$$

where $\bar{V}^*$ is the asymptotic minimax risk in the Gaussian setting. The above is easily shown for scalar θ by transforming the state variable x to $\dot{\mu}_0 x$ and replacing σ^2 with $\dot{\mu}_0^2\sigma^2$, following which the infinitesimal generator (5.6) becomes equivalent to the one in (2.8) since $\mu(s) = \dot{\mu}_0 h(s)$. The argument for vector θ is given below.

C.0.1. *Proof of (C.1) for vector θ.* We employ the same notation as in Section 5.3. It is without loss of generality to suppose $\Sigma = I$, otherwise, we can perform the subsequent analysis after applying the transformations $h \leftarrow \Sigma^{-1/2}h, x \leftarrow \Sigma^{-1/2}x$ and $\dot{\mu}_0 \leftarrow \Sigma^{1/2}\dot{\mu}_0$. Consider the class, $\bar{\mathcal{P}}$, of priors, m_0, over h supported on $\mu \cdot \dot{\mu}_0/(\dot{\mu}_0^\intercal\dot{\mu}_0)$, where $\mu \in \mathbb{R}$ can take on various values (so m_0 is, in essence, a prior on μ). For these priors, $\dot{\mu}_0^\intercal h = \mu$. Recall that under the approximate posterior, $\tilde{p}_n(h|x,q) \propto \mathcal{N}(x|qh, q\Sigma) \cdot m_0(h)$. It is then easily verified that, for the class $\bar{\mathcal{P}}$, $\tilde{p}_n(h|x,q)$ depends on x only through $\dot{\mu}_0^\intercal x$. Furthermore, we also have $h(s) = \mu(s) \cdot \dot{\mu}_0/(\dot{\mu}_0^\intercal \mu_0)$, where $\mu(s), h(s)$ are the posterior means of μ, h under $\tilde{p}_n(\cdot|x,q)$.

Choose $\{\phi_i\}_{i=1}^{d-1}$ such that $\{\dot{\mu}_0/\dot{\mu}_0^\intercal \mu_0, \phi_1, \dots, \phi_{d-1}\}$ are orthonormal and span $\mathbb{R}^d$. Suppose we transform the state variables x to z as $z = Px$, where $P^\intercal = [\dot{\mu}_0, \phi_1, \dots, \phi_{d-1}]$. Clearly, P is invertible, and the first component of z is $\bar{x} := \dot{\mu}_0^\intercal x$. Consider the generator $L[\cdot]$ in (5.7). Following the transformation of variables,

$$h(s)^\intercal D_x f = \frac{\mu(s)}{\dot{\mu}_0^\intercal \dot{\mu}_0} \dot{\mu}_0^\intercal \cdot P^\intercal D_z f = \mu(s) \cdot \left[1, \mathbf{0}_{1\times(d-1)}\right] \cdot D_z f = \mu(s) \partial_{\bar{x}} f,$$

and $\mathrm{Tr}\,[D_x^2 f] = \mathrm{Tr}\,[PP^\intercal \cdot D_z^2 f]$. Clearly, $PP^\intercal$ is block diagonal, with diagonal entries $\dot{\mu}_0^\intercal \dot{\mu}_0$ and $I_{(d-1)}$. Hence, we can write $\mathrm{Tr}\,[D_x^2 f] = (\dot{\mu}_0^\intercal \dot{\mu}_0) \cdot \partial_{\bar{x}}^2 f + \mathrm{Tr}\,[D_{\tilde{x}}^2 f]$ where $\tilde{x}$ is the part of z excluding the first component. Combining the above, and defining $\sigma^2 := \dot{\mu}_0^\intercal \dot{\mu}_0$ (more generally, for $\Sigma \neq I$, this would be $\dot{\mu}_0^\intercal \Sigma \dot{\mu}_0$), we have thus shown $L[f](s) = \partial_q f + \mu(s) \partial_{\bar{x}} f + \frac{1}{2}\sigma^2 \partial_{\bar{x}}^2 f + \frac{1}{2}\mathrm{Tr}\,[D_{\tilde{x}}^2 f]$.

The minimal Bayes risk, $V^*(s; m_0)$, solves the PDE:

$$\partial_t f(s) + \mu^+(s) + \min\left\{-\mu(s) + L[f](s), 0\right\} = 0 \text{ if } t < 1; \quad f(s) = 0 \text{ if } t = 1.$$

Now, $\tilde{p}_n(h|x,q)$ depends on x only though $\bar{x}$, so $\mu(s) \equiv \tilde{\mathbb{E}}[\mu|s], \mu^+(s) \equiv \tilde{\mathbb{E}}[\mu \mathbb{I}\{\mu \geq 0\}|s]$ are functions only of $\bar{x}, q$. Hence, by similar viscosity solution arguments as in the proof of Theorem 6 (Appendix F), it follows that $V^*(s; m_0)$ solves

$$\partial_t f(\bar{s}) + \mu^+(\bar{s}) + \min\left\{-\mu(\bar{s}) + \bar{L}[f](\bar{s}), 0\right\} = 0 \text{ if } t < 1; \quad f(\bar{s}) = 0 \text{ if } t = 1,$$

where $\bar{s} := (\bar{x}, q, t)$ and $\bar{L}[f](\bar{s}) = \partial_q f + \mu(\bar{s}) \partial_{\bar{x}} f + \frac{1}{2}\sigma^2 \partial_{\bar{x}}^2 f$. But the above has the same form as PDE (2.8) in the Gaussian setting if we interpret m_0 as a prior on μ. Hence, $\sup_{m_0 \in \bar{\mathcal{P}}} V^*(0; m_0) = \bar{V}^*$, the minimax risk in the Gaussian regime.

Since $\bar{\mathcal{P}} \subset \mathcal{P}$, the set of all compactly supported priors on $\boldsymbol{h}$, we have thereby derived a lower bound on minimax risk. As an aside, we note that our proof also goes through after replacing $\bar{\mathcal{P}}$ with the class of product priors defined in Section 6; the argument would then be similar to the proof of Theorem 6, see Appendix F.

Appendix D. Proof of Theorem 5

Recall that $\mathbf{y}_i = \{Y_k\}_{k=1}^i$ denotes the rewards after i pulls of the arms. Denote by $\mathbb{E}_{(\mathbf{y}_n,h)}[\cdot]$ the expectation under the 'true' joint density $dS_n(\mathbf{y}_n, h) := \left\{\prod_{i=1}^n p_{\theta_0+h/\sqrt{n}}(Y_i)\right\} \cdot m_0(h)$. Let $\nu(\mathbf{y}_n) := \prod_{i=1}^n \nu(Y_i)$, $p_{n,\theta}(\mathbf{y}_n) := \prod_{k=1}^n p_\theta(Y_k)$ and $\bar{P}_n$ be the probability measure corresponding to the 'true' marginal density $d\bar{P}_n(\mathbf{y}_n) := \int p_{n,\theta_0+h/\sqrt{n}}(\mathbf{y}_n) \cdot m_0(h)d\nu(h)$. We use $\bar{\mathbb{E}}_n[\cdot]$ to denote its corresponding expectation. As first defined in Appendix A.3, let $\tilde{\bar{P}}_n$ denote the measure (but not necessarily a probability) corresponding to the density $d\tilde{\bar{P}}_n(\mathbf{y}_n) := \int d\Lambda_{n,h}(\mathbf{y}_n) \cdot m_0(h)d\nu(h)$. In what follows, we denote $d\Lambda_{n,h}(\mathbf{y}_n)$ by $\lambda_{n,h}(\mathbf{y}_n)$ for ease of notation, and note that

$$\lambda_{n,h}(\mathbf{y}_n) := d\Lambda_{n,h}(\mathbf{y}_n) \equiv \frac{d\Lambda_{n,h}(\mathbf{y}_n)}{d\nu(\mathbf{y}_n)} = \exp\left\{\frac{1}{\sigma^2}hx_n - \frac{1}{2\sigma^2}h^2\right\} p_{n,\theta_0}(\mathbf{y}_n).$$

Finally, $\|\cdot\|_{\text{TV}}$ denotes the total variation metric between two measures.

The proof follows the basic outline established in Appendix A.3. Recall the notation used there, as well as the expressions for $V_{\pi,n}(0), \tilde{V}_{\pi,n}(0)$ given in (A.7) and (A.8).

Step 1 (Approximation of $V_{\pi,n}(0)$ with $\tilde{V}_{\pi,n}(0)$): We start by proving some convergence properties of $\tilde{\bar{P}}_n$ and $\tilde{p}_n(\cdot|\mathbf{y}_{nq})$ to $\bar{P}_n$ and $p_n(\cdot|\mathbf{y}_{nq})$. The proofs here make heavy use of the SLAN property (5.2) established in Lemma 2. Let A_n denote the event $\left\{\mathbf{y}_n : \sup_q |x_{nq}| \le M\right\}$. For any measure P, define $P \cap A_n$ as the restriction of P to the set A_n. By Lemma 6 in Appendix E, for any $\epsilon > 0$ there exists $M < \infty$ such that

$$\lim_{n\to\infty} \bar{P}_n\left(A_n^c\right) \le \epsilon, \tag{D.1}$$

$$\lim_{n\to\infty} \left\|\bar{P}_n \cap A_n - \tilde{\bar{P}}_n \cap A_n\right\|_{\text{TV}} = 0, \text{ and} \tag{D.2}$$

$$\lim_{n\to\infty} \sup_q \bar{\mathbb{E}}_n\left[\mathbb{I}_{A_n} \|p_n(\cdot|\mathbf{y}_{nq}) - \tilde{p}_n(\cdot|\mathbf{y}_{nq})\|_{\text{TV}}\right] = 0. \tag{D.3}$$

The measures $\Lambda_{n,h}(\cdot), \tilde{\bar{P}}_n(\cdot)$ are not probabilities as they need not integrate to 1. But Lemma 6 also shows the following: $\Lambda_{n,h}(\cdot), \tilde{\bar{P}}_n(\cdot)$ are σ-finite and contiguous with respect to P_{n,θ_0}, and letting $\mathcal{Y}_n$ denote the sample space of $\mathbf{y}_n$,

$$\lim_{n\to\infty} \tilde{\bar{P}}_n(\mathcal{Y}_n) = 1 \quad \text{and} \quad \lim_{n\to\infty} \tilde{\bar{P}}_n(A_n^c) \le \epsilon. \tag{D.4}$$

The first result in (D.4) implies that $\tilde{\bar{P}}_n$ is almost a probability measure.

Based on the above, we show that

$$\lim_{n\to\infty}\sup_{\pi\in\Pi}\left|V_{\pi,n}(0)-\tilde{V}_{\pi,n}(0)\right|=0 \tag{D.5}$$

by bounding each term in the following expansion:

$$\begin{aligned}
&V_{\pi,n}(0)-\tilde{V}_{\pi,n}(0)\\
&=\bar{\mathbb{E}}_n\left[\mathbb{I}_{A_n^c}\frac{1}{n}\sum_{j=0}^{n-1}\mathbb{E}\left[R_n(h,\pi_{j+1})|\,\mathbf{y}_{nq_j(\pi)}\right]\right]+\tilde{\mathbb{E}}_n\left[\mathbb{I}_{A_n^c}\frac{1}{n}\sum_{j=0}^{n-1}\tilde{\mathbb{E}}\left[R_n(h,\pi_{j+1})|\,\mathbf{y}_{nq_j(\pi)}\right]\right]\\
&\quad+\left(\bar{\mathbb{E}}_n-\tilde{\mathbb{E}}_n\right)\left[\mathbb{I}_{A_n}\frac{1}{n}\sum_{j=0}^{n-1}\tilde{\mathbb{E}}\left[R_n(h,\pi_{j+1})|\,\mathbf{y}_{nq_j(\pi)}\right]\right]\\
&\quad+\bar{\mathbb{E}}_n\left[\mathbb{I}_{A_n}\frac{1}{n}\sum_{j=0}^{n-1}\left\{\mathbb{E}\left[R_n(h,\pi_{j+1})|\,\mathbf{y}_{nq_j(\pi)}\right]-\tilde{\mathbb{E}}\left[R_n(h,\pi_{j+1})|\,\mathbf{y}_{nq_j(\pi)}\right]\right\}\right].
\end{aligned} \tag{D.6}$$

Because of the compact support of the prior, the posteriors $p_n(\cdot|\mathbf{y}_{nq}),\tilde{p}_n(\cdot|\mathbf{y}_{nq})$ are also compactly supported on $|h|\le\Gamma$ for all q. On this set $|R_n(h,\pi_j)|\le b\Gamma$ for some $b<\infty$ by Assumption 1(iii). The first two quantities in (D.6) are therefore bounded by $b\Gamma\bar{P}_n(A_n^c)$ and $b\Gamma\tilde{\bar{P}}_n(A_n^c)$. By (D.1) and (D.4), these can be made arbitrarily small by choosing a suitably large M in the definition of A_n. The third term in (D.6) is bounded by $b\Gamma\left\|\bar{P}_n\cap A_n-\tilde{\bar{P}}_n\cap A_n\right\|_{\mathrm{TV}}$. By (D.2) it converges to 0 as $n\to\infty$. The expression within {} brackets in the fourth term of (D.6) is smaller than $b\Gamma\left\|p_n(\cdot|\mathbf{y}_{nq_j(\pi)})-\tilde{p}_n(\cdot|\mathbf{y}_{nq_j(\pi)})\right\|_{\mathrm{TV}}$. Hence, by the linearity of expectations, the term overall is bounded (uniformly over $\pi\in\Pi$) by

$$b\Gamma\sup_q\bar{\mathbb{E}}_n\left[\mathbb{I}_{A_n}\left\|p_n(\cdot|\mathbf{y}_{nq})-\tilde{p}_n(\cdot|\mathbf{y}_{nq})\right\|_{\mathrm{TV}}\right],$$

which is $o(1)$ because of (D.3). We have thus shown (D.5).

