

Information-Filtration Trade-Off in Thompson Sampling for MNL Bandits

Iván Vidal-Romero, Denis Sauré

Department of Industrial Engineering, University of Chile,
Santiago, Chile, ivan.vidal@ug.uchile.cl, dsauré@uchile.cl

Learning in dynamic assortment planning under multinomial logit (MNL) demand couples revenue and information through the denominator shared by every displayed product, which fixes what an assortment earns and what each choice reveals. We study Thompson sampling, where every decision optimizes an attraction vector drawn from the representation it maintains. That representation is a design choice: filtering the likelihood to product-versus-outside comparisons gives one conjugate law per product but discards the substitution evidence coupling the attractions, which the complete likelihood keeps at a cost. We analyze both ends of this information-filtration trade-off, and one approximation in between. At the filtered end, the probability that several independent draws are optimistic at once can decay exponentially in the display capacity K ; existing analyses avoid this by correlating, resampling, or inflating draws. None of that is necessary: the linear-fractional form of MNL revenue tests whether correcting one omitted draw would improve the sampled optimum; with scalar posterior odds and per-product clocks, coordinate repair gives gap-free frequentist regret $\tilde{O}(\sqrt{NKT})$ for independent BetaPrime draws, without resampling or variance inflation. At the complete end, a Gamma-Exponential augmentation yields a Gibbs sampler for the joint posterior, and it attains Bayesian regret $\tilde{O}(\sqrt{NT})$, uniformly in K . These appear to be the first such guarantees for product-wise independent draws and the complete MNL posterior. A mean-field representation attains gap-free frequentist regret $\tilde{O}(K\sqrt{NT})$ once its sampling center is certified and its dispersion is corrected for the coupling it discards; a computational study maps the statistical-computational trade-off.

Key words: dynamic assortment planning, multinomial logit, Thompson sampling, posterior sampling

1. Introduction

Motivation. Dynamic assortment planning asks which products a retailer should display when the demand for them must be learned from the purchases they generate. The problem is traditionally cast in the bandit paradigm (Caro and Gallien 2007, Rusmevichientong et al. 2010, Sauré and Zeevi 2013), with feedback that is structured and *censored*: only the alternative a customer selects is observed. In each period, the retailer selects at most K products from a catalog of size N , where the display capacity K reflects limited shelf space or screen positions. The arriving customer either purchases one of the displayed products or selects the *outside option* and leaves without buying. Under the multinomial logit (MNL) demand model (Luce 1959, McFadden 1978, Train 2003), each product i carries an initially unknown *attraction* $v_i > 0$, with larger values indicating greater appeal,

and the probability that a customer buys product i from an assortment S is $v_i/(1 + \sum_{j \in S} v_j)$, where the outside option is normalized to attraction one. The retailer faces two intertwined tasks: earning revenue from the assortments it displays, and learning the attractions that determine which assortments are profitable. The denominator shared by every displayed product enters both tasks. It determines the revenue of an assortment, because adding a product diverts demand from those already shown; and it governs what is learned, because one observed choice is informative about every alternative available and rejected. Over a horizon of T customers, every displayed assortment both earns revenue and generates the substitution evidence on which later ones depend: the exploration–exploitation trade-off in structured form.

Many sequential learning policies address uncertainty through the principle of optimism. An upper confidence bound (UCB) policy constructs a favorable model from a confidence set and acts as though that model were true (Auer et al. 2002); in the assortment setting, Agrawal et al. (2019) build such bounds on the attractions themselves. Thompson sampling (Thompson 1933) instead draws a plausible parameter vector and optimizes for the draw. The intuition is that a decision taken under an optimistic view of demand either performs well, in which case little is lost, or is contradicted by the observations it generates, in which case the optimism is corrected; regret, the revenue forgone relative to a benchmark that knows the true attractions, accumulates only through the second case. A common way to make this precise is to show that the benchmark’s assortment is represented optimistically often enough and that the supporting errors accumulate slowly.

For a randomized policy, how those draws are generated is itself a design choice, and two distinct forms of dependence are at stake. The first is statistical. Because one choice is informative about every displayed alternative, the observations leave the attractions dependent; a policy that maintains one state per product and samples each coordinate independently forgoes that dependence, whereas a policy that retains the complete MNL likelihood preserves it, at a computational cost. The second is algorithmic: even with a product-wise representation, a policy may deliberately depart from a plain draw off its current posterior in order to coordinate exploration, whether by correlating the draws across products, by inflating each product’s sampling variance beyond what its posterior would produce so that a favorable value becomes more likely, or by combining the two, as when a coordinate-wise maximum over several correlated draws is used to secure optimism. The distinction matters because independent draws face an obstacle: if the optimal assortment contains several products, requiring every sampled coordinate to be optimistic at the same time is an event whose probability can shrink exponentially in the display capacity K .