Step 2 (Approximating $V_n^*(0)$ with a recursive formula): The measure, $\tilde{\bar{P}}_n$, used in the outer expectation in the definition of $\tilde{V}_{\pi,n}(0)$ is not a probability. This can be rectified as follows: First, note that the density $\lambda_{n,h}(\cdot)$ can be written as

$$\lambda_{n,h}(\mathbf{y}_n)=\prod_{i=1}^{n}\left\{\exp\left\{\frac{h}{\sqrt{n}}\psi(Y_i)-\frac{h^2}{2\sigma^2 n}\right\}p_{\theta_0}(Y_i)\right\}=\prod_{i=1}^{n}\tilde{p}_n(Y_i|h), \tag{D.7}$$

where[19]

$$\tilde{p}_n(Y_i|h) := \exp\left\{\frac{h}{\sqrt{n}}\psi(Y_i) - \frac{h^2}{2\sigma^2 n}\right\} p_{\theta_0}(Y_i).$$

Using (D.7), Lemma 7 shows that $\tilde{\tilde{P}}_n$ can be disintegrated as

$$d\tilde{\tilde{P}}_n(\mathbf{y}_n) = \prod_{i=1}^{n}\left\{\int \tilde{p}_n(Y_i|h)\tilde{p}_n(h|\mathbf{y}_{i-1})d\nu(h)\right\}, \tag{D.8}$$

with $\tilde{p}_n(h|\mathbf{y}_0) := m_0(h)$. Now define $c_{n,i} := \int \left\{\int \tilde{p}_n(Y_i|h)d\nu(Y_i)\right\} \tilde{p}_n(h|\mathbf{y}_{i-1})d\nu(h)$, and let $\tilde{\mathbb{P}}_n$ denote the probability measure

$$\tilde{\mathbb{P}}_n(\mathbf{y}_n) = \prod_{i=1}^{n}\tilde{\mathbb{P}}_n(Y_i|\mathbf{y}_{i-1}), \text{ where}$$
$$d\tilde{\mathbb{P}}_n(Y_i|\mathbf{y}_{i-1}) := \frac{1}{c_{n,i}}\int \tilde{p}_n(Y_i|h)\tilde{p}_n(h|\mathbf{y}_{i-1})d\nu(h). \tag{D.9}$$

Note that $c_{n,i}$ is a random (because it depends on $\mathbf{y}_{i-1}$) integration factor ensuring $\tilde{\mathbb{P}}_n(y_{i+1}|\mathbf{y}_i)$, and therefore $\tilde{\mathbb{P}}_n$, is a probability. In Lemma 8, it is shown that there exists some non-random $C < \infty$ such that

$$\sup_i |c_{n,i} - 1| \le Cn^{-c} \text{ for any } c < 3/2, \tag{D.10}$$

and furthermore, $\left\|\tilde{\mathbb{P}}_n - \tilde{\tilde{P}}_n\right\|_{\mathrm{TV}} \to 0$ as $n \to \infty$. Hence, letting

$$\breve{V}_{\pi,n}(0) := \mathbb{E}_{\tilde{\mathbb{P}}_n}\left[\frac{1}{n}\sum_{j=0}^{n-1}\tilde{\mathbb{E}}\left[R_n(h,\pi_{j+1})|\,\mathbf{y}_{nq_j(\pi)}\right]\right],$$

where $\mathbb{E}_{\tilde{\mathbb{P}}_n}[\cdot]$ is the expectation with respect to $\tilde{\mathbb{P}}_n$, one obtains the approximation

$$\sup_{\pi\in\Pi}\left|\tilde{V}_{\pi,n}(0) - \breve{V}_{\pi,n}(0)\right| \le b\Gamma\left\|\tilde{\mathbb{P}}_n - \tilde{\tilde{P}}_n\right\|_{\mathrm{TV}} \to 0. \tag{D.11}$$

See the arguments following (D.6) for the definition of b.

Since $\tilde{p}_n(h|\mathbf{y}_{i-1}) \equiv \tilde{p}_n(h|x = x_{i-1}, q = (i-1)/n)$ by (5.5) with $\tilde{p}_n(h|x = 0, q = 0) := m_0(h)$, it follows from (D.9) that $\tilde{\mathbb{P}}_n(Y_i|\mathbf{y}_{i-1}) \equiv \tilde{\mathbb{P}}_n(Y_i|x = x_{i-1}, q = (i-1)/n)$.

Define $\breve{V}_n^*(0) = \inf_{\pi\in\Pi}\breve{V}_{\pi,n}(0)$. Recall that for a given $\pi \in \{0,1\}$, $\tilde{\mathbb{E}}\left[R_n(h,\pi)|\,\mathbf{y}_{nq_j}\right] \equiv \tilde{\mathbb{E}}\left[R_n(h,\pi)|\,x_{nq_j}, q_j\right]$ by (5.5). Furthermore, we have noted above that the conditional distribution of the future values of the rewards, $\tilde{\mathbb{P}}_n(Y_{nq_j+1}|\mathbf{y}_{nq_j})$, also depends only on (x_{nq_j}, q_j). Based on this, standard backward induction/dynamic programming arguments imply $\breve{V}_n^*(0)$ can be obtained as the solution at $(x,q,t) = (0,0,0)$

[19] Despite the notation, $\tilde{p}_n(Y_i|h)$ is not a probability density.

of the recursive problem

$$\breve{V}_n^*(x,q,t) = \min_{\pi\in\{0,1\}} \left\{ \frac{\tilde{\mathbb{E}}\left[R_n(h,\pi)\middle| x,q\right]}{n} + \mathbb{E}_{\tilde{\mathbb{P}}_n}\left[\mathbb{I}_n \cdot \breve{V}_n^*\left(x + \frac{\pi\sigma^2\psi(Y_{nq+1})}{\sqrt{n}}, q+\frac{\pi}{n}, t+\frac{1}{n}\right)\middle| s\right]\right\};$$

$$\text{if } t<1,$$

$$\breve{V}_n^*(x,q,1) = 0, \tag{D.12}$$

where $\mathbb{E}_{\tilde{\mathbb{P}}_n}[\cdot|\, s]$ denotes the expectation under $\tilde{\mathbb{P}}_n(Y_{nq+1}|\mathbf{y}_{nq}) \equiv \tilde{\mathbb{P}}_n(Y_{nq+1}|x = x_{nq}, q)$ and $\mathbb{I}_n = \mathbb{I}\{t \le 1-1/n\}$.

Now, Step 2 and (D.11) imply $\lim_{n\to\infty} |V_n^*(0) - \breve{V}_n^*(0)| = 0$. But, the value $\pi^* \in \{0,1\}$ that attains the minimum in (D.12) depends only on s. We would have thus obtained the approximation, $\breve{V}_n^*(0)$, to $V_n^*(0)$ even if we restricted the policy class to $\Pi^{\mathcal{S}}$. This proves the first claim of the theorem.

Step 3 (Auxiliary results for showing PDE approximation of (D.12)): We now state a couple of results that will be used to show that the solution, $\breve{V}_n^*(\cdot)$, to (D.12) converges to the solution of a PDE.

The first result is that, for any given $\pi \in \{0,1\}$, $\tilde{\mathbb{E}}[R_n(h,\pi)|\, x,q]$ can be approximated by $\mu^+(s) - \pi\mu(s)$ uniformly over (x,q). To this end, denote $\bar{R}(h,\pi) = \dot{\mu}_0 h\left(\mathbb{I}(\dot{\mu}_0 h > 0) - \pi\right)$. Assumption 1(iii) implies $\sup_{|h|\le\Gamma} |\mu_n(h) - \dot{\mu}_0 h/\sqrt{n}| \le \Gamma^2\delta_n/\sqrt{n}$. Combining this with Lipschitz continuity of $x\mathbb{I}(x>0) - \pi x$ gives

$$\sup_{|h|\le\Gamma;\pi\in\{0,1\}} \left| R_n(h,\pi) - \bar{R}(h,\pi)\right| \le 2\Gamma^2\delta_n.$$

Recalling the definitions of $\mu^+(s), \mu(s)$ from the main text, the above implies

$$\sup_{(x,q);\pi\in\{0,1\}} \left|\tilde{\mathbb{E}}\left[R_n(h,\pi)\middle| x,q\right] - \left(\mu^+(s) - \pi\mu(s)\right)\right| \le 2\Gamma^2\delta_n \to 0. \tag{D.13}$$

The next result is given as Lemma 9 in Appendix E. It states that there exists $\xi_n \to 0$ independent of both s and $\pi \in \{0,1\}$ such that

$$\sqrt{n}\sigma^2\mathbb{E}_{\tilde{\mathbb{P}}_n}\left[\pi\psi(Y_{nq+1})\middle| s\right] = \pi h(s) + \xi_n, \text{ and} \tag{D.14}$$

$$\sigma^4\mathbb{E}_{\tilde{\mathbb{P}}_n}\left[\pi\psi^2(Y_{nq+1})\middle| s\right] = \pi\sigma^2 + \xi_n. \tag{D.15}$$

Furthermore,

$$\mathbb{E}_{\tilde{\mathbb{P}}_n}\left[\left|\psi(Y_{nq+1})\right|^3\middle| s\right] < \infty. \tag{D.16}$$

Step 4 (PDE approximation of (D.12)): The unique solution, $\breve{V}_n^*(s)$, to (D.12) converges locally uniformly to $V_n^*(s)$, the viscosity solution to PDE (2.8). This follows by similar arguments as in the proof of Theorem 2:

Clearly the scheme defined in (D.12) is monotonic. Assumption 1(iii) implies there exists $b < \infty$ such that $\sup_{\pi,|h|\le\Gamma} |R_n(h,\pi)| \le b\Gamma$. Hence, the solution to (D.12) is uniformly bounded, with $|\breve{V}_n^*(s)| \le b\Gamma$ independent of s and n. This proves stability. Finally, consistency of the scheme follows by similar arguments as in the proof of Theorem 2, after making use of (D.13) and (D.14) - (D.16).

This completes the proof of the second claim of the theorem.

Step 5 (Proof of the third claim): Steps 1 and 2 imply $\lim_{n\to\infty} V_{\pi^*_{\Delta t},n}(0) - \breve{V}_{\pi^*_{\Delta t}n}(0) = 0$. In addition, we can follow the arguments in Step 2 to express $\breve{V}_{\pi^*_{\Delta t}n}(0)$ in recursive form, in a manner similar to the definition of $V^*_{\Delta t,n,l}(\cdot)$ in the proof of Theorem 4; the only difference is that the operator $\tilde{S}_{\Delta t}[\phi](x,q)$ in that proof should now read as the solution at $(x,q,\Delta t)$ of the recursive equation

$$f(x,q,\tau) = \frac{\tilde{\mathbb{E}}[R_n(h,1)|\,x,q]}{n} + \mathbb{E}_{\tilde{\mathbb{P}}_n}\left[f\left(x + \frac{\sigma^2\psi(Y_{nq+1})}{\sqrt{n}}, q + \frac{1}{n}, \tau - \frac{1}{n}\right)\bigg|\, s\right];\ \tau > 0$$

$$f(x,q,0) = \phi(x,q).$$

Now, an application of Barles and Jakobsen (2007, Theorem 3.1), using (D.13) - (D.16) to verify the requirements (cf. Appendix B), gives $\left|\tilde{S}_{\Delta t}\left[V^*_{\Delta t,l+1}\right] - S_{\Delta t}\left[V^*_{\Delta t,l+1}\right]\right| \lesssim \min\left\{n^{-1/14}, \xi_n, \delta_n\right\}$. The rest of the proof is analogous to that of Theorem 4.

Appendix E. Supporting lemmas for the proof of Theorem 5

We implicitly assume Assumption 1 for all the results in this section apart from Lemma 1.

Lemma 1. *Let $p(Y|h)$ denote the likelihood of Y given some parameter h with prior distribution $m_0(h)$. Under the one-armed bandit experiment, the posterior distribution, $p_n(\cdot|\mathcal{F}_t)$, of h given all information until time t satisfies*

$$p_n(h|\mathcal{F}_t) \propto \left\{\prod_{i=1}^{\lfloor nq(t)\rfloor} p(Y_i|h)\right\}\cdot m_0(h). \tag{E.1}$$

In particular, the posterior distribution is independent of the past values of actions.

Proof. Note that $\mathcal{F}_t$ is the sigma-algebra generated by $\xi_t \equiv \{\{A_j\}_{j=1}^{\lfloor nt \rfloor}, \{Y_i\}_{i=1}^{\lfloor nq(t) \rfloor}\}$; here, j refers to the time period while i refers to number of pulls of the arm. The claim is shown using induction. Clearly, it is true for $t = 1$. For any $t > 1$, we can think of $p_n(h|\xi_{t-1})$ as the revised prior for μ. Suppose that $A_t = 1$. Then $nq(t) = nq(t-1) + 1$, and

$$\begin{aligned} p_n(h|\xi_t) &\propto p(Y_t, A_t = 1|\xi_t, h) \cdot p_n(h|\xi_{t-1}) \\ &\propto \pi(A_t = 1|\xi_{t-1}) \cdot p(Y_t|h) \cdot p_n(h|\xi_{t-1}) \\ &\propto p(Y_t|h) \cdot p_n(h|\xi_{t-1}) = \left\{ \prod_{i=1}^{\lfloor nq(t) \rfloor} p(Y_i|h) \right\} \cdot m_0(h). \end{aligned}$$

Alternatively, suppose $A_t = 0$. Then, $nq(t) = nq(t-1)$, and $p(A_t = 0|\xi_t, h) = \pi(A_t = 0|\xi_t)$ is independent of h, so

$$\begin{aligned} p_n(h|\xi_t) &\propto p(A_t = 0|\xi_t, h) \cdot p_n(h|\xi_{t-1}) \\ &\propto p_n(h|\xi_{t-1}) = \left\{ \prod_{i=1}^{\lfloor nq(t) \rfloor} p(Y_i|h) \right\} \cdot m_0(h). \end{aligned}$$

Thus the induction step holds under both possibilities, and the claim follows. □

Lemma 2. *Suppose P_θ is quadratic mean differentiable as in (5.1). Then P_θ satisfies the SLAN property as defined in (5.2).*

Proof. The proof builds on Van der Vaart (2000, Theorem 7.2). Set $p_n := dP_{\theta_0 + h/\sqrt{n}}/d\nu$, $p_0 := dP_{\theta_0}/d\nu$ and $W_{ni} := 2\left[\sqrt{p_n/p_0}(Y_i) - 1\right]$. We use $E[\cdot]$ to denote expectations with respect to P_{n,θ_0}. Quadratic mean differentiability implies $E[\psi(Y_i)] = 0$ and $E[\psi^2(Y_i)] = 1/\sigma^2$, see Van der Vaart (2000, Theorem 7.2).