Objectives. The closest MNL Thompson-sampling antecedent is the epoch construction of Agrawal et al. (2026), described below, which accumulates purchases while an assortment is held

fixed and turns them into product-level conjugate updates. What matters here is the distinction between that mechanism and its regret analysis: the worst-case guarantee is proved not for independent draws from those product-wise laws, but for a Gaussian approximation that uses common shocks and a coordinate-wise maximum over several draws, precisely in order to avoid the exponentially unlikely event above. Saha and Gaillard (2025) show that nonexponential capacity dependence is attainable, and with a sharper, near-optimal rate, but they do so within the UCB approach, which sidesteps joint optimistic sampling altogether: a confidence-based rule commits to a single favorable model and never has to make several independently drawn coordinates optimistic at once. Nonexponential capacity dependence in the MNL assortment bandit has thus been reached either by algorithmically coordinating the draws or by avoiding randomized draws in the first place. Whether independent product draws can coordinate through the MNL revenue function itself remains open, and answering it in the affirmative is a contribution of this paper.

Methodological approach. We start from the policy of Agrawal et al. (2026), which holds an assortment fixed until a customer selects the outside option and updates from the purchases accumulated in between. It is best described as Thompson sampling for a *filtered* likelihood: those epoch statistics retain only the comparisons between each displayed product and the outside option, discarding the periods in which a product was displayed and a different product was bought. The closed-form BetaPrime laws it samples from are therefore conjugate updates for the comparison model rather than the marginals of the MNL posterior. The policy applies Bayes’ rule to the retained comparison factors. We keep that comparison but replace the common epoch clock by one clock per product: the clock of product i advances whenever i or the outside option is selected. We show that the construction preserves the conjugate updates while removing the requirement that an assortment remain fixed for a random stretch of time, so that every customer contributes information as soon as the choice is observed and the assortment may be revised just as often. We then move to a full Thompson sampling policy, which discards no data, carries the correlation the MNL likelihood induces rather than imposing it, and remains implementable.

Contributions. The three contributions are organized by how much of the MNL likelihood a policy retains. Retaining less buys tractability: the product-wise clocks admit conjugate updates by using only the Bernoulli comparison factors and discarding the likelihood contributions of periods in which a displayed product lost the sale to another displayed product, and those are the same periods that couple the attractions to one another. Retaining the complete likelihood keeps that substitution evidence, and the posterior dependence it induces, at a computational cost. The first contribution analyzes the filtered end of this range and the second the complete end; the third places an intermediate approximation between them, bounds the regret of a corrected form of it, and measures numerically what separates the three.

1. *Coordinate repair with independent draws.* We identify the structure that lets independent per-product draws produce coordinated exploration. The MNL assortment revenue with heterogeneous product revenues is neither linear nor coordinatewise monotone in the attractions: increasing the attraction of a low-revenue product can divert demand from more profitable alternatives. The structure that yields coordination for linear combinatorial rewards (Zhang and Combes 2024) and for monotone nonlinear ones (Jin et al. 2026) therefore does not apply directly. We recover it through the linear-fractional form of MNL revenue. Write $R(S, x)$ for the expected revenue of assortment S when the attractions are x , and let r_i be the known revenue earned when product i is purchased. The classical transformation of linear-fractional objectives (Charnes and Cooper 1962) gives, for any candidate revenue level λ ,

$$R(S, x) \geq \lambda \quad \Longleftrightarrow \quad \sum_{i \in S} (r_i - \lambda)x_i \geq \lambda.$$

At a fixed level λ , the comparison is linear in the attractions. We use this structure to test whether replacing one omitted product’s sampled attraction by a value slightly below its true attraction would improve the sampled optimal revenue. If it would, a sufficiently large draw for that product forces its selection; if no omitted product passes this test and the displayed coordinates are accurate, the selected assortment has small regret. Scalar posterior-odds bounds control how often a useful product is missed, while its clock controls how often it can be displayed without generating another retained comparison. This *coordinate-repair* argument gives frequentist regret $\tilde{O}(\sqrt{NKT})$ for our *asynchronous*-clock policy with its original independent BetaPrime draws, without variance inflation, whenever attractions are bounded away from zero; here $\tilde{O}(\cdot)$ suppresses logarithmic factors. The guarantee is gap-free, in that it does not deteriorate as the revenue gap between the best assortment and its competitors vanishes, and it carries no exponential dependence on K . To the best of our knowledge, this is the first gap-free, nonexponential (in K) frequentist guarantee for product-wise independent Thompson sampling for the MNL assortment bandit. The contribution here is the sampling mechanism; on capacity dependence, the correlated policy of Agrawal et al. (2026) and the UCB policy of Saha and Gaillard (2025) remain sharper.

2. *The value of the complete posterior.* A clock for a given product retains periods in which said product or the outside option is selected, but omits periods in which it was displayed and another product was purchased. The complete MNL likelihood preserves this substitution evidence and thereby couples the attraction parameters. We adapt a Gamma–Exponential augmentation, which renders the coordinates conditionally independent without destroying their marginal dependence, to obtain a Gibbs sampler that targets the complete posterior, and we analyze

full-likelihood Thompson sampling, which draws exactly from that posterior in every period. Under a compactly supported prior, independent across products, the policy has Bayesian regret $\tilde{O}(\sqrt{NT})$, uniformly over the display capacity. To the best of our knowledge, this is the first capacity-uniform guarantee of this order for Thompson sampling with exact draws from the complete item-level MNL posterior; related posterior-sampling and confidence-based policies for MNL demand have reached the same scale under different models and information structures, as Section 2 discusses. The advance is therefore in the posterior and the proof architecture, not in the rate. That architecture is worth stating in outline. Posterior symmetry turns regret into a comparison between two draws from the same posterior. We measure this nonseparable comparison in the Fisher metric that the MNL model itself induces, and then accumulate along the realized path the information that the sampled assortments generate.

3. *An intermediate representation and empirical evidence.* Exact posterior draws are not free, so the statistical target and what a policy can actually compute must be kept apart. The implementable counterpart, *M-sweep Gibbs Thompson sampling*, replaces the exact draw by finitely many Gibbs sweeps before each decision. We also introduce a third representation, *mean-field filtering* (Section 6), which sits between the two ends of the range. Like the clock policies, it keeps one independent law per product; unlike them it does not discard the shared denominator, but replaces it by a correction computed from the current mean attractions. Computationally, it is as expensive as the clock policies. As implemented it carries no guarantee, but a covariance-corrected variant of it does: certifying the sampling center and correcting the product-wise dispersion yields gap-free regret of order $\tilde{O}(K\sqrt{NT})$. Our numerical experiments compare these three policies against each other and against the correlated Thompson-sampling and UCB benchmarks, and yield three findings. First, finite-sweep Gibbs attains the lowest regret in most designs, but it requires the complete history and one to two orders of magnitude more computation per decision. Second, mean-field filtering recovers most of that gain at clock-like cost. Third, it does so unevenly: it falls behind when revenue gaps are small and at large display capacities, while under intense substitution it matches Gibbs. The numerical contribution is thus a map of the statistical–computational trade-off.

Taken together, the three results separate objects that are easily conflated: the dependence of choices created by the MNL denominator, the posterior dependence created by retaining the complete likelihood, and correlation deliberately introduced across randomized exploration draws. Coordinate repair shows that the third is not necessary for a nonexponential independent-product guarantee in this structured nonlinear problem. The complete-posterior result shows how retaining

the second changes both the attainable analysis and the computational burden. The asynchronous clock also removes a requirement that a competitive extension could not maintain, since no firm can hold the full assortment fixed for a random stretch of time if rivals revise their own offerings on their own schedules; developing that extension is future work.

Organization. Section 2 reviews the related assortment-learning, posterior-sampling, and inference literature. Section 3 formulates the MNL assortment bandit and its regret criterion. The next three sections develop one policy each: Section 4 the asynchronous-clock policy, Section 5 full-likelihood Thompson sampling, and Section 6 mean-field filtering. Section 7 reports the computational study, and Section 8 concludes. The e-companion collects the notation, supporting lemmas, and technical detail that the two performance-bound proofs set aside, together with the supporting constructions and the proofs of the lemmas.

Notation. We reserve v for the true attraction vector, x for a generic vector, and $\tilde{v}$ for a posterior sample. We write $\mathbb{P}(\cdot)$ for probabilities and $\mathbb{E}[\cdot]$ for expectations. Their subscripts specify the governing law or the conditioning information; in $\mathbb{P}_v$ and $\mathbb{E}_v$, the attractions are fixed at v . For an event A , $\mathbb{1}\{A\}$ is its indicator: one if the event occurs and zero otherwise. For every positive integer m , $[m] := \{1, \dots, m\}$; if B is finite, $|B|$ denotes its cardinality. We define $\mathbb{R}_+ := [0, \infty)$ and $\mathbb{R}_{++} := (0, \infty)$, and $\sigma(\cdot)$ for the σ -algebra generated by its argument. Finally, for $\alpha, \beta > 0$ we write $X \sim \text{BetaPrime}(\alpha, \beta)$ when $X/(1 + X) \sim \text{Beta}(\alpha, \beta)$; writing $\alpha = \mu(\beta - 1)$, one has $\mathbb{E}[X] = \mu$ for $\beta > 1$ and $\text{Var}(X) = \mu(1 + \mu)/(\beta - 2)$ for $\beta > 2$.

2. Literature Review

Assortment Optimization and Dynamic Learning. When attractions are known, van Ryzin and Mahajan (1999) study revenue maximization under MNL demand, and Talluri and van Ryzin (2004) does so under a general discrete-choice model; under a display capacity, Rusmevichientong et al. (2010) give an algorithm based on the geometry of lines that returns a revenue-maximizing assortment, which we take as an oracle. See Kök et al. (2008) for a survey of assortment planning in retail. Static solvability does not determine how attractions should be learned: once they are estimated from sales, the offered assortments determine both current revenue and later information. Caro and Gallien (2007) first formulate dynamic assortment learning as a finite-horizon Bayesian bandit, assuming independent demands. Under MNL demand, Rusmevichientong et al. (2010) and Sauré and Zeevi (2013) are the first to study regret directly; both explore and then exploit, obtaining asymptotic regret $O(N \log T)$ when the revenue gap between the optimal assortment and the best alternative is known. Yang (2021) refines this analysis product by product, and Lattimore and Szepesvári (2020) survey the underlying bandit problems.