It is without loss of generality for this proof to take the domain of q to be $\{0, 1/n, 2/n, \ldots, 1\}$. For any such q,

$$E\left[\sum_{i=1}^{nq} W_{ni}\right] = 2nq\left(\int \sqrt{p_n \cdot p_0}\, d\nu - 1\right) = -nq \int \left(\sqrt{p_n} - \sqrt{p_0}\right)^2 d\nu.$$

Now, (5.1) implies there exists $\epsilon_n \to 0$ such that

$$\left| n \int \left(\sqrt{p_n} - \sqrt{p_0}\right)^2 d\nu - \frac{h^2}{4\sigma^2} \right| \lesssim \epsilon_n h^2.$$

Hence, for any given h,

$$\sup_q \left| E\left[\sum_{i=1}^{nq} W_{ni}\right] - \frac{qh^2}{4\sigma^2}\right| \to 0. \tag{E.2}$$

Next, denote $Z_{ni} = W_{ni} - h\psi(Y_i)/\sqrt{n} - E[W_{ni}]$ and $S_{nq} = \sum_{i=1}^{nq} Z_{ni}$. Observe that $E[Z_{ni}] = 0$ since $E[\psi(Y_i)] = 0$. Furthermore, by (5.1),

$$\mathrm{Var}[\sqrt{n}Z_{ni}] = E\left[\left(\sqrt{n}W_{ni} - h\psi(Y_i)\right)^2\right] \lesssim \epsilon_n h^2 \to 0. \tag{E.3}$$

Now, an application of Kolmogorov's maximal inequality for partial sum processes gives

$$P\left(\sup_q |S_{nq}| \geq \lambda\right) \leq \frac{1}{\lambda^2}\mathrm{Var}\left[\sum_{i=1}^{n} Z_{ni}\right] = \frac{1}{\lambda^2}\mathrm{Var}[\sqrt{n}Z_{ni}].$$

Combined with (E.2) and (E.3), the above implies

$$\sum_{i=1}^{nq} W_{ni} = \frac{h}{\sqrt{n}}\sum_{i=1}^{nq}\psi(Y_i) - \frac{qh^2}{4\sigma^2} + o_{P_{n,\theta_0}}(1) \text{ uniformly over q.} \tag{E.4}$$

We now employ a Taylor expansion of the logarithm $\ln(1+x) = x - \frac{1}{2}x^2 + x^2 R(2x)$ where $R(x) \to 0$ as $x \to 0$, to expand the log-likelihood as

$$\begin{aligned}\ln \prod_{i=1}^{nq} \frac{p_n}{p_0}(Y_i) &= 2\sum_{i=1}^{nq} \ln\left(1 + \frac{1}{2}W_{ni}\right) \\ &= \sum_{i=1}^{nq} W_{ni} - \frac{1}{4}\sum_{i=1}^{nq} W_{ni}^2 + \frac{1}{2}\sum_{i=1}^{nq} W_{ni}^2 R(W_{ni}). \end{aligned}\tag{E.5}$$

Because of (E.3), we can write $\sqrt{n}W_{ni} = h\psi(Y_i) + C_{ni}$ where $E[|C_{ni}|^2] \to 0$. Defining $A_{ni} := 2h\psi(Y_i)C_{ni} + C_{ni}^2$, some straightforward algebra then gives $nW_{ni}^2 = h^2\psi^2(Y_i) + A_{ni}$ with $E[|A_{ni}|] \to 0$. Now, by the uniform law of large numbers for partial sum processes, see e.g., Bass and Pyke (1984), $n^{-1}\sum_{i=1}^{nq} h^2\psi^2(Y_i)$ converges uniformly in P_{n,θ_0}-probability to qh^2/σ^2. Furthermore, $E\left[\sup_q n^{-1}\sum_{i=1}^{nq}|A_{ni}|\right] \leq E\left[n^{-1}\sum_{i=1}^{n}|A_{ni}|\right] = E[|A_{ni}|] \to 0$ and therefore $n^{-1}\sum_{i=1}^{nq} A_{ni}$ converges uniformly in P_{n,θ_0}-probability to 0. These results yield

$$\sum_{i=1}^{nq} W_{ni}^2 = \frac{qh^2}{\sigma^2} + o_{P_{n,\theta_0}}(1) \text{ uniformly over q.}$$

Next, by the triangle inequality and Markov's inequality

$$\begin{aligned} nP_{n,\theta_0}\left(|W_{ni}| > \varepsilon\sqrt{2}\right) &\le nP_{n,\theta_0}\left(h^2\psi^2(Y_i) > n\varepsilon^2\right) + nP_{n,\theta_0}\left(|A_{ni}| > n\varepsilon^2\right) \\ &\le \varepsilon^{-2}h^2 E\left[\psi^2(Y_i)\mathbb{I}\left\{\psi^2(Y_i) > n\varepsilon^2\right\}\right] + \varepsilon^{-2}E\left[|A_{ni}|\right] \to 0 \end{aligned}$$

for any given h. The above implies $\max_{1\le i\le n}|W_{ni}| = o_{P_{n,\theta_0}}(1)$ and consequently, $\max_{1\le i\le n}|R(W_{ni})| = o_{P_{n,\theta_0}}(1)$. The last term on the right hand side of (E.5) is bounded by $\max_{1\le i\le n}|R(W_{ni})| \cdot \sum_{i=1}^{n} W_{ni}^2$ and is therefore $o_{P_{n,\theta_0}}(1)$ by the above results. We thus conclude

$$\ln \prod_{i=1}^{nq} \frac{p_n}{p_0}(Y_i) = \sum_{i=1}^{nq} W_{ni} - \frac{qh^2}{4\sigma^2} + o_{P_{n,\theta_0}}(1) \text{ uniformly over q.}$$

The claim follows by combining the above with (E.4). □

Lemma 3. *For any $\epsilon > 0$, there exist $M(\epsilon), N(\epsilon) < \infty$ such that $M \ge M(\epsilon)$ and $n \ge N(\epsilon)$ implies $\bar{P}_n(A_n^c) < \epsilon$. Furthermore, letting $A_n^q = \left\{\mathbf{y}_{nq} : \sup_{\tilde{q}\le q}|x_{n\tilde{q}}| < M\right\}$, and $\mathbb{E}_{n,0}[\cdot]$, the expectation under P_{n,θ_0},*

$$\sup_q \mathbb{E}_{n,0}\left[\mathbb{I}_{A_n^q}\left\|\frac{dP_{nq,\theta_0+h/\sqrt{n}}}{dP_{nq,\theta_0}}(\mathbf{y}_{nq}) - \frac{d\Lambda_{nq,h}}{dP_{nq,\theta_0}}(\mathbf{y}_{nq})\right\|\right] = o(1)\ \forall\ \{h : |h| \le \Gamma\}.$$

Proof. Set $A_{n,M} = \left\{\mathbf{y}_n : \sup_q |x_{nq}| < M\right\}$ and $P_{nq,h} = P_{nq,\theta_0+h/\sqrt{n}}$. Note that x_{nq} is a partial sum process with mean 0 under $P_{n,0} := P_{n,\theta_0}$. By Kolmogorov's maximal inequality, $P_{n,0}\left(\sup_q|x_{nq}| \ge M\right) \le M^{-1}\text{Var}[x_n] = M^{-1}\sigma^2$. Hence, $P_{n.0}(A_{n,M_n}^c) \to 0$ for any $M_n \to \infty$. But by (5.2) and standard arguments involving Le Cam's first lemma, $P_{n,h}$ is contiguous to $P_{n,0}$ for all h. This implies $\bar{P}_n := \int P_{n,h}dm_0(h)$ is also contiguous to $P_{n,0}$ (this can be shown using the dominated convergence theorem; see also, Le Cam and Yang, p.138). Consequently, $\bar{P}_n(A_{n,M_n}^c) \to 0$ for any $M_n \to \infty$. The first claim is a straightforward consequence of this.

For the second claim, we follow Le Cam and Yang (2000, Proposition 6.2):

We first argue that $P_{nq_n,h}$ is contiguous to $P_{nq_n,0}$ for any deterministic sequence $\{q_n\}$ such that $q_n \to \bar{q} \in [0,1]$. We have

$$\begin{aligned} \ln\frac{dP_{nq_n,h}}{dP_{nq_n,0}} &= \frac{1}{\sigma^2}hx_{nq_n} - \frac{q_n}{2\sigma^2}h^2 + o_{P_{n,0}}(1) \\ &\xrightarrow[P_{n,0}]{d} N\left(-\frac{\bar{q}h^2}{2\sigma^2}, \frac{\bar{q}h^2}{\sigma^2}\right), \end{aligned} \tag{E.6}$$

where the equality follows from (5.2), and the weak convergence limit follows from: (i) weak convergence of x_{nq} under $P_{n,0}$ to a Brownian motion process $W(q)$, see e.g., Van Der Vaart and Wellner (1996, Chapter 2.12), and (ii) the extended continuous mapping theorem, see Van Der Vaart and Wellner (1996, Theorem 1.11.1). Since $E_{P_{n,0}}[f(\mathbf{y}_{nq_n})] = E_{P_{nq_n,0}}[f(\mathbf{y}_{nq_n})]$ for any $f(\cdot)$, we conclude from (E.6) and the definition of weak convergence that

$$\ln \frac{dP_{nq_n,h}}{dP_{nq_n,0}} \xrightarrow[P_{nq_n,0}]{d} N\left(-\frac{\bar{q}h^2}{2\sigma^2}, \frac{\bar{q}h^2}{\sigma^2}\right).$$

An application of Le Cam's first lemma then implies $P_{nq_n,h}$ is contiguous to $P_{nq_n,0}$.

Now, let $q_n \in [0,1]$ denote a quantity such that

$$\sup_q \mathbb{E}_{n,0}\left[\mathbb{I}_{A_n^q}\left\|\frac{dP_{nq,h}}{dP_{nq,0}} - \frac{d\Lambda_{nq,h}}{dP_{nq,0}}\right\|\right] \leq \mathbb{E}_{n,0}\left[\mathbb{I}_{A_n^{q_n}}\left\|\frac{dP_{nq_n,h}}{dP_{nq_n,0}} - \frac{d\Lambda_{nq_n,h}}{dP_{nq_n,0}}\right\|\right] + \epsilon$$

for some arbitrarily small $\epsilon \geq 0$ (such a q_n, ϵ always exist by the definition of the supremum). Without loss of generality, we may assume q_n converges to some $\bar{q} \in [0,1]$; otherwise we can employ a subsequence argument since q_n lies in a bounded set. Define

$$G_n(q) := \mathbb{I}_{A_n^{q_n}}\left\|\frac{dP_{nq,h}}{dP_{nq,0}} - \frac{d\Lambda_{nq,h}}{dP_{nq,0}}\right\|.$$

The claim follows if we show $\mathbb{E}_{n,0}[G_n(q_n)] \to 0$. By Lemma 2 and the definition of $\Lambda_{nq,h}(\cdot)$,

$$G_n(q) = \mathbb{I}_{A_n^{q_n}} \cdot \exp\left\{\frac{1}{\sigma^2}hx_{nq} - \frac{q}{2\sigma^2}h^2\right\}(\exp\delta_{n,q} - 1),$$

where $\sup_q |\delta_{n,q}| = o(1)$ under $P_{n,0}$. Since $\mathbb{I}_{A_n^{q_n}} \cdot \exp\left\{\frac{1}{\sigma^2}hx_{nq_n} - \frac{q_n}{2\sigma^2}h^2\right\}$ is bounded for $|h| \leq \Gamma$ by the definition of $\mathbb{I}_{A_n^q}$, this implies $G_n(q_n) = o(1)$ under $P_{n,0}$. Next, we argue $G_n(q_n)$ is uniformly integrable. The term $\mathbb{I}_{A_n^{q_n}} \cdot d\Lambda_{nq_n,h}/dP_{nq_n,0}$ in the definition of $G_n(q_n)$ is bounded, and therefore uniformly integrable, for $|h| \leq \Gamma$. We now prove uniform integrability of $dP_{nq_n,h}/dP_{nq_n,0}$, and thereby that of the remaining term, $\mathbb{I}_{A_n^{q_n}} \cdot dP_{nq_n,h}/dP_{nq_n,0}$, in the definition of $G_n(q_n)$. For any $b < \infty$,

$$\begin{aligned}\mathbb{E}_{n,0}\left[\frac{dP_{nq_n,h}}{dP_{nq_n,0}}\mathbb{I}\left\{\frac{dP_{nq_n,h}}{dP_{nq_n,0}} > b\right\}\right] &= \int \frac{dP_{nq_n,h}}{dP_{nq_n,0}}\mathbb{I}\left\{\frac{dP_{nq_n,h}}{dP_{nq_n,0}} > b\right\} dP_{nq_n,0} \\ &\leq P_{nq_n,h}\left(\frac{dP_{nq_n,h}}{dP_{nq_n,0}} > b\right).\end{aligned}$$

But,

$$P_{nq_n,0}\left(\frac{dP_{nq_n,h}}{dP_{nq_n,0}}>b\right)\leq b^{-1}\int\frac{dP_{nq_n,h}}{dP_{nq_n,0}}dP_{nq_n,0}\leq b^{-1},$$

so the contiguity of $P_{nq_n,h}$ with respect to $P_{nq_n,0}$ implies we can choose b and $\bar{n}$ large enough such that

$$\limsup_{n\geq\bar{n}} P_{nq_n,h}\left(\frac{dP_{nq_n,h}}{dP_{nq_n,0}}>b\right)<\epsilon$$

for any arbitrarily small ϵ. These results demonstrate uniform integrability of $G_n(q_n)$ under $P_{n,0}$. Since convergence in probability implies convergence in expectation for uniformly integrable random variables, we have thus shown $\mathbb{E}_{n,0}\left[G_n(q_n)\right]\to 0$, which concludes the proof. ☐

Lemma 4. $\lim_{n\to\infty}\left\|\bar{P}_n\cap A_n-\tilde{\bar{P}}_n\cap A_n\right\|_{TV}=0.$

Proof. Set $P_{n,h}:=P_{n,\theta_0+h/\sqrt{n}}$. By the properties of the total variation metric, contiguity of $\bar{P}_n$ with respect to $P_{n,0}$ and the absolute continuity of $\Lambda_{n,h}$ with respect to $P_{n,0}$,

$$\begin{aligned}&\lim_{n\to\infty}\left\|\bar{P}_n\cap A_n-\tilde{\bar{P}}_n\cap A_n\right\|_{\text{TV}}\\&=\frac{1}{2}\lim_{n\to\infty}\int\left\{\int\mathbb{I}_{A_n}\left|\frac{dP_{n,h}}{dP_{n,0}}(\mathbf{y}_n)-\frac{d\Lambda_{n,h}}{dP_{n,0}}(\mathbf{y}_n)\right|dP_{n,0}(\mathbf{y}_n)\right\}m_0(h)d\nu(h).\end{aligned}$$

In the last expression, denote the term within the $\{\}$ brackets by $f_n(h)$. By Lemma 3, $f_n(h)\to 0$ for each h. Additionally, $\mathbb{I}_{A_n}\cdot(d\Lambda_{n,h}/dP_{n,0})$ is bounded because of the definition of A_n and the fact $|h|\leq\Gamma$, while

$$\int\mathbb{I}_{A_n}\left|\frac{dP_{n,h}}{dP_{n,0}}\right|dP_{n,0}\leq\int\frac{dP_{n,h}}{dP_{n,0}}dP_{n,0}\leq 1.$$

Hence, $f_n(h)$ is dominated by a (suitably large) constant for all n. The dominated convergence theorem then implies $\int f_n(h)m_0(h)d\nu(h)\to 0$. This proves the claim.