Optimism and Confidence Bounds. The principle of optimism under uncertainty, illustrated by the UCB policies of Auer et al. (2002), evaluates each action through an estimate that is favorable yet still compatible with the observations. In the MNL-bandit, Agrawal et al. (2019) exploit the structure of the model to construct upper confidence bounds on the attractions and offer the assortment that is optimal under them, attaining worst-case regret of order $\tilde{O}(\sqrt{NT} + N)$. Saha and Gaillard (2025) use rank-breaking to update product-versus-outside comparisons after every customer and obtain a leading term of order $\tilde{O}(\sqrt{NT})$ when no product is more attractive than the outside option. For $T = \Omega(N)$ and $K \leq N/4$, Chen and Wang (2018) prove a lower bound matching the $\sqrt{NT}$ scale up to logarithmic factors.

Thompson Sampling. Rather than selecting the most favorable parameter within a confidence region, Thompson sampling maintains a distribution over the parameters, draws a sample, and decides as if that sample described the true demand. Introduced by Thompson (1933), the rule attains near-optimal gap-free guarantees in the classical bandit (Agrawal and Goyal 2017), and Russo and Van Roy (2014) relate posterior sampling to optimism, deriving Bayesian bounds whose size depends on how closely the revenues of different actions are related; Russo et al. (2018) survey the area. For MNL assortments, Agrawal et al. (2026) organize observations into outside-option epochs and recover product-level BetaPrime conjugacy, but obtain their worst-case guarantee through a Gaussian approximation with cross-product correlation and a coordinate-wise maximum over several draws. In the joint assortment-and-pricing problem, Miao and Chao (2021) obtain Bayesian regret of order $\tilde{O}(\sqrt{NT} + N)$, which is $\tilde{O}(\sqrt{NT})$ when $T \gtrsim N$, for a cycle-based Thompson-sampling policy whose sampling construction differs from ours.

Contextual Formulations. A more general class links product utilities through a parameter shared across items. Oh and Iyengar (2019) analyze a policy that samples from the posterior of a d -dimensional utility parameter and obtain Bayesian regret $\tilde{O}(d\sqrt{T})$, assuming access to the required posterior draws; with one-hot product features their model permits item-specific utilities but has $d = N$, so the displayed rate becomes $\tilde{O}(N\sqrt{T})$. Oh and Iyengar (2021) obtain a frequentist $\tilde{O}(\sqrt{dT})$ guarantee through a more involved UCB construction, which specializes to $\tilde{O}(\sqrt{NT})$ when $d = N$, and Lee and Oh (2025) derive capacity-free, variance-dependent frequentist bounds by combining reward covariance, MNL Fisher widths, and an elliptical potential.

Combinatorial Bandits. The coordination difficulty we address also appears when an action is a set of elementary components. Wang and Chen (2018) analyze Thompson sampling with independent draws for those components. Zhang and Combes (2024) obtain polynomial, gap-dependent regret from variance-inflated independent Gaussian draws under linear rewards, and for monotone

nonlinear rewards Jin et al. (2026) show that independent sampling can make the probability of exploring the optimal action exponentially small, replacing the separate draws by a shared seed. A structured exception appears in cascading bandits: Zhong et al. (2021) derive a problem-independent, nonexponential bound for independent Beta draws when the optimal items have equal click probabilities, whereas their sharper policy coordinates exploration through a common Gaussian perturbation. MNL revenue falls outside both the linear and the coordinatewise-monotone cases these results exploit.

Posterior Inference and Computation. Thompson sampling presumes that the policy can sample the law representing its uncertainty, and with joint posteriors this step is not a neutral implementation detail. Phan et al. (2019) show that an approximation which appears statistically close to its target can still induce linear regret, whereas Mazumdar et al. (2020) construct Langevin methods that preserve the optimal instance-dependent regret order under suitable log-concavity and concentration conditions. To retain the MNL posterior dependence, we adapt the data-augmentation principle of Tanner and Wong (1987) and the Gamma–Exponential construction of Caron and Doucet (2012) for generalized Bradley–Terry and Plackett–Luce models.

Positioning. Agrawal et al. (2026) is the closest in sampling mechanism: product-level conjugacy recovered from epoch statistics, with optimism secured by correlating the draws. Saha and Gaillard (2025) is the closest in information and timing: the same product-versus-outside comparisons, revised after every customer, with optimism secured through confidence bounds. Our policies take a rather more direct approach. The asynchronous-clock policy keeps the per-customer timing and the product-wise filter, and obtains its guarantee from independent BetaPrime draws without variance inflation; the full-likelihood policy abandons the filter, retaining the dependence that the MNL likelihood induces at a computational cost. The three constructions differ in what information is retained and in how optimism is produced, and the paper measures both.

3. Problem Formulation

Consider a retailer who operates over a horizon of T periods, one customer per period, with a product catalog $\mathcal{N} := [N]$. In period $t \in [T]$ it displays an *assortment* $S_t \subseteq \mathcal{N}$; then customer t buys one product of S_t or selects the *outside option*, denoted by 0. We write $I_t \in S_t \cup \{0\}$ for the alternative chosen in period t .