☐

Lemma 5. $\sup_q\bar{\mathbb{E}}_n\left[\mathbb{I}_{A_n}\left\|p_n(\cdot|\mathbf{y}_{nq})-\tilde{p}_n(\cdot|\mathbf{y}_{nq})\right\|_{TV}\right]=o(1).$

Proof. Set $P_{n,h}=P_{n,\theta_0+h/\sqrt{n}}$, $p_{nq,h}(\mathbf{y}_{nq})=dP_{nq,h}(\mathbf{y}_{nq})/d\nu$, $\lambda_{nq,h}(\mathbf{y}_{nq})=d\Lambda_{nq,h}(\mathbf{y}_{nq})/d\nu$, $\bar{p}_{nq}(\mathbf{y}_{nq})=d\bar{P}_{nq}(\mathbf{y}_{nq})/d\nu$ and $\tilde{\bar{p}}_{nq}(\mathbf{y}_{nq})=d\tilde{\bar{P}}_{nq}(\mathbf{y}_{nq})/d\nu$. Let S_{nq} and $\tilde{S}_{nq}$ denote joint measures over $(\mathbf{y}_{nq},h)$, corresponding to $dS_{nq}(\mathbf{y}_{nq},h)=p_{nq,h}(\mathbf{y}_{nq})\cdot m_0(h)$ and $d\tilde{S}_{nq}(\mathbf{y}_{nq},h)=\lambda_{nq,h}(\mathbf{y}_{nq})\cdot m_0(h)$.

In the main text, we introduced the approximate posterior $\tilde{p}_n(h|\mathbf{y}_{nq})$. Formally, this is defined via the disintegration $d\tilde{S}_{nq}(\mathbf{y}_{nq}, h) = \tilde{p}_n(h|\mathbf{y}_{nq}) \cdot d\tilde{\bar{P}}_n(\mathbf{y}_{nq})$, where $d\tilde{\bar{P}}_n(\mathbf{y}_{nq}) := \int \left\{ d\tilde{S}_{nq}(\mathbf{y}_{nq}, h) \right\} d\nu(h)$. Such a conditional probability always exists, see, e.g., Le Cam and Yang (2000, p. 136). In a similar vein, we can disintegrate $dS_{nq} = p_n(h|\mathbf{y}_{nq}) \cdot \bar{p}_{nq}(\mathbf{y}_{nq})$. Since $p_n(h|\mathbf{y}_{nq}), \tilde{p}_n(h|\mathbf{y}_{nq})$ are both conditional probabilities, we obtain $\bar{p}_{nq}(\mathbf{y}_{nq}) = \int p_{nq,h}(\mathbf{y}_{nq}) m_0(h) d\nu(h)$ and $\tilde{\bar{p}}_{nq}(\mathbf{y}_{nq}) = \int \lambda_{nq,h}(\mathbf{y}_{nq}) m_0(h) d\nu(h)$.

Define $\Omega_n \equiv \{\mathbf{y}_n : p_{n,0}(\mathbf{y}_n) \neq 0\}$. Since the total variation metric is bounded by 1 and $\bar{P}_n$ is contiguous with respect to $P_{n,0}$,

$$\sup_q \bar{\mathbb{E}}_n \left[\mathbb{I}_{A_n} \|p_n(\cdot|\mathbf{y}_{nq}) - \tilde{p}_n(\cdot|\mathbf{y}_{nq})\|_{\text{TV}} \right] = \sup_q \bar{\mathbb{E}}_n \left[\mathbb{I}_{A_n \cap \Omega_n} \|p_n(\cdot|\mathbf{y}_{nq}) - \tilde{p}_n(\cdot|\mathbf{y}_{nq})\|_{\text{TV}} \right] + o(1).$$

Now, by the properties of the total variation metric and the disintegration formula,

$$\begin{aligned} 2\|p_n(\cdot|\mathbf{y}_{nq}) - \tilde{p}_n(\cdot|\mathbf{y}_{nq})\|_{\text{TV}} &= \int |p_n(h|\mathbf{y}_{nq}) - \tilde{p}_n(h|\mathbf{y}_{nq})| \, d\nu(h) \\ &= \int \left| \frac{p_{nq,h}(\mathbf{y}_{nq}) \cdot m_0(h)}{\bar{p}_{nq}(\mathbf{y}_{nq})} - \frac{\lambda_{nq,h}(\mathbf{y}_{nq}) \cdot m_0(h)}{\tilde{\bar{p}}_{nq}(\mathbf{y}_{nq})} \right| d\nu(h). \end{aligned}$$

Hence,

$$\begin{aligned} &2\bar{\mathbb{E}}_n \left[\mathbb{I}_{A_n \cap \Omega_n} \|p_n(\cdot|\mathbf{y}_{nq}) - \tilde{p}_n(\cdot|\mathbf{y}_{nq})\|_{\text{TV}} \right] \\ &\leq \bar{\mathbb{E}}_n \left[\mathbb{I}_{A_n \cap \Omega_n} \int \frac{|p_{nq,h}(\mathbf{y}_{nq}) - \lambda_{nq,h}(\mathbf{y}_{nq})|}{\bar{p}_{nq}(\mathbf{y}_{nq})} m_0(h) d\nu(h) \right] \\ &\quad + \bar{\mathbb{E}}_n \left[\mathbb{I}_{A_n \cap \Omega_n} \int \lambda_{nq,h}(\mathbf{y}_{nq}) \left| \frac{1}{\bar{p}_{nq}(\mathbf{y}_{nq})} - \frac{1}{\tilde{\bar{p}}_{nq}(\mathbf{y}_{nq})} \right| m_0(h) d\nu(h) \right] \\ &:= B_{1n}(q) + B_{2n}(q) \end{aligned}$$

We start by bounding $\sup_q B_{1n}(q)$. Recall the definition of $A_n^q \supseteq A_n$ from the statement of Lemma 3. By Fubini's theorem and the definition of $\bar{p}_{nq}(\cdot)$ as the density of $\bar{P}_{nq}$,

$$\begin{aligned} B_{1n}(q) &\leq \int \left\{ \int \mathbb{I}_{A_n^q \cap \Omega_n} |p_{nq,h}(\mathbf{y}_{nq}) - \lambda_{nq,h}(\mathbf{y}_{nq})| \, d\nu(\mathbf{y}_{nq}) \right\} m_0(h) d\nu(h) \\ &\leq \int \left\{ \int \mathbb{I}_{A_n^q} \left| \frac{dP_{nq,h}}{dP_{nq,0}}(\mathbf{y}_{nq}) - \frac{d\Lambda_{nq,h}}{dP_{nq,0}}(\mathbf{y}_{nq}) \right| dP_{nq,0}(\mathbf{y}_{nq}) \right\} m_0(h) d\nu(h), \quad \text{(E.7)} \end{aligned}$$

the change of measure to $P_{nq,0}$ in the last inequality being allowed under Ω_n. Hence,

$$\sup_q B_{1n}(q) \leq \int \left\{ \sup_q \int \mathbb{I}_{A_n^q} \left| \frac{dP_{nq,h}}{dP_{nq,0}}(\mathbf{y}_{nq}) - \frac{d\Lambda_{nq,h}}{dP_{nq,0}}(\mathbf{y}_{nq}) \right| dP_{nq,0}(\mathbf{y}_{nq}) \right\} m_0(h) d\nu(h).$$

In the above expression, denote the term within the {} brackets by $g_n(h)$. By Lemma 3, $g_n(h) \to 0$ for each h. Furthermore, by similar arguments as in the proof of Lemma 4, $g_n(h)$ is bounded by a constant for all n (it is easy to see that the bound derived there applies uniformly over all q). The dominated convergence theorem then gives $\int g_n(h) m_0(h) d\nu(h) \to 0$, and therefore, $\sup_q B_{1n}(q) = o(1)$.

We now turn to $B_{2n}(q)$. The disintegration formula implies $\lambda_{nq,h}(\mathbf{y}_{nq}) \cdot m_0(h) = \tilde{\bar{p}}_{nq}(\mathbf{y}_{nq}) \cdot \tilde{p}_n(h|\mathbf{y}_{nq})$. So,

$$\begin{aligned} B_{2n}(q) &= \bar{\mathbb{E}}_n \left[\mathbb{I}_{A_n \cap \Omega_n} \int \tilde{p}_n(h|\mathbf{y}_{nq}) \left| \frac{\tilde{\bar{p}}_{nq}(\mathbf{y}_{nq}) - \bar{p}_{nq}(\mathbf{y}_{nq})}{\bar{p}_{nq}(\mathbf{y}_{nq})} \right| d\nu(h) \right] \\ &= \bar{\mathbb{E}}_n \left[\mathbb{I}_{A_n \cap \Omega_n} \left| \frac{\tilde{\bar{p}}_{nq}(\mathbf{y}_{nq}) - \bar{p}_{nq}(\mathbf{y}_{nq})}{\bar{p}_{nq}(\mathbf{y}_{nq})} \right| \right] \\ &\leq \int \mathbb{I}_{A_n^q \cap \Omega_n} \left| \tilde{\bar{p}}_{nq}(\mathbf{y}_{nq}) - \bar{p}_{nq}(\mathbf{y}_{nq}) \right| d\nu(\mathbf{y}_{nq}). \end{aligned} \tag{E.8}$$

By the integral representation for $\tilde{\bar{p}}_{nq}(\mathbf{y}_{nq}), \bar{p}_{nq}(\mathbf{y}_{nq})$ the right hand side of (E.8) equals

$$\begin{aligned} &\int \mathbb{I}_{A_n^q \cap \Omega_n} \left| \int \frac{d\Lambda_{nq,h}}{dP_{nq,0}}(\mathbf{y}_{nq}) dm_0(h) - \int \frac{dP_{nq,h}}{dP_{nq,0}}(\mathbf{y}_{nq}) dm_0(h) \right| dP_{nq,0}(\mathbf{y}_{nq}) \\ &\leq \int \left\{ \int \mathbb{I}_{A_n^q} \left| \frac{d\Lambda_{nq,h}}{dP_{nq,0}}(\mathbf{y}_{nq}) - \frac{dP_{nq,h}}{dP_{nq,0}}(\mathbf{y}_{nq}) \right| dP_{nq,0}(\mathbf{y}_{nq}) \right\} m_0(h) d\nu(h), \end{aligned} \tag{E.9}$$

where the second step makes use of Fubini's theorem. The right hand side of (E.9) is the same as in (E.7). So, by the same arguments as before, $\sup_q B_{2n}(q) = o(1)$. The claim can therefore be considered proved. □

Lemma 6. *Let $\mathcal{Y}_n$ denote the domain of $\mathbf{y}_n$. Then, $\lim_{n\to\infty} \sup_{|h|\leq\Gamma} \Lambda_{n,h}(\mathcal{Y}_n) = 1$, and $\Lambda_{n,h}$ is contiguous to P_{n,θ_0}. Furthermore, $\lim_{n\to\infty} \tilde{\bar{P}}_n(\mathcal{Y}_n) = 1$, $\tilde{\bar{P}}_n$ is contiguous to P_{n,θ_0} and for each $\epsilon > 0$ there exists $M(\epsilon), N(\epsilon) < \infty$ such that $\tilde{\bar{P}}_n(A_n^c) < \epsilon$ for all $M \geq M(\epsilon)$ and $n \geq N(\epsilon)$.*

Proof. Set $P_{n,h} := P_{n,\theta_0+h/\sqrt{n}}$ and $p_{n,h} = dP_{n,h}/d\nu$. Note that $p_{n,0}(\mathbf{y}_n) = \prod_{i=1}^n p_0(Y_i)$, where $p_0(\cdot)$ is the density function of $P_{\theta_0}(Y)$. Then, by the definition of $\Lambda_{n,h}$ and

$\lambda_{n,h}(\cdot)$, we can write $\Lambda_{n,h}(\mathcal{Y}_n) \equiv \int \lambda_{n,h}(\mathbf{y}_n) d\nu(\mathbf{y}_n)$ as

$$\Lambda_{n,h}(\mathcal{Y}_n) = (a_n(h))^n \text{ where}$$
$$a_n(h) := \int \exp\left\{\frac{h}{\sqrt{n}}\psi(Y_i) - \frac{h^2}{2\sigma^2 n}\right\} p_0(Y_i) d\nu(Y_i).$$

Denote $g_n(h,Y) = \frac{h}{\sqrt{n}}\psi(Y) - \frac{h^2}{2\sigma^2 n}$, $\delta_n(h,Y) = \exp\{g_n(h,Y)\} - \{1 + g_n(h,Y) + g_n^2(h,Y)/2\}$ and $\mathbb{E}_{p_0}[\cdot]$, the expectation corresponding to $p_0(Y)$. Then,

$$\begin{aligned} a_n(h) &= \mathbb{E}_{p_0}\left[\exp\left\{\frac{h}{\sqrt{n}}\psi(Y) - \frac{h^2}{2\sigma^2 n}\right\}\right] \\ &= \mathbb{E}_{p_0}\left[1 + g_n(h,Y) + \frac{1}{2}g_n^2(h,Y)\right] + \mathbb{E}_{p_0}\left[\delta_n(h,Y)\right] \\ &:= Q_{n1}(h) + Q_{n2}(h). \end{aligned} \tag{E.10}$$

Since $\psi(\cdot)$ is the score function at θ_0, $\mathbb{E}_{p_0}[\psi(Y)] = 0$ and $\mathbb{E}_{p_0}[\psi^2(Y)] = 1/\sigma^2$. Using these results and the fact $|h| \le \Gamma$, straightforward algebra implies

$$Q_{n1}(h) = 1 + b_n, \text{ where } b_n \le \Gamma^4/8\sigma^4 n^2.$$

We can expand Q_{n2} as follows:

$$Q_{n2}(h) = \mathbb{E}_{p_0}\left[\mathbb{I}_{\psi(Y)\le K}\delta_n(h,Y)\right] + \mathbb{E}_{p_0}\left[\mathbb{I}_{\psi(Y)>K}\delta_n(h,Y)\right]. \tag{E.11}$$

Since $|h| \le \Gamma$ and $e^x - (1+x+x^2/2) = O(|x|^3)$, the first term in (E.11) is bounded by $K^3\Gamma^2 n^{-3/2}$. Furthermore, for large enough n, the second term in (E.11) is bounded by $\mathbb{E}_{p_{\theta_0}}[\exp|\psi(Y)|]/\exp(aK)$ for any $a < 1$. Hence, setting $K = (3/2a)\ln n$ gives $\sup_{|h|\le\Gamma} Q_{n2}(h) = O\left(\ln^3 n/n^{3/2}\right)$. In view of the above,

$$\sup_{|h|\le\Gamma} |a_n(h) - 1| = O(n^{-c}) \text{ for any } c < 3/2.$$

Thus, $\sup_{|h|\le\Gamma} |\Lambda_{n,h}(\mathcal{Y}_n) - 1| = |\{1 + O(n^{-c})\}^n - 1| = O(n^{-(c-1)})$. Since it is possible to choose any $c < 3/2$, this proves the first claim.

Under $P_{n,0}$, the likelihood $d\Lambda_{n,h}/dP_{n,0}$ converges weakly to some V satisfying $\mathbb{E}_{P_{n,0}}[V] = 1$ (the argument leading to this is standard, see, e.g., Van der Vaart, 2000, Example 6.5). Since $\Lambda_{n,h}(\mathcal{Y}_n) \to 1$, an application of Le Cam's first lemma implies $\Lambda_{n,h}$ is contiguous with respect to $P_{n,0}$.

Because $m_0(\cdot)$ is supported on $|h| \le \Gamma$, $|\tilde{\tilde{P}}_n(\mathcal{Y}_n) - 1| \le \int |\Lambda_{n,h}(\mathcal{Y}_n) - 1| m_0(h) d\nu(h) = O(n^{-(c-1)})$. Thus, $\lim_{n\to\infty} \tilde{\tilde{P}}_n(\mathcal{Y}_n) = 1$. Contiguity of $\tilde{\tilde{P}}_n$ with respect to $P_{n,0}$ follows from the contiguity of $\Lambda_{n,h}$ with respect to $P_{n,0}$. The final claim, that $\tilde{\tilde{P}}_n(A_n^c) < \epsilon$, follows by similar arguments as in the proof of Lemma 3. □

Lemma 7. *The measure, $\tilde{\tilde{P}}_n$, can be disintegrated as in equation (D.8).*

Proof. Let $\lambda_{nq,h}(\cdot)$, $\tilde{S}_{nq}$ be defined as in the proof of Lemma 5. Equation (D.7) implies

$$\lambda_{n,h}(\mathbf{y}_n) \cdot m_0(h) = \lambda_{n-1,h}(\mathbf{y}_{n-1}) \cdot m_0(h) \cdot \tilde{p}(Y_n|h). \tag{E.12}$$

Let $\tilde{S}_{n-1}$ denote the probability measure corresponding to the density $d\tilde{S}_{n-1} = \lambda_{n-1,h}(\mathbf{y}_{n-1}) \cdot m_0(h)$. As argued in the proof of Lemma 5, one can disintegrate this as $d\tilde{S}_{n-1} = p_n(h|\mathbf{y}_{n-1}) \cdot \tilde{\bar{p}}_{n-1}(\mathbf{y}_{n-1})$, where $p_n(h|\mathbf{y}_{n-1})$ is a conditional probability density and $\tilde{\bar{p}}_{n-1}(\mathbf{y}_{n-1}) = \int \lambda_{n-1,h}(\mathbf{y}_{n-1}) m_0(h) d\nu(h)$. Thus,

$$\lambda_{n-1,h}(\mathbf{y}_{n-1}) \cdot m_0(h) = p_n(h|\mathbf{y}_{n-1}) \cdot \tilde{\bar{p}}_{n-1}(\mathbf{y}_{n-1}).$$

Combining the above with (E.12) gives

$$\lambda_{n,h}(\mathbf{y}_n) \cdot m_0(h) = p_n(h|\mathbf{y}_{n-1}) \cdot \tilde{\bar{p}}_{n-1}(\mathbf{y}_{n-1}) \cdot \tilde{p}(Y_n|h).$$

Taking the integral with respect h on both sides, and making use of the definition of $\tilde{\bar{p}}_n(\cdot)$,

$$\tilde{\bar{p}}_n(\mathbf{y}_n) = \tilde{\bar{p}}_{n-1}(\mathbf{y}_{n-1}) \cdot \int \tilde{p}(Y_n|h) p_n(h|\mathbf{y}_{n-1}) d\nu(h). \tag{E.13}$$

There is nothing special about the choice of n here, so iterating the above expression gives the desired result, (D.8). □

Lemma 8. *Let $c_{n,i}$ and $\tilde{\mathbb{P}}_n$ denote the quantities defined in Step 4 of the proof of Theorem 5. There exists some non-random $C < \infty$ such that $\sup_i |c_{n,i} - 1| \le Cn^{-c}$ for any $c < 3/2$. Furthermore, $\lim_{n\to\infty} \left\| \tilde{\mathbb{P}}_n - \tilde{\tilde{P}}_n \right\|_{TV} = 0$.*

Proof. Denote

$$a_n(h) := \int \tilde{p}_n(Y_i|h) d\nu(Y_i) = \int \exp\left\{ \frac{h}{\sqrt{n}} \psi(Y_i) - \frac{h^2}{2\sigma^2 n} \right\} p_0(Y_i) d\nu(Y_i).$$

It is shown in the proof of Lemma 6 that $\sup_{|h|\le\Gamma}|a_n(h)-1| = O(n^{-c})$ for any $c < 3/2$. Since $c_{n,i} = \int a_n(h)\tilde{p}_n(h|\mathbf{y}_{i-1})d\nu(h)$, and $\tilde{p}_n(h|\mathbf{y}_{i-1})$ is a probability density, this proves the first claim.

For the second claim, denote $\tilde{p}_n(Y_i|\mathbf{y}_{i-1}) := \int \tilde{p}_n(Y_i|h)\tilde{p}_n(h|\mathbf{y}_{i-1})d\nu(h)$. We also write $c_{n,i}(\mathbf{y}_{i-1})$ for $c_{n,i}$ to make it explicit that this quantity depends on $\mathbf{y}_{i-1}$. The properties of the total variation metric, along with (D.8) and (D.9) imply

$$\begin{aligned}\left\|\tilde{\mathbb{P}}_n - \tilde{\tilde{P}}_n\right\|_{\text{TV}} &= \frac{1}{2}\int\left|\frac{d\tilde{\mathbb{P}}_n}{d\nu} - \frac{d\tilde{\tilde{P}}_n}{d\nu}\right|d\nu \\ &= \frac{1}{2}\int\prod_{i=1}^n \tilde{p}_n(Y_i|\mathbf{y}_{i-1})\left|\prod_{i=1}^n \frac{1}{c_{n,i}(\mathbf{y}_{i-1})} - 1\right|d\nu(\mathbf{y}_n) \\ &\le \frac{1}{2}\sup_{\mathbf{y}_n}\left|\prod_{i=1}^n \frac{1}{c_{n,i}(\mathbf{y}_{i-1})} - 1\right|\cdot\int\prod_{i=1}^n \tilde{p}_n(Y_i|\mathbf{y}_{i-1})d\nu(\mathbf{y}_n).\end{aligned}$$

Recall from (D.8) that $\prod_{i=1}^n \tilde{p}_n(Y_i|\mathbf{y}_{i-1})$ is the density (wrt ν) of $\tilde{\tilde{P}}_n$, so the integral in the above expression equals $\int d\tilde{\tilde{P}}_n = \tilde{\tilde{P}}_n(\mathcal{Y}) \to 1$ by Lemma 6. Furthermore, using the first claim of the present lemma, it is straightforward to show

$$\sup_{\mathbf{y}_n}\left|\prod_{i=1}^n \frac{1}{c_{n,i}(\mathbf{y}_{i-1})} - 1\right| = O(n^{-(c-1)}).$$

Thus, $\left\|\tilde{\mathbb{P}}_n - \tilde{\tilde{P}}_n\right\|_{\text{TV}} = O(n^{-(c-1)})$ and the claim follows. ∎

Lemma 9. *For the probability measure $\tilde{\mathbb{P}}_n$ defined in Step 4 of the proof of Theorem 5, there exists a deterministic sequence $\xi_n \to 0$ independent of s and $\pi \in \{0,1\}$ such that equations (D.14) - (D.16) hold.*

Proof. Start with (D.14). We have

$$\begin{aligned}\mathbb{E}_{\tilde{\mathbb{P}}_n}\left[\psi(Y_{nq+1})|\, s\right] &= c_{n,nq+1}^{-1}\int\left\{\int \psi(Y_{nq+1})\tilde{p}_n(Y_{nq+1}|h)d\nu(Y_{nq+1})\right\}\tilde{p}(h|x,q)d\nu(h) \\ &= c_{n,nq+1}^{-1}\int \mathbb{E}_{p_{\theta_0}}\left[\psi(Y)\exp\left\{\frac{h}{\sqrt{n}}\psi(Y) - \frac{h^2}{2\sigma^2 n}\right\}\right]\tilde{p}(h|x,q)d\nu(h) \\ &= \left(1 + O(n^{-c})\right)\cdot\int \mathbb{E}_{p_{\theta_0}}\left[\psi(Y)\exp\left\{\frac{h}{\sqrt{n}}\psi(Y) - \frac{h^2}{2\sigma^2 n}\right\}\right]\tilde{p}(h|x,q)d\nu(h),\end{aligned}$$

where the second equality follows by the definition of $\tilde{p}(Y_i|h)$, and the third equality follows by (D.10), where it may be recalled we can choose any $c \in (0, 3/2)$. Define

$g_n(h,Y) = \frac{h}{\sqrt{n}}\psi(Y) - \frac{h^2}{2\sigma^2 n}$ and $\delta_n(h,Y) = \exp\{g_n(h,Y)\} - \{1+g_n(h,Y)\}$. Then,

$$\mathbb{E}_{p_{\theta_0}}\left[\psi(Y)\exp\left\{\frac{h}{\sqrt{n}}\psi(Y) - \frac{h^2}{2\sigma^2 n}\right\}\right]$$
$$= \mathbb{E}_{p_{\theta_0}}\left[\psi(Y)\left\{1 + \frac{h}{\sqrt{n}}\psi(Y) - \frac{h^2}{2\sigma^2 n}\right\}\right] + \mathbb{E}_{p_{\theta_0}}\left[\psi(Y)\delta_n(h,Y)\right].$$

Assumption 1(i) implies, see e.g., Van der Vaart (2000, Theorem 7.2), $\mathbb{E}_{p_{\theta_0}}[\psi(Y)] = 0$ and $\mathbb{E}_{p_{\theta_0}}[\psi^2(Y)] = 1/\sigma^2$. Hence, the first term in the above expression equals $h/(\sqrt{n}\sigma^2)$. For the second term,

$$\mathbb{E}_{p_{\theta_0}}\left[\psi(Y)\delta_n(h,Y)\right] = \mathbb{E}_{p_{\theta_0}}\left[\mathbb{I}_{\psi(Y)\le K}\psi(Y)\delta_n(h,Y)\right] + \mathbb{E}_{p_{\theta_0}}\left[\mathbb{I}_{\psi(Y)>K}\psi(Y)\delta_n(h,Y)\right]. \tag{E.14}$$

Since $|h| \le \Gamma$ and $e^x - (1+x) = o(x^2)$, the first term in in (E.14) is bounded by $K^3\Gamma^2 n^{-1}$. The second term in (E.14) is bounded by $\mathbb{E}_{p_{\theta_0}}[\exp|\psi(Y)|]/\exp(aK)$ for any $a < 1$. Hence, setting $K = (1/a)\ln n$ gives $\sup_{|h|\le\Gamma}|\mathbb{E}_{p_{\theta_0}}[\psi(Y)\delta_n(h,Y)]| = O(\ln^3 n/n)$. Combining the above results and noting that $|h| \le \Gamma$, we obtain

$$\sqrt{n}\sigma^2\mathbb{E}_{\tilde{\mathbb{P}}_n}\left[\psi(Y_{nq+1})|\, s\right] = \left(1+O(n^{-c})\right)\cdot\left\{\int h\tilde{p}(h|x,q)d\nu(h) + O(\ln n/\sqrt{n})\right\} = h(s)+\xi_n,$$

where $\xi_n \asymp \ln n/\sqrt{n}$. This proves (D.14). The proofs of (D.15) and (D.16) are similar. □

Appendix F. Additional details and proof of Theorem 6 for non-parametric models

We start with a formal definition of the parametric sub-models and priors used in our setup.

F.0.1. *Parametric sub-models and priors on tangent spaces.* Following Van der Vaart (2000), we define one-dimensional parametric sub-models,$\{P_{t,\boldsymbol{h}} : t \le \eta\}$, to be the class of probability densities such that

$$\int \left[\frac{\left(dP_{t,\boldsymbol{h}}^{1/2} - dP_0^{1/2}\right)}{t} - \frac{1}{2}\boldsymbol{h}dP_0^{1/2}\right]^2 d\nu \to 0 \text{ as } t \to 0, \tag{F.1}$$

for some measure function $\boldsymbol{h}(\cdot)$. It is well known, see e.g., Van der Vaart (2000), that (F.1) implies $\int \boldsymbol{h}dP_0 = 0$ and $\int \boldsymbol{h}^2 dP_0 < \infty$. As mentioned in the main text, the set of all such candidate h is termed the tangent space $T(P_0)$. This is a subset of the Hilbert space $L^2(P_0)$, endowed with the inner product $\langle f, g\rangle = \mathbb{E}_{P_0}[fg]$ and norm $\|f\| = \mathbb{E}_{P_0}[f^2]^{1/2}$. As in Section 5, (F.1) implies the SLAN property that for all $\boldsymbol{h} \in T(P_0)$,

$$\sum_{i=1}^{\lfloor nq\rfloor} \ln \frac{dP_{1/\sqrt{n},\boldsymbol{h}}}{dP_0}(Y_i) = \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nq\rfloor} \boldsymbol{h}(Y_i) - \frac{q}{2}\|\boldsymbol{h}\|^2 + o_{P_0}(1), \quad \text{uniformly over } q. \tag{F.2}$$

Asymptotic Bayes risk is defined in terms of priors on the tangent space $T(P_0)$. To define this formally, we start by selecting $\{\phi_1, \phi_2, \dots\} \in T(P_0)$ such that $\{\psi/\sigma, \phi_1, \phi_2, \dots\}$ form an orthonormal basis for the closure of $T(P_0)$; the division of ψ by σ is simply to ensure $\|\psi/\sigma\|^2 = \int x^2/\sigma^2 dP_0(x) = 1$. By the Hilbert space isometry, each $\boldsymbol{h} \in T(P_0)$ can then be associated with an element from the l_2 space of square integrable sequences, $(h_0/\sigma, h_1, \dots)$, where $h_0 = \langle \psi, \boldsymbol{h}\rangle$ and $h_k = \langle \phi_k, \boldsymbol{h}\rangle$ for all $k \neq 0$. A prior on $T(P_0)$ therefore corresponds to a prior on l_2.

Let $(\varrho(1), \varrho(2), \dots)$ denote an arbitrary permutation of $(1, 2, \dots)$. As mentioned in the main text, we impose two restriction on ρ_0. The first is that ρ_0 is supported on a finite dimensional sub-space,

$$\mathcal{H}_I \equiv \left\{\boldsymbol{h} \in T(P_0) : \boldsymbol{h} = \frac{1}{\sigma}\langle \psi, \boldsymbol{h}\rangle \frac{\psi}{\sigma} + \sum_{k=1}^{I-1}\left\langle \phi_{\varrho(k)}, \boldsymbol{h}\right\rangle \phi_{\varrho(k)}\right\}$$

of $T(P_0)$, or equivalently, on a subset of l_2 of finite dimension I. Crucially, the first component of $\boldsymbol{h} \in l_2$, corresponding to h_0/σ, is always included in the support of the prior. This important as $h_0 = \langle \psi, \boldsymbol{h}\rangle$ is exactly the mean reward (upto a $\sqrt{n}$ scaling). The second restriction is that it is possible to decompose $\rho_0 = m_0 \times \lambda$, where m_0 is a prior on h_0 and λ is a prior on $(h_{\varrho(1)}, h_{\varrho(2)}, \dots)$. Recall that $\mu_n(\boldsymbol{h}) := \mu(P_{1/\sqrt{n},\boldsymbol{h}}) \approx h_0/\sqrt{n}$. Thus m_0 is effectively equivalent to a prior on the scaled rewards $\sqrt{n}\mu_n$, just as in Section 2.