Choices follow a multinomial logit (MNL) model. Each product $i \in \mathcal{N}$ carries an unknown *attraction* $v_i > 0$, and the outside option is normalized to $v_0 \equiv 1$ for identification. Given the offered

assortment S and the attraction vector $v := [v_i]_{i \in \mathcal{N}}$, the probability that the customer selects $i \in S \cup \{0\}$ is

$$p_i(S, v) = \frac{v_i}{1 + \sum_{j \in S} v_j},$$

with $p_i(S, v) = 0$ for $i \in \mathcal{N} \setminus S$. This remains the conditional choice law given the past history and the policy's current randomization. The denominator aggregates the attractions of every alternative present in the choice set, so adding a product to S , holding the other attractions fixed, lowers the choice probability of every product already displayed. Each product i yields a known revenue $r_i \in [0, 1]$ when purchased, so the expected revenue of assortment S under v is

$$R(S, v) = \sum_{i \in S} r_i p_i(S, v) = \frac{\sum_{i \in S} r_i v_i}{1 + \sum_{j \in S} v_j}. \quad (1)$$

Adding a product both contributes its own revenue and enlarges the denominator faced by the others. An optimal assortment therefore balances profitability against composition and is not, in general, the set of highest-revenue products. That same denominator is a source of statistical difficulty: a purchase is evidence about the product bought and also about the products that were displayed and declined. Retaining this cross-product information yields a posterior faithful to the model, but it couples every pair of products that has ever been displayed together; discarding it permits product-wise updates at the cost of part of the substitution signal. We call this the *information-filtration trade-off*.

Display space is limited: a capacity $K \leq N$ restricts the feasible decisions to $\mathcal{S} := \{S \subseteq \mathcal{N} : |S| \leq K\}$. For $x \in \mathbb{R}_+^N$ we write

$$S^*(x) \in \arg \max \{R(S, x) : S \in \mathcal{S}\}$$

for the optimal assortment under x . Ties are broken by two conventions applied in order. Among the maximizers we retain those that are inclusion-wise minimal, that is, those from which no product can be deleted without losing optimality; if several remain, we take the smallest under a fixed total order on the subsets of $\mathcal{N}$, say the lexicographic order induced by the product labels. This makes $S^*(x)$ a single assortment and prevents it from carrying products that contribute no revenue, a property the analysis uses repeatedly. The *full-information oracle* is the hypothetical retailer that knows v and offers $S^*(v)$ in every period.

The retailer does not know v , and observes only the customer's choice from the assortment it displayed. Let $\mathcal{H}_t := \{(S_\tau, I_\tau) : \tau \leq t\}$ be the history through period t and $\mathcal{F}_t := \sigma(\mathcal{H}_t)$ the σ -algebra it generates, with $\mathcal{F}_0$ trivial, so that the information available when period t begins is $\mathcal{F}_{t-1}$. A possibly randomized *policy* is a sequence of probability kernels on $\mathcal{S}$, one per period, indexed by the realized histories $\mathcal{H}_{t-1}$. It is admissible if each kernel depends only on the observed history

and is implemented using randomization independent of v , past observations, and the randomness generating customer choices.

Performance is measured against the oracle. The expected *regret* of a policy is

$$\text{Reg}(T, v) = \mathbb{E} \left[\sum_{t=1}^T R(S^*(v), v) - R(S_t, v) \right], \quad (2)$$

the expectation running over the customer choices and over any randomization the policy uses. Regret is the cost of having to learn v while serving customers: early on the policy may sacrifice sales to explore demand, and later it may still lose revenue by acting on an imperfect estimate. A policy attains *sublinear regret* when $\text{Reg}(T, v) = o(T)$.

The policies studied below share the same structure: each maintains a prior over the attractions, updates it with the choices observed so far, draws a sample $\tilde{v}$ from the resulting law, and offers the assortment $S^*(\tilde{v})$ that is optimal under that draw. They differ in how much of the MNL likelihood the representation they update retains. Asynchronous clocks keep product-versus-outside comparisons and admit product-wise BetaPrime updates; full-likelihood Thompson sampling keeps the complete MNL likelihood; and mean-field filtering keeps one law per product but corrects the shared denominator. The three sections that follow take them in that order, placing each at a different point of the information-filtration trade-off.

4. Asynchronous Clocks and Independent BetaPrime Sampling

Our starting point is the epoch policy of Agrawal et al. (2026), which can be described as a Thompson sampling policy for a *filtered* likelihood: posteriors (and assortments) are computed only when the outside alternative is chosen (i.e. an epoch), at which point only the comparisons between each displayed product and the outside option are retained. This section develops the first policy and its regret guarantee. Section 4.1 replaces the common epoch clock by one clock per product and records the conjugate update; Section 4.2 introduces the one-coordinate repair argument and states the bound for the conjugate sampling law; and Section 4.3 proves it.

4.1. The Asynchronous Clock Policy

For $i \in \mathcal{N}$, define

$$V_{i,t} := \sum_{\tau=1}^{t-1} \mathbb{1}\{I_\tau = i\}, \quad N_{i,t} := \sum_{\tau=1}^{t-1} \mathbb{1}\{i \in S_\tau, I_\tau = 0\},$$

the number of periods before t in which i was purchased and the number in which i was on display and the outside option was taken. If product $i \in \mathcal{N}$ is offered in period t , then its probability of being chosen over the outside alternative is independent of the assortment offered, i.e.

$$\mathbb{P}_v(I_t = i \mid I_t \in \{i, 0\}, S_t) = \frac{v_i}{1 + v_i}.$$

Together with a BetaPrime prior, this comparison yields a product-level conjugate update. Write $\mathcal{G}_{i,t}$ for the information carried by the comparisons involving i before period t : their outcomes, taken in order, together with their number, but not the periods in which they occurred; we call each such outcome a *tick* of product i , and the comparison that produced it a *retained* comparison; product i retains a comparison in exactly those periods that display it and end with a sale to i or to the outside option, and we use the two terms interchangeably in what follows. The individual tick times are deliberately excluded, since a period in which i was displayed but no comparison was recorded would reveal that another displayed product won the sale, which is the evidence the filter discards. Agrawal et al. (2026) show that if $v_i \sim \text{BetaPrime}(\alpha_0, \beta_0)$ a priori, with $\alpha_0, \beta_0 > 0$ here and in all that follows, then the conjugate update based on these comparisons stays in that family, with the two counts added to the prior parameters, so these counts are all that the policy needs to carry for product i . Their argument covers the epoch construction, that is, only the periods in which a customer selects the outside option, and only when the assortment has been held fixed since the previous such period. Note that the update uses the Bernoulli factors of the comparisons recorded in $\mathcal{G}_{i,t}$; their product need not be the likelihood of $\mathcal{G}_{i,t}$, since the number of recorded comparisons can also carry information about v_i . Given $\mathcal{F}_{t-1}$, the attractions follow the MNL posterior, which couples the products and is not BetaPrime. Proposition 1 relaxes both conditions: the update remains valid in every period, not only in those in which the outside option is chosen, and whatever assortments were offered in the meantime.

PROPOSITION 1 (Product-Level BetaPrime Update). *Let $i \in \mathcal{N}$. Conditional on v , the ticks successively recorded by the clock of i form an initial segment of an i.i.d. $\text{Bernoulli}(v_i/(1 + v_i))$ sequence. If $v_i \sim \text{BetaPrime}(\alpha_0, \beta_0)$ a priori, then the conjugate update based on the recorded Bernoulli comparisons gives, in each period $t \in [T]$,*

$$\tilde{v}_{i,t} \sim \text{BetaPrime}(\alpha_0 + V_{i,t}, \beta_0 + N_{i,t}).$$

Our first policy, *asynchronous-clock Thompson sampling*, depicted in algorithmic form in Algorithm 1, exploits this: it redraws $\tilde{v}$ and reoptimizes the assortment in every period. We call the pair $(V_{i,t}, N_{i,t})$ the *clock* of product i , since its counts advance at their own pace, on a schedule set by the policy's past decisions. Drawing the coordinates independently does not make the observations independent: one outside-option choice advances the clock of every displayed product.

4.2. One-Coordinate Repair and the Regret Bound

Algorithm 1 draws each $\tilde{v}_{i,t}$ from the conjugate BetaPrime sampling law of Proposition 1, independently across products, and the regret guarantee below applies to that draw exactly as specified:

Algorithm 1 Asynchronous-Clock Thompson Sampling

-
- 1: **Input:** horizon T , capacity K , revenues $\{r_i\}_{i=1}^N$, prior (α_0, β_0) .
 - 2: **Initialize:** set $V_i \leftarrow 0$ and $N_i \leftarrow 0$, $\forall i \in \mathcal{N}$.
 - 3: **for** $t = 1, \dots, T$ **do**
 - 4: Sample $\tilde{v}_i \sim \text{BetaPrime}(\alpha_0 + V_i, \beta_0 + N_i)$ independently, $\forall i \in \mathcal{N}$, and offer $S_t \leftarrow S^*(\tilde{v})$.
 - 5: Observe $I_t \in S_t \cup \{0\}$ and tick the clocks: $V_i \leftarrow V_i + \mathbb{1}\{I_t = i\}$ and $N_i \leftarrow N_i + \mathbb{1}\{i \in S_t, I_t = 0\}$, $\forall i \in \mathcal{N}$.
 - 6: **end for**
-

it neither inflates these sampling variances nor induces correlation across products, for instance by redrawing several times and retaining a coordinate-wise maximum.

For our analysis, we retain the normalization that no product is more attractive than the outside option and assume a uniform lower bound $v_{\min} \in (0, 1)$ on the attractions.

ASSUMPTION 1. *The attraction vector satisfies $v \in [v_{\min}, 1]^N$, and $\beta_0 > 0$.*

The lower bound makes the comparison between attraction errors and revenue losses uniform across instances. The upper bound controls the probability that a product advances its clock.

Outline of the Argument. A Thompson-sampling guarantee is usually obtained through an optimism event: the sampled vector must represent the oracle assortment $S^*(v)$ favorably often enough. Under independent product-wise draws, that event asks all $|S^*(v)|$ coordinates to be favorable at once, and its probability can decay exponentially in K . Our proof of Theorem 1 replaces it by a condition on one coordinate at a time. The mechanism is *one-coordinate repair*: an appreciable revenue loss must involve either a selected product whose sampled attraction is inaccurate, or an omitted product whose inclusion can be forced by correcting only that product's sampled attraction. The first failure is a matter of posterior concentration. For the second, conditional independence compares the probability of a failed repair with the probability of selecting the same product, and the analysis charges these selection opportunities to the comparisons that product's clock retains.