F.0.2. *Heuristics.* We now provide an informal account of why the second component, λ, of the product prior $\rho_0 := m_0 \times \lambda$ does not feature in asymptotics and it is sufficient, asymptotically, to restrict the state variables to x_{nq}, q, t.

By construction, the prior ρ_0 is supported on a finite-dimensional subset of the tangent space of the form $\left\{\boldsymbol{h}^\intercal \boldsymbol{\chi}(Y_i) : \boldsymbol{h} \in \mathbb{R}^I\right\}$, where $\boldsymbol{\chi} := (\psi/\sigma, \phi_{\varrho(1)}, \dots, \phi_{\varrho(I-1)})$. In what follows, we drop the permutation ϱ for simplicity. Consider the posterior density, $p_n(\cdot|\mathcal{F}_t)$, of the vector $\boldsymbol{h}$ given $\mathcal{F}_t$, where the filtration $\mathcal{F}_t$ is defined as in Section 5. By Lemma 1,

$$p_n(\cdot|\mathcal{F}_t) = p_n(\cdot|\mathbf{y}_{nq(t)}) \propto \left\{ \prod_{i=1}^{\lfloor nq(t) \rfloor} dP_{1/\sqrt{n},\boldsymbol{h}^\intercal \boldsymbol{\chi}}(Y_i) \right\} \cdot \rho_0(\boldsymbol{h}). \tag{F.3}$$

Here, as before, $q(t) = n^{-1} \sum_{j=1}^{\lfloor nt \rfloor} \mathbb{I}(A_j = 1)$. Now, (F.2) suggests that the likelihood term in (F.3) can be approximated by a new likelihood, the density of the 'tilted' measure $\Lambda_{nq,\boldsymbol{h}}(\cdot)$ defined as

$$d\Lambda_{nq,\boldsymbol{h}}(\mathbf{y}_{nq}) := \exp\left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nq \rfloor} \boldsymbol{h}^\intercal \boldsymbol{\chi}(Y_i) - \frac{q}{2} \left\| \boldsymbol{h} \right\|^2 \right\} dP_{1/\sqrt{n},0}(\mathbf{y}_{nq}). \tag{F.4}$$

Le $\boldsymbol{\chi}_{nq} := n^{-1/2} \sum_{i=1}^{\lfloor nq \rfloor} \boldsymbol{\chi}(Y_i)$. Then, taking $\tilde{p}_n(\cdot|\mathbf{y}_{nq})$ to be the corresponding approximate posterior density as in Section 5, we have:

$$\begin{aligned} \tilde{p}_n(\boldsymbol{h}|\mathbf{y}_{nq}) &\propto d\Lambda_{nq,\boldsymbol{h}}(\mathbf{y}_{nq}) \cdot \rho_0(\boldsymbol{h}) \\ &\propto \tilde{p}_q(\boldsymbol{\chi}_{nq}|\boldsymbol{h}) \cdot \rho_0(\boldsymbol{h}); \text{ where } \tilde{p}_q(\cdot|\boldsymbol{h}) \equiv \mathcal{N}(\cdot|q\boldsymbol{h}, qI). \end{aligned} \tag{F.5}$$

The approximate posterior of $\boldsymbol{h}$ depends on the I dimensional quantity $\boldsymbol{\chi}_{nq}$. However, it is possible to achieve achieve further dimension reduction for the marginal posterior density, $\tilde{p}_n(h_0|\mathbf{y}_{nq})$, of h_0. Indeed, for any $\boldsymbol{h} \in T(P_0)$,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nq \rfloor} \boldsymbol{h}(Y_i) - \frac{q}{2} \left\| \boldsymbol{h} \right\|^2 = \frac{h_0}{\sigma\sqrt{n}} \sum_{i=1}^{\lfloor nq \rfloor} Y_i - \frac{q}{2\sigma^2} h_0^2 + (\text{terms independent of } h_0)$$

where the equality follows from the Hilbert space isometry which implies $\boldsymbol{h} = (h_0/\sigma)(\psi/\sigma) + \sum_{k=1}^{I} h_k \phi_k$, and $\left\| \boldsymbol{h} \right\|^2 = (h_0/\sigma)^2 + \sum_{k=1}^{I} h_k^2$. So, defining $x_{nq} = n^{-1/2} \sum_{i=1}^{\lfloor nq \rfloor} Y_i$, we obtain from (F.4) and (F.5) that

$$\begin{aligned} \tilde{p}_n(h_0|\mathbf{y}_{nq}) &\propto \exp\left\{ \frac{h_0}{\sigma^2} x_{nq} - \frac{q}{2\sigma^2} h_0^2 \right\} \cdot m_0(h_0) \\ &\propto \tilde{p}_q(x_{nq}|h_0) \cdot m_0(h_0), \quad \text{where} \quad \tilde{p}_q(\cdot|h_0) \equiv \mathcal{N}(\cdot|qh_0, q\sigma^2). \end{aligned} \tag{F.6}$$

In other words, one can approximate the posterior distribution of h_0 under $\mathcal{F}_t$ by $\tilde{p}_n(h_0|x_{nq(t)}, q(t)) \equiv \tilde{p}_n(h_0|\mathbf{y}_{nq(t)}) \propto p_{q(t)}(x_{nq(t)}|h_0) \cdot m_0(h_0)$, just as in Section 5. Since the expected reward depends only on h_0 due to (6.1), this suggests that it is sufficient, asymptotically, to restrict the state variables to $x_{nq(t)}, q(t), t$.

F.0.3. *Assumptions.* Set $\tilde{\mathbb{E}}[\cdot|s]$ to be the expectation under $\tilde{p}_n(h_0|x,q)$, $\mu^+(s) := \tilde{\mathbb{E}}\left[h_0\mathbb{I}\{h_0>0\}|s\right]$ and $\mu(s) := \tilde{\mathbb{E}}[h_0|s]$. Note that by (F.6), these terms are the same as in Section 2.2.2. Also, set $h(\boldsymbol{\chi}_{nq},q) := \tilde{\mathbb{E}}\left[\boldsymbol{h}|\boldsymbol{\chi}_{nq},q\right]$ where $\tilde{\mathbb{E}}[\cdot|\boldsymbol{\chi}_{nq},q]$ is the expectation under $\tilde{p}_n(\boldsymbol{h}|\boldsymbol{\chi}_{nq},q)$, defined in (F.5). We employ the following assumptions for Theorem 6:

Assumption 2. *(i) The sub-models $\{P_{t,h};h\in T(P_0)\}$ satisfy (F.1). (ii) $\mathbb{E}_{P_0}[|Y|^3]<\infty$. (iii) There exists $\delta_n\to 0$ such that $\sqrt{n}\mu(P_{1/\sqrt{n},\boldsymbol{h}}))=h_0+\delta_n\left\Vert\boldsymbol{h}\right\Vert^2\ \forall\ \boldsymbol{h}\in T(P_0)$. (iv) $\rho_0(\cdot)$ is supported on $\mathcal{H}_I(\Gamma)\equiv\{\boldsymbol{h}\in\mathcal{H}_{\mathcal{I}}:\mathbb{E}_{P_0}\left[\exp|\boldsymbol{h}|\right]\le\Gamma\}$ for some $\Gamma<\infty$. (v) $\mu(\cdot)$ and $\mu^+(\cdot)$ are Hölder continuous and $\sup_s\varpi(s)\le C<\infty$. Furthermore, $h(\boldsymbol{\chi},q)$ is also Hölder continuous.*

Assumption 2(iii) is a stronger version of (6.1), but is satisfied for all commonly used sub-models. For instance, if $dP_{1/\sqrt{n},\boldsymbol{h}} := (1+n^{-1/2}\boldsymbol{h})dP_0$ as in Van der Vaart (2000, Example 25.16), $\sqrt{n}\mu(P_{1/\sqrt{n},\boldsymbol{h}})=\langle\psi,\boldsymbol{h}\rangle=h_0$. Assumption 2(iv) requires the prior to be supported on score functions with finite exponential moments. As with Assumptions 1(ii) & 1(iv), it ensures the tilt $d\Lambda_{nq,\boldsymbol{h}}(\mathbf{y}_{nq})/dP_{1/\sqrt{n},0}(\mathbf{y}_{nq})$ in (F.4) is uniformly bounded. It is somewhat restrictive as it implies $\mathbb{E}_{P_0}[\exp|h_0Y|]<\infty$ for all $h_0\in\text{supp}(m_0)$. However, similar to Assumptions 1(ii) & 1(iv), we suspect it can be relaxed at the expense of more intricate proofs. Finally, Assumption 2(v) differs from Assumption 1(v) only in requiring continuity of $h(\boldsymbol{\chi},q)$. While $h(\boldsymbol{\chi},q)$ is not present in PDE (2.8), it arises in the course of various PDE approximations in the proof. The form of the posterior in (F.5) implies this should be satisfied under mild assumptions on ρ_0. It is certainly satisfied for Gaussian ρ_0.

F.0.4. *Proof of Theorem 6.* The proof consists of two steps. First, we show that $V_n^*(0)$ converges to the solution of a PDE with state variables $(\boldsymbol{\chi},q,t)$ where $\boldsymbol{\chi}(t) := \boldsymbol{\chi}_{nq(t)}$ with $\boldsymbol{\chi}_{nq}$ defined in Section 6. Recall that the first component of $\boldsymbol{\chi}$ is x/σ. Next, we show that the PDE derived in the first step can be reduced to one involving just the state variables $s=(x,q,t)$.

The first step follows the proof of Theorem 5 with straightforward modifications. Indeed, the setup is equivalent to taking $\boldsymbol{\chi}(Y_i)$ to be the vector-valued score function in the parametric setting (see, Section 5.3). The upshot of these arguments is that $V_n^*(0)$ converges to $V^*(0)$, where $V^*(\cdot)$ solves the PDE

$$\partial_t f(\boldsymbol{\chi},q,t)+\mu^+(x,q)+\min\left\{-\mu(x,q)+\bar{L}[f](\boldsymbol{\chi},q,t),0\right\}=0 \text{ if } t<1 \qquad \text{(F.7)}$$
$$f(\boldsymbol{\chi},q,t)=0 \text{ if } t=1,$$

with the infinitesimal generator (here $\triangle$ denotes the Laplace operator)

$$\bar{L}[f](\boldsymbol{\chi},q,t):=\partial_q f+h(\boldsymbol{\chi},q)^{\intercal}D_{\boldsymbol{\chi}}f+\frac{1}{2}\triangle_{\boldsymbol{\chi}}f.$$

See Section 6 for the definition of $h(\boldsymbol{\chi},q)$. Note that $\mu^+(\cdot),\mu(\cdot)$ are functions only of (x,q). This is because they depend only on the first component, h_0/σ, of $\boldsymbol{h}$ and its posterior distribution can be approximated by $\tilde{p}_n(h_0|x,q)$, defined in (F.6).

By the arguments leading to (F.6), the first component of the vector $h(\boldsymbol{\chi},q)$ is $\sigma^{-1}\tilde{\mathbb{E}}[h_0|\boldsymbol{\chi},q]=\sigma^{-1}\tilde{\mathbb{E}}[h_0|x,q]=\sigma^{-1}\mu(x,q)$. Let $\boldsymbol{\chi}^c$, $h^c(\boldsymbol{\chi},q)$ denote $\boldsymbol{\chi},h(\boldsymbol{\chi},q)$ without their first components $\chi_1=x/\sigma$ and $h_1(\boldsymbol{\chi},q)=\sigma^{-1}\mu(x,q)$. Then, defining

$$L[f](x,q,t):=\partial_q f+\mu(x,q)\partial_x f+\frac{1}{2}\sigma^2\partial_x^2 f,$$

we see that $\bar{L}[f]=L[f]+h^c(\boldsymbol{\chi},q)^{\intercal}D_{\boldsymbol{\chi}^c}f+\frac{1}{2}\triangle_{\boldsymbol{\chi}^c}f$. Note that in defining $L[f](\cdot)$, we made use of the change of variables $\partial_{\chi_1}f=\sigma\partial_x f$ and $\partial_{\chi_1}^2 f=\sigma^2\partial_x^2 f$. We now claim that the solution of PDE (F.7) is the same as that of PDE (2.8), reproduced here:

$$\partial_t f(x,q,t)+\mu^+(x,q)+\min\left\{-\mu(x,q)+L[f](x,q,t),0\right\}=0 \text{ if } t<1 \qquad \text{(F.8)}$$
$$f(x,q,t)=0 \text{ if } t=1.$$

Intuitively, this is because the state variables in $\boldsymbol{\chi}^c$ do not affect instantaneous payoffs $\mu^+(x,q)-\mu(x,q),\mu^+(x,q)$, nor do they affect the boundary condition, so these state variables are superfluous. The formal proof makes use of the theory of viscosity solutions: Under Assumption 2(v), Theorem 1 implies there exists a unique viscosity solution to (F.7), denoted by $V^*(\boldsymbol{\chi},q,t)$. Then, it is straightforward to show that $\bar{V}^*(x,q,t)=\sup_{\boldsymbol{\chi}^c}V^*(\boldsymbol{\chi},q,t)$ is a viscosity sub-solution to (F.8).[20] In a

[20] See Crandall et al. (1992) for the definition of viscosity sub- and super-solutions using test functions. To show $\bar{V}^*$ is a sub-solution one can argue as follows: First, $\bar{V}^*(x,q,t)$ is upper-semicontinuous because of the continuity of the solution $V^*(\boldsymbol{\chi},q,t)$ to PDE (F.7). Second, $\bar{V}^*$ satisfies the boundary condition in PDE (F.8) by construction. Third, let $\phi\in\mathcal{C}^\infty(\mathcal{X},\mathcal{Q},\mathcal{T})$ denote a test function such that $\phi\geq\bar{V}^*$ everywhere. By the definition of $\bar{V}^*$ we also have

similar fashion, $\underline{V}^*(x,q,t) = \inf_{\boldsymbol{\chi}^c} V^*(\boldsymbol{\chi},q,t)$ is a viscosity super-solution to (F.8). Under Assumption 2(v), a comparison principle (see, Crandall et al., 1992) holds for (F.8) implying any super-solution is larger than a solution, which is in turn larger than a sub-solution. But $\bar{V}^*(x,q,t) \geq \underline{V}^*(x,q,t)$ by definition, so it must be the case $\bar{V}^*(x,q,t) = \underline{V}^*(x,q,t) = V^*(x,q,t)$, where $V^*(x,q,t)$ is the unique viscosity solution to (F.8). This proves $V^*(\boldsymbol{\chi},q,t) = V^*(x,q,t)$, as claimed.

Appendix G. Theory for MAB and its generalizations

G.1. Multi-armed bandits.

Existence of a solution to PDE (2.7). By Barles and Jakobsen (2007, Theorem A.1), there exists a unique viscosity solution to PDE (2.7) if $\mu^{\max}(\cdot)$ and $\mu_k(\cdot)$ are Hölder continuous for all k.

Convergence to the PDE. Let $V_n^*(\cdot)$ denote the minimal Bayes risk function in the Gaussian setting. The following analogue of Theorem 2 can then be shown with a straightforward modification to the proof:

Theorem 7. *Suppose $\mu(\cdot)$ and $\mu^{\max}(\cdot)$ are Hölder continuous and the prior m_0 is such that $\mathbb{E}[|\mu|^3|s] < \infty$ at each s. Then, as $n \to \infty$, $V_n^*(\cdot)$ converges locally uniformly to $V^*(\cdot)$, the unique viscosity solution of PDE (2.7).*

Piece-wise constant policies. The construction of piece-wise constant policies in the multi-armed setting is analogous to Section 3.3. Following Barles and Jakobsen (2007, Theorem 3.1), Theorems 3 and 4 can be shown to hold under Lipschitz continuity of $\mu^{\max}(\cdot), \mu_k(s)$ and $\sup_s \{\mu^{\max}(s) - \max_k \mu(s)\} < \infty$.

Parametric and non-parametric distributions. Let $P_\theta^{(k)}$ denote the probability distribution over the rewards from arm k. It is without loss of generality to assume the distributions across arms are independent of each other as we only ever observe the outcomes from a single arm. The parameter $\theta \in \mathbb{R}^d$ may have some components that are shared across all the arms. As in the one-armed bandit setting, we

$\phi(x,q,t) \geq V^*(\boldsymbol{\chi},q,t)$ everywhere. Since $V^*(\boldsymbol{\chi},q,t)$ is a solution to PDE (F.7), ϕ must satisfy the viscosity requirement for a sub-solution to PDE (F.7). But because ϕ is constant in $\boldsymbol{\chi}^c$, this implies it also satisfies the viscosity requirement for a sub-solution to PDE (F.8). These three facts suffice to show $\bar{V}^*$ is a sub-solution.

choose a reference θ_0 such that $\mathbb{E}_{P_{\theta_0}^{(k)}}[Y_k] = 0$, and focus on local perturbations of the form $\{\theta_{n,h} \equiv \theta_0 + h/\sqrt{n} : h \in \mathbb{R}^d\}$. We then place a non-negligible prior M_0 on the local parameter h.

To simplifty notation, suppose that θ is scalar. Let $\nu := \nu_1 \times \nu_2$, where ν_1 is a dominating measure for $\{P_\theta^{(k)} : \theta \in \mathbb{R}, k = 0, \dots, K-1\}$ and ν_2 is a dominating measure for the prior M_0 on h. Define $p_\theta^{(k)} = dP_\theta^{(k)}/d\nu$, $m_0 = dM_0/d\nu$ (in the sequel, we shorten the Radon-Nikodym derivative $dP/d\nu$ to just dP). As in Section 5, we require the class $\{P_\theta^{(k)}\}$ to be quadratic mean differentiable (q.m.d) around θ_0 for each k. This in turn implies the SLAN property that, for each k,

$$\sum_{i=1}^{\lfloor nq_k \rfloor} \ln \frac{dp_{\theta_0+h/\sqrt{n}}^{(k)}}{dp_{\theta_0}^{(k)}} = \frac{1}{\sigma_k^2} h x_{k,nq_k} - \frac{q_k}{2\sigma_k^2} h^2 + o_{P_{n,\theta_0}^{(k)}}(1), \text{ uniformly over } q_k, \tag{G.1}$$

where

$$x_{k,nq} := \sigma_k^2 \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nq \rfloor} \psi_k(Y_i^{(k)}),$$

$\psi_k(\cdot)$ is the score function corresponding to $P_{\theta_0}^{(k)}$, and σ_k^2 is the corresponding inverse information matrix, i.e., $\sigma_k^2 = \left(\mathbb{E}_{P_{\theta_0}^{(k)}}[\psi_k^2]\right)^{-1}$.

Recall that $\mathbf{y}_n^{(k)} := (Y_1^{(k)}, \dots, Y_n^{(k)})$ denotes the vector of stacked outcomes for each arm k. Then, in the fixed n setting, the posterior distribution of h is (compare the equation below with (5.3))

$$p_n(h|\mathcal{F}_t) = p_n\left(h\middle|\left\{\mathbf{y}_{nq_k(t)}^{(k)}\right\}_k\right) \propto \left[\prod_{k=0}^{K-1} \prod_{i=1}^{\lfloor nq_k(t) \rfloor} p_{\theta_0+h/\sqrt{n}}^{(k)}(Y_i^{(k)})\right] \cdot m_0(h).$$

As in Section 5, we approximate the likelihood (the bracketed term in the above expression) with an approximation implied by (G.1). So, the approximate posterior is

$$\tilde{p}_n(h|s) \propto \left[\prod_{k=0}^{K-1} \tilde{p}_{q_k}(x_k|h)\right] \cdot m_0(h); \text{ where } \tilde{p}_{q_k}(\cdot|h) \equiv \mathcal{N}(\cdot|q_k h, q_k \sigma_k^2). \tag{G.2}$$

The above suggests Theorem 5 can be extended to the K armed case. This is done under the following assumptions: Define $\mu_n^{(k)}(h) = \mathbb{E}_{P_{\theta_0+h/\sqrt{n}}^{(k)}}\left[Y_i^{(k)}\right]$.

Assumption 3. *(i) The class $\{P_\theta^{(k)}\}$ is q.m.d around θ_0 for each k. (ii) $\mathbb{E}_{P_{\theta_0}^{(k)}}[\exp|\psi_k(Y)|] < \infty$ for each k. (iii) For each k, there exists $\dot{\mu}_0^{(k)} < \infty$ such that $\sqrt{n}\mu_n^{(k)}(h) = \dot{\mu}_0^{(k)} h + o(|h|^2)$. (iv) The support of $m_0(\cdot)$ is a compact set $\{h : |h| \leq \Gamma\}$ for some $\Gamma < \infty$.*

(v) $\mu(\cdot)$ *and* $\mu^{\max}(\cdot)$ *are Hölder continuous. Additionally,* $\sup_s \{\mu^{\max}(s) - \max_k \mu(s)\} \leq C < \infty$*.*

Let $V_{\pi,n}(\cdot)$ denote the Bayes risk of policy π and $V_n^*(\cdot)$ the minimal Bayes risk, both under fixed n. Define $\Pi^{\mathcal{S}}$ as the class of all sequentially measurable policies that are functions only of $s = \{\{x_k, q_k\}_k, t\}$, and $V_n^{\mathcal{S}*}(0)$ the fixed n minimal Bayes risk when the policies are restricted to $\Pi^{\mathcal{S}}$. Also, take $\pi^*_{\Delta t}$ to be the optimal piecewise constant policy with Δt increments. Finally, denote by $L_k[\cdot]$ the infinitesimal generator

$$L_k[f] := \partial_{q_k} f + h(s)\partial_{x_k} f + \frac{1}{2}\sigma_k^2 \partial_{x_k}^2 f, \tag{G.3}$$

where $h(s) := \tilde{\mathbb{E}}[h|s]$ and $\tilde{\mathbb{E}}[\cdot|s]$ is the expectation under $\tilde{p}_n(\cdot|s)$, defined in (G.2).

Theorem 8. *Suppose that Assumption 3 holds. Then: (i)* $\lim_{n\to\infty} \left| V_n^*(0) - V_n^{\mathcal{S}*}(0) \right| = 0$*. (ii)* $\lim_{n\to\infty} V_n^*(0) = V^*(0)$*, where* $V^*(\cdot)$ *solves PDE (2.7) with the infinitesimal generators given by (G.3). (iii) If, further,* $\mu(\cdot)$*,* $\mu^{\max}(\cdot)$ *are Lipschitz continuous,* $\lim_{n\to\infty} |V_{\pi^*_{\Delta t},n}(0) - V^*(0)| \lesssim \Delta t^{1/4}$ *for any fixed* Δt*.*

The proof is analogous to that of Theorem 5, with the key difference being that the relevant likelihood is

$$\prod_{k=0}^{K-1} \prod_{i=1}^{\lfloor nq_k(t) \rfloor} p^{(k)}_{\theta_0 + h/\sqrt{n}}(Y_i^{(k)})$$

instead of $\prod_{i=1}^{\lfloor nq(t) \rfloor} p_{\theta_0 + h/\sqrt{n}}(Y_i)$. The independence of the reward distributions across arms is convenient here, and helps simplify the proof.[21] See Adusumilli (2022b) for an example of the formal argument.

Similar adaptations can be made for the results in Section 6.

G.2. **Best arm identification.** Best arm identification describes a class of sequential experiments in which the DM is allowed to experiment among K arms of a bandit until a set time $t = 1$ (corresponding to n time periods). At the end of the experimentation phase, an arm is selected for final implementation. Statistical loss is determined by expected payoffs during the implementation phase,

[21] For instance, it implies that the joint probability $\prod_{k=0}^{K-1} P^{(k)}_{nq_{nk},h}$ is contiguous to $\prod_{k=0}^{K-1} P^{(k)}_{nq_k,0}$ for any $(q_{n0}, \ldots, q_{n(K-1)}) \to (q_0, \ldots, q_K)$ as $n \to \infty$, as long as $P^{(k)}_{nq_{nk},h}$ is contiguous to $P^{(k)}_{nq_k,0}$ for each k. This enables us to prove an analogue to Lemma 3, which is a key step in the proof.

but not on payoffs generated during experimentation, i.e., there is no exploitation motive. In the Gaussian setting, it is sufficient to use the same state variables $s = \{\{x_k, q_k\}_k, t\}$ as in K armed bandits.

Let $\boldsymbol{\mu} := (\mu_0, \dots, \mu_{K-1})$ denote the mean rewards of each arm, and $\pi^{(I)} \in \{0, \dots, K-1\}$ the action of the DM in the implementation phase. Following the best arm identification literature, see, e.g., Kasy and Sautmann (2021), we take the loss function to be expected regret in the implementation phase (also known as "simple regret")

$$L(\pi^{(I)}, \boldsymbol{\mu}) = \max_k \mu_k - \sum_k \mu_k \mathbb{I}(\pi^{(I)} = k).$$

Suppose that the state variable at the end of experimentation is s. The Bayes risk of policy $\pi^{(I)}$ given the terminal state s is

$$V_{\pi^{(I)}}(s) = \mathbb{E}\left[L(\pi^{(I)}, \boldsymbol{\mu})|s\right] = \mu^{\max}(s) - \sum_k \mu_k(s)\mathbb{I}(\pi^{(I)} = k).$$

Hence, the optimal Bayes policy is $\pi^{(I)} = \arg\max_k \mu_k(s)$ and the minimal Bayes risk at the end of experimentation, i.e., when $t = 1$, is $V^*(s) = \mu^{\max}(s) - \max_k \mu_k(s)$. This determines the boundary condition at $t = 1$.

We can obtain a PDE characterization of $V^*(\cdot)$ through similar heuristics as in Section 2.2. By (2.1), the change to q_k and x_k in a short time period Δt following state s is approximately

$$\Delta q_k \approx \pi_k \Delta t; \quad \Delta x_k \approx \pi_k \mu_k \Delta t + \sigma_k \sqrt{\pi_k} \Delta W(t).$$

Now, for 'interior states' with $t < 1$, the recursion

$$V^*(s) = \inf_{\pi \in [0,1]^K} \mathbb{E}\left[V^*\left(\{x_k + \Delta x_k, q_k + \Delta q_k\}_k, t + \Delta t\right)\middle| s\right]$$

must hold for any small time increment Δt. Thus, by similar (heuristic) arguments as in Section 2.2, $V^*(\cdot)$ satisfies

$$\partial_t V^* + \min_k L_k[V^*](s) = 0 \qquad \text{if } t < 1; \tag{G.4}$$
$$V^*(s) = \varpi(s) \text{ if } t = 1,$$

where $\varpi(s) := \mu^{\max}(s) - \max_k \mu_k(s)$.

As we show below, all previous theoretical results (including for parametric and non-parametric models) continue to apply with minor modifications to the statements and the proofs. See also Adusumilli (2022a) for the derivation of the minimax optimal policy in the two arm case. The assumptions required are the same as that for multi-armed bandits.

Existence of a solution to PDE (G.4). This is again a direct consequence of Barles and Jakobsen (2007, Theorem A.1).

Convergence to the PDE. Recall that the relevant state variables are $s = \{\{x_k, q_k\}_k, t\}$. In analogy with (3.1), the Bayes risk in the fixed n setting is given by

$$V_n^*(x_1, q_1, \dots, x_K, q_K, t) = \mathbb{I}_n^c \cdot \varpi(s) + \dots$$
$$\dots + \min_{\pi_1,\dots,\pi_K \in [0,1]} \mathbb{E}\left[\mathbb{I}_n \cdot V_n^*\left(\left\{x_k + \frac{\pi_k Y_{nq_k+1}^{(k)}}{\sqrt{n}}, q_k + \frac{\pi_k}{n}\right\}_k, t + \frac{1}{n}\right)\middle| s\right] \tag{G.5}$$

where $\mathbb{I}_n := \mathbb{I}\{t \geq 1/n\}$. The solution, $V_n^*(\cdot)$, of the above converges locally uniformly to the viscosity solution, $V^*(\cdot)$, of PDE (G.4). We can show this by modifying the proof of Theorem 2 to account for the non-zero boundary condition. As in that proof, after a change of variables $\tau = 1 - t$, we can characterize $V_n^*(\cdot)$ as the solution to $S_n(s, \phi(s), [\phi]) = 0$, where for any $u \in \mathbb{R}$ and $\phi : \mathcal{S} \to \mathbb{R}$, and $\mathbb{I}_n := \mathbb{I}\{\tau > 1/n\}$,

$$S_n(s, u, [\phi]) := -\mathbb{I}_n^c \cdot \frac{(\varpi(s) - u)}{n} - \dots$$
$$\dots - \mathbb{I}_n \cdot \min_{\pi_1,\dots,\pi_K \in [0,1]} \mathbb{E}\left[\phi\left(\left\{x_k + \frac{\pi_k Y_{nq_k+1}^{(k)}}{\sqrt{n}}, q_k + \frac{\pi_k}{n}\right\}_k, \tau - \frac{1}{n}\right) - u \middle| s\right].$$

Define $F(D^2\phi, D\phi, s) = \partial_\tau \phi - \min_k L_k[\phi](s)$.