The Repair Event. For a vector x , write $x_{-i} := (x_j)_{j \neq i}$ for its coordinates other than i , and write $R^*(x) := R(S^*(x), x)$ and $R^* := R^*(v)$. To express repair without using the coordinate being tested, let ξ be an attraction-error tolerance with $0 < \xi \leq v_{\min}/2$. For each $x \in \mathbb{R}_{++}^N$, let $x^{(i)}$ agree with x , except that its i -th coordinate is replaced by $v_i - \xi$; we say that product i is *repairable* at x when correcting that one coordinate makes every optimal assortment contain it. The repair event compares the resulting optimum with the best revenue attainable without product i :

$$\mathcal{D}_i(\xi) := \left\{ x_{-i} : x \in \mathbb{R}_{++}^N, R^*(x^{(i)}) > \max_{\substack{S \in \mathcal{S} \\ i \notin S}} R(S, x) \right\}. \quad (3)$$

For $i \notin S$, $R(S, x) = R(S, x^{(i)})$, so the condition depends only on x_{-i} . The strict inequality rules out any optimum that omits i after the repair.

LEMMA 1 (One-Coordinate Coverage). *Let $0 < \xi \leq v_{\min}/2$ and $x \in \mathbb{R}_{++}^N$.*

(i) *If $x_{-i} \in \mathcal{D}_i(\xi)$ and $x_i \geq v_i - \xi$, then $i \in S^*(x)$.*

(ii) *If $|x_i - v_i| \leq \xi$ for every $i \in S^*(x)$ and $x_{-i} \notin \mathcal{D}_i(\xi)$ for every $i \notin S^*(x)$, then $R^* - R(S^*(x), v) \leq 2\xi/v_{\min}$.*

Part (ii) is the formal content of the mechanism described above: away from these two failures, the sampled assortment is nearly optimal. Part (i) is what makes a failure checkable one coordinate at a time. Both parts are proved in the e-companion.

Together the two parts reduce the regret to two counts: the periods in which some selected product has an inaccurate draw, and those in which a repairable product is omitted. Concentration of the BetaPrime draws bounds the first; the posterior odds of drawing below $v_i - \xi$, accumulated over the comparisons that the clock of i retains, bound the second. Both bounds hold for every tolerance ξ , so choosing ξ proportionally to a revenue-gap level and integrating the resulting count over that level yields the following guarantee.

THEOREM 1 (Regret of the Asynchronous-Clock Policy). *Suppose Assumption 1 holds. There exists $C < \infty$, depending only on $(v_{\min}, \alpha_0, \beta_0)$, such that Algorithm 1 satisfies*

$$\text{Reg}(T, v) \leq C\sqrt{NKT \log(eNT)} + C.$$

In particular, uniformly for $1 \leq K \leq N$, the regret is of order $\tilde{O}(\sqrt{NKT})$.

The guarantee preserves both the conjugate sampling law and independent product-wise draws. One-coordinate repair controls the cost of omitted products, while concentration controls the selected ones; the clocks account for both costs without synchronizing the assortment decisions. Our proof must also deal with the randomness in the clocks themselves: a deterministic prefix of the retained comparisons has a binomial count, but the count observed at a policy-dependent time need not have a binomial law conditional on its length. It therefore uses the deterministic prefixes only after a pathwise accounting step.

4.3. Proof of Theorem 1

We bound regret by counting periods with an appreciable revenue loss and then integrating this count. Let ξ be an attraction-error tolerance with $0 < \xi \leq v_{\min}/2$. By Lemma 1(ii), a loss above $2\xi/v_{\min}$ has one of two causes: a selected product i with $|\tilde{v}_{i,t} - v_i| > \xi$, or an omitted product whose inclusion can be forced by correcting its draw. Errors of the first type are counted by Lemma 2, stated next. Steps 1 and 2 control the omitted products; Step 3 combines the two counts. Two further auxiliary results, Proposition EC.1 and Lemma EC.2, remain in Section EC.3.1; each is described where it is used.

Throughout, a subscript t on probability or expectation denotes conditioning on $\mathcal{F}_{t-1}$ under the fixed true vector v . Constants $C > 0$ may change from line to line and depend only on $(v_{\min}, \alpha_0, \beta_0)$.

LEMMA 2 (Selected-Coordinate Error). *For the asynchronous-clock policy and $0 < \xi \leq v_{\min}/2$,*

$$\mathbb{E}_v \left[\sum_{t=1}^T \mathbf{1}\{\exists i \in S_t : |\tilde{v}_{i,t} - v_i| > \xi\} \right] \leq \frac{CN(K+1)}{\xi^2} \log(eNT) + C, \quad (4)$$

where C depends only on (α_0, β_0) .

The count runs over periods: a product with few ticks can be selected repeatedly, so the clock-advance probability in (EC.2) converts control per comparison into the capacity factor $K+1$. The lemma is proved in the e-companion.

Step 1 (Comparing Omission with Selection). Hold the other sampled attractions fixed. The repair event $\mathcal{D}_i(\xi)$ of (3) records that i is repairable at the current draw, and does not involve the draw of coordinate i . A failed repair occurs when this condition holds but the actual sampled assortment omits i ; write $\mathcal{B}_{i,t}(\xi) := \{\tilde{v}_{-i,t} \in \mathcal{D}_i(\xi), i \notin S_t\}$.

On the repair event, any draw at least $v_i - \xi$ forces the selection of i , by Lemma 1(i). Thus omission requires a draw below this threshold. Set $q_{i,t}(\xi) := \mathbb{P}_t(\tilde{v}_{i,t} \geq v_i - \xi)$ and $\omega_{i,t}(\xi) := (1 - q_{i,t}(\xi))/q_{i,t}(\xi)$. These odds compare missing the threshold with reaching it. They depend only on the current counts, and $q_{i,t}(\xi) > 0$ because the BetaPrime density is positive. Conditional independence gives

$$\mathbb{P}_t(\mathcal{B}_{i,t}(\xi)) \leq (1 - q_{i,t}(\xi))\mathbb{P}_t(\tilde{v}_{-i,t} \in \mathcal{D}_i(\xi)) \leq \omega_{i,t}(\xi)\mathbb{P}_t(i \in S_t). \quad (5)$$

For the second inequality, the repair event together with a draw above the threshold has probability $q_{i,t}(\xi)\mathbb{P}_t(\tilde{v}_{-i,t} \in \mathcal{D}_i(\xi))$ and forces selection. We have therefore bounded an omission probability by a selection probability weighted by scalar posterior odds.

Step 2 (From Selections to Retained Comparisons). The odds change only when product i records a comparison. Several calendar periods can therefore incur the same odds cost before its clock advances. A selected product records a comparison with probability at least $1/(K+1)$, by (EC.2). This bound will control the repeated cost of each prefix; the scalar posterior-odds lemma will then sum over prefixes.

Write $n_{i,t} := V_{i,t} + N_{i,t}$ for the comparisons retained before t and $\theta_i := v_i/(1+v_i)$. Proposition EC.1 represents the purchase count as $V_{i,t} = \sum_{k=1}^{n_{i,t}} Z_{i,k}$, where $\sum_{k=1}^n Z_{i,k} \sim \text{Binomial}(n, \theta_i)$ at every deterministic length n . Under (EC.1), the transformed draw $\tilde{v}_{i,t}/(1+\tilde{v}_{i,t})$ has the Beta law with these success and failure counts.

To apply the scalar lemma, express the repair threshold in the same scale. The value $v_i - \xi$ becomes $\theta_i - \epsilon_i$, where $\epsilon_i = \xi/\{(1+v_i)(1+v_i-\xi)\}$. Since $\xi \leq v_i/2$ and $v_i \leq 1$, we have $0 < \epsilon_i < \theta_i \leq 1/2$ and $\epsilon_i \geq \xi/4$, which are the conditions under which Lemma EC.2 sums these odds over deterministic prefixes. For n comparisons and s successes, denote the corresponding odds by

$$\omega_n(s, \epsilon_i) := \mathbb{P}(\text{Beta}(\alpha_0 + s, \beta_0 + n - s) \geq \theta_i - \epsilon_i)^{-1} - 1.$$

Then $\omega_{i,t}(\xi) = \omega_{n_{i,t}}(V_{i,t}, \epsilon_i)$. We use the binomial law only after replacing the adaptive prefix lengths by a sum over deterministic ones.

We can now bound the total failed-repair count. Taking expectations in (5) replaces its conditional selection probability by a selection indicator, since $\omega_{i,t}(\xi)$ is $\mathcal{F}_{t-1}$ -measurable. Conditional also on the current draw, the weight and assortment are fixed before the customer chooses, and a selected product records a comparison with probability $(1+v_i)/(1+\sum_{j \in S_t} v_j) \geq 1/(K+1)$. Consequently,

$$\begin{aligned} \mathbb{E}_v \left[\sum_{t=1}^T \mathbb{1}\{\mathcal{B}_{i,t}(\xi)\} \right] &\leq \mathbb{E}_v \left[\sum_{t=1}^T \omega_{i,t}(\xi) \mathbb{1}\{i \in S_t\} \right] \\ &\leq (K+1) \mathbb{E}_v \left[\sum_{t=1}^T \omega_{i,t}(\xi) \mathbb{1}\{i \in S_t, I_t \in \{i, 0\}\} \right] \\ &= (K+1) \mathbb{E}_v \left[\sum_{n=0}^{n_{i,T+1}-1} \omega_n \left(\sum_{k=1}^n Z_{i,k}, \epsilon_i \right) \right] \\ &\leq (K+1) \sum_{n=0}^{T-1} \mathbb{E}_v \left[\omega_n \left(\sum_{k=1}^n Z_{i,k}, \epsilon_i \right) \right] \leq \frac{C(K+1)}{\xi^2} \log(eT). \end{aligned} \quad (6)$$

The equality counts realized comparisons: each advances the local count by one and uses a distinct prefix, from 0 through $n_{i,T+1} - 1$. Since $n_{i,T+1} \leq T$, adding the unvisited prefixes increases the non-negative