We need to verify monotonicity, stability and consistency of $S_n(\cdot)$. Monotonicity of $S_n(s, u, [\phi])$ is clearly satisfied. Stability is also straightforward under the assumption $\sup_s \varpi(s) < \infty$. The consistency requirement is more subtle. For interior values, i.e., when $s := (x, q, \tau)$ is such that $\tau > 0$, the usual conditions (A.3) and (A.4) are required to hold with the definitions of $S_n(\cdot), F(\cdot)$ above. These can be shown using the same Taylor expansion arguments as in the proof of Theorem 2. For boundary values, $s \in \partial\mathcal{S} \equiv \{(x, q, 0) : x \in \mathcal{X}, q \in [0, 1]\}$, the consistency requirements are (see, Barles and Souganidis, 1991)

$$\limsup_{\substack{n\to\infty\\ \rho\to 0\\ z\to s\in\partial\mathcal{S}}} nS_n(z,\phi(z)+\rho,[\phi+\rho]) \le \max\left\{F(D^2\phi(s),D\phi(s),s),\phi(s)-\varpi(s)\right\}, \tag{G.6}$$

$$\liminf_{\substack{n\to\infty\\ \rho\to 0\\ z\to s\in\partial\mathcal{S}}} nS_n(z,\phi(z)+\rho,[\phi+\rho]) \ge \min\left\{F(D^2\phi(s),D\phi(s),s),\phi(s)-\varpi(s)\right\}. \tag{G.7}$$

We can show (G.6) as follows (the proof of (G.7) is similar): By the definition of $S_n(\cdot)$, for every sequence $(n\to\infty,\rho\to 0,z\to s\in\partial\mathcal{S})$, there exists a sub-sequence such that either $nS_n(z,\phi(z)+\rho,[\phi+\rho])=\phi+\rho-\varpi(z)$ or

$$nS_n(z,\phi(z)+\rho,[\phi+\rho]) = -\min_{\pi_1,\dots,\pi_K\in[0,1]} \mathbb{E}\left[\phi\left(\left\{x_k+\frac{\pi_k Y^{(k)}_{nq_k+1}}{\sqrt{n}},q_k+\frac{\pi_k}{n}\right\}_k,\tau-\frac{1}{n}\right)-u\middle|s\right].$$

In the first instance, $nS_n(z,\phi(z)+\rho,[\phi+\rho])\to\phi(s)-\varpi(s)$ by the continuity of $\varpi(\cdot)$, while the second instance gives rise to the same expression for $S_n(\cdot)$ as being in the interior, so that $nS_n(z,\phi(z)+\rho,[\phi+\rho])\to F(D^2\phi(s),D\phi(s),s)$ by similar arguments as in the proof of Theorem 2. Thus, in all cases, the limit along subsequences is smaller than the right hand side of (G.6).

Piecewise-constant policies. The results on piece-wise constant policies continue to apply since Barles and Jakobsen (2007, Theorem 3.1) holds under any continuous boundary condition.

Parametric and non-parametric distributions. The analogues of Theorems 5 and 6 follow by the same reasoning as that employed for multi-armed bandits in Appendix G.1. In fact, the proofs are even simpler since the loss function is just the regret payoff at $t=1$.

G.3. **Discounting.** Our methods also apply to bandit problems without a definite end point. Suppose the rewards in successive periods are discounted by $e^{-\beta/n}$ for some $\beta>0$. Here, n is to be interpreted as a scaling of the discount factor; it is the number of periods of experimentation in unit time when the DM experiments in regular time increments and intends to discount rewards by the fraction $e^{-\beta}$ after $\Delta t=1$. Discounting ensures the cumulative regret is finite. It also changes the considerations of the DM, who will now be impatient to start 'exploitation' sooner as future rewards are discounted. Popular bandit algorithms such as Thompson sampling do not admit discounting and will therefore be substantially sub-optimal.

In the Gaussian setting with one arm, the relevant state variables under discounting are $s := (x, q)$, where x, q are defined in the same manner as before, but q can now take values above 1 (it is the number of times the arm is pulled divided by n). The counterpart of PDE (2.8) for discounted rewards is

$$\beta V^* - \mu^+(s) - \min\{-\mu(s) + L[V^*](s), 0\} = 0. \tag{G.8}$$

Note that PDE (G.8) does not require a boundary condition.

All the previous theoretical results continue to apply to discounted bandits, as we demonstrate below. The assumptions required are the same as in Theorems 1-6 in the main text, along with $\beta > 0$.

Existence of a solution to PDE (G.8). By Barles and Jakobsen (2007, p. 29), there exists a unique viscosity solution to PDE (G.8).

Convergence to the PDE. The analogue to (3.1) under discounting is

$$V_n^*(x, q) = \min_{\pi \in [0,1]} \mathbb{E}\left[\frac{\mu^+(s) - \pi\mu(s)}{n} + e^{-\beta/n} V_n^* \left(x + \frac{A_\pi Y_{nq+1}}{\sqrt{n}}, q + \frac{A_\pi}{n} \right) \middle| s \right]. \tag{G.9}$$

A straightforward modification of the proof of Theorem 2 then shows $V_n^*(\cdot)$ converges locally uniformly to $V^*(\cdot)$, the viscosity solution of PDE (G.8). There is no analogue to piece-wise constant policies in the discounted setting.

Parametric and non-parametric distributions. The proofs of Theorems 5 and 6 are slightly complicated by the fact q is now unbounded. While the SLAN property (5.2) applies even if $q > 1$, it does require $q < \infty$. We can circumvent this issue by exploiting the fact that the infinite horizon problem is equivalent to a finite horizon problem with a very large time limit. In other words, we prove the relevant results for the PDE

$$\begin{aligned} \partial_t V^* - \beta V^* + \mu^+(s) + \min\{-\mu(s) + L[V^*](s), 0\} = 0 \text{ if } t < 1, \\ V^*(s) = 0 \text{ if } t = T, \end{aligned} \tag{G.10}$$

with the boundary condition set at $t = T$, and then let $T \to \infty$.

Let $V^*(0), V^*(0;T)$ denote the viscosity solutions to PDEs (G.8) and (G.10), evaluated at s_0. Following the first step in Appendix (A.3), the Bayes risk under a policy π in the fixed n setting with discounting can be shown to be

$$V_{\pi,n}(0) = \mathbb{E}_{(\mathbf{y}_n,h)} \left[\frac{1}{n} \sum_{j=1}^{\infty} e^{-\beta j/n} R_n(h, \pi_j) \right]. \tag{G.11}$$

Analogously, if we terminate the experiment at a suitably large T, we have

$$V_{\pi,n}(0;T) = \mathbb{E}_{(\mathbf{y}_n,h)} \left[\frac{1}{n} \sum_{j=1}^{nT} e^{-\beta j/n} R_n(h, \pi_j) \right].$$

Under Assumption 1, $R_n(h,\pi) \le C < \infty$ (due to the compactness of the prior m_0), so $\sup_{\pi \in \Pi} |V_{\pi,n}(0) - V_{\pi,n}(0;T)| \lesssim e^{-\beta T}$. Now, a straightforward modification of the proof of Theorem 5 implies $\lim_{n\to\infty} \inf_{\pi\in\Pi} V_{\pi,n}(0;T) = V^*(0;T)$, where $V^*(0;T)$ is the viscosity solution to PDE (G.10) evaluated at s_0. Finally, it can be shown, e.g., by approximating the PDEs with dynamic programming problems as in Theorem 2, that $|V^*(0;T) - V^*(0)| \lesssim e^{-\beta T}$. Since we can choose T as large as we want, it follows $\lim_{n\to\infty} \inf_{\pi\in\Pi} V_{\pi,n}(0) = V_n^*(0)$. The proof of Theorem 6 can be modified in a similar manner.

Appendix H. Computation using finite-difference methods

As mentioned in the main text, PDE (2.8) also be solved using 'upwind' finite-difference methods. The method is more accurate than the Monte-Carlo algorithm (Algorithm 1) but scales less favorably with increasing number of arms. To implement this method we first discretize both the spatial (i.e., $\mathcal{X}$ and $\mathcal{Q}$) and time domains. Let i, j index the grid points for x, q respectively, with the grid lengths being $\Delta x, \Delta q$. PDEs of the form (2.8) are always solved backward in time, so, for this section, we switch the direction of time (i.e., $t = 1$ earlier is now $t = 0$) and discretize it as $0, \Delta t, \dots, m\Delta t, \dots, 1$. Denote $V_{i,j}^m$ as the approximation to the PDE solution V^* at grid points i, j and time period $m\Delta t$.

We approximate the second derivative $\partial_x^2 V^*$ using

$$\partial_x^2 V^* \approx \frac{V_{i+1,j}^m + V_{i-1,j}^m - 2V_{i,j}^m}{(\Delta x)^2}.$$

As for the first order derivatives, we approximate by either $\frac{V^m_{i+1,j}-V^m_{i,j}}{\Delta x}$ or $\frac{V^m_{i,j}-V^m_{i-1,j}}{\Delta x}$ depending on whether the associated drift, i.e., the coefficient multiplying $\partial_x V^*$ is positive or negative. This is known as up-winding and is crucial for ensuring the resulting approximation procedure is 'monotone' (see Appendix A.1, and also Achdou et al. (2022) for a discussion of monotonicity, and its necessity for showing convergence of the approximation procedures). In our setting, this implies

$$\begin{aligned}\partial_x V^* &\approx \frac{V^m_{i+1,j}-V^m_{i,j}}{\Delta x}\mathbb{I}(\mu(s)\geq 0)+\frac{V^m_{i,j}-V^m_{i-1,j}}{\Delta x}\mathbb{I}(\mu(s)<0)\\ &:= \left(\frac{V^m_{i+1,j}-V^m_{i,j}}{\Delta x}\right)_+,\end{aligned}$$

while $\partial_q V^*$, which is associated with the coefficient 1, is approximated as

$$\partial_q V^* \approx \frac{V^m_{i,j+1}-V^m_{i,j}}{\Delta q}.$$

Finally, let $\mu^+_{i,j}, \mu_{i,j}$ denote the values of $\mu^+(\cdot), \mu(\cdot)$ evaluated at the grid points i, j.

Following the derivative approximations, the PDE can be solved using explicit, implicit or hybrid schemes. The previous version of this manuscript discussed these different approaches and their convergence properties.[22] Our recommendation is to use the hybrid scheme. It is faster than the standard implicit scheme as it does not require policy iteration. At the same time, it is more numerically stable than the explicit scheme as it does not require the CFL condition that $\Delta t \leq 0.5\min\{(\Delta x)^2, (\Delta q)^2\}$; instead, we only need $\Delta t \to 0$.

The algorithm is based on a recursion whereby $V^0_{i,j}=0$, and an estimate of the action-value function, $\tilde{V}^{m+1,1}_{i,j}$, corresponding to the case where the arm was pulled in step $m+1$, is computed in terms of $V^m_{i,j} := \min\left\{\tilde{V}^{m,1}_{i,j}, \tilde{V}^{m,0}_{i,j}\right\}$ as the solution to

$$\begin{aligned}\tilde{V}^{m+1,1}_{i,j} &= V^m_{i,j}+\mu^+_{i,j}-\mu_{i,j}+\frac{\tilde{V}^{m+1,1}_{i,j+1}-\tilde{V}^{m+1,1}_{i,j}}{\Delta q}\\ &+\mu_{i,j}\left(\frac{\tilde{V}^{m+1,1}_{i+1,j}-\tilde{V}^{m+1,1}_{i,j}}{\Delta x}\right)_+ +\frac{1}{2}\sigma^2\frac{\tilde{V}^{m+1,1}_{i+1,j}+\tilde{V}^{m+1,1}_{i-1,j}-2\tilde{V}^{m+1,1}_{i,j}}{(\Delta x)^2}=0. \quad \text{(H.1)}\end{aligned}$$

As for the action-value function corresponding to the case where the arm was not pulled, we have

$$\tilde{V}^{m+1,0}_{i,j} := V^m_{i,j}+\mu^+_{i,j}.$$

[22] This version can be accessed at arXiv:2112.06363v14.

We then set $V_{i,j}^{m+1} := \min\left\{\tilde{V}_{i,j}^{m+1,1}, \tilde{V}_{i,j}^{m+1,0}\right\}$ and continue the iterations until $m = M-1$. The pseudo-code for the hybrid FD scheme is described in Algorithm 2.

Algorithm 2 Hybrid FD

Require: M (number of time periods)
1: **initialize** $V_{i,j}^0 = 0$
2: **for** $m = 0, \ldots, M-1$: **do**
3: Write (H.1) as $A\tilde{\mathbf{V}}_{m+1}^1 - \mathbf{V}_m + \mathbf{X} = 0$ where $\tilde{\mathbf{V}}_m^{(1)} = \mathrm{vec}(\tilde{V}_{i,j}^{m,1}; i, j)$
4: $\tilde{\mathbf{V}}_{m+1}^1 = A^{-1}(\mathbf{V}_m - \mathbf{X})$
5: $\tilde{\mathbf{V}}_{m+1}^0 = \mathbf{V}_m + \boldsymbol{\mu}^+$ where $\boldsymbol{\mu}^+ = \mathrm{vec}(\mu_{i,j}^+; i, j)$
6: $\mathbf{V}_{m+1} = \min\left\{\tilde{\mathbf{V}}_{m+1}^1, \tilde{\mathbf{V}}_{m+1}^0\right\}$ where the minimum is computed element-wise
7: **end for**

H.0.1. *Implementation details for Section 4.2.* For the empirical illustration in Section 4.2, we used $\Delta x = 1/1500$, $\Delta q = 1/600$ and $\Delta t = 1/1000$. Since x is unbounded, for the purposes of computation we set its upper and lower bounds to $l - 3\sigma$ and $u + 3\sigma$, where l and u are the support points of the least favorable prior.

References

Y. Achdou, J. Han, J.-M. Lasry, P.-L. Lions, and B. Moll, "Income and wealth distribution in macroeconomics: A continuous-time approach," *The Review of Economic Studies*, vol. 89, no. 1, pp. 45–86, 2022.

K. Adusumilli, "How to sample and when to stop sampling: The generalized Wald problem and minimax policies," *arXiv preprint arXiv:2210.15841*, 2022.

——, "Minimax policies for best arm identification with two arms," *arXiv preprint arXiv:2204.05527*, 2022.

G. Barles and E. Jakobsen, "Error bounds for monotone approximation schemes for parabolic hamilton-jacobi-bellman equations," *Mathematics of Computation*, vol. 76, no. 260, pp. 1861–1893, 2007.

G. Barles and P. E. Souganidis, "Convergence of approximation schemes for fully nonlinear second order equations," *Asymptotic Analysis*, vol. 4, no. 3, pp. 271–283, 1991.

R. F. Bass and R. Pyke, "A strong law of large numbers for partial-sum processes indexed by sets," *The Annals of Probability*, pp. 268–271, 1984.

M. G. Crandall, H. Ishii, and P.-L. Lions, "User's guide to viscosity solutions of second order partial differential equations," *Bulletin of the American Mathematical Society*, vol. 27, no. 1, pp. 1–67, 1992.

M. Kasy and A. Sautmann, "Adaptive treatment assignment in experiments for policy choice," *Econometrica*, vol. 89, no. 1, pp. 113–132, 2021.

L. Le Cam and G. L. Yang, *Asymptotics in Statistics: Some basic concepts.* Springer Science & Business Media, 2000.

A. W. Van der Vaart, *Asymptotic statistics.* Cambridge university press, 2000.

A. W. Van Der Vaart and J. Wellner, *Weak convergence and empirical processes: with applications to statistics.* Springer Science & Business Media, 1996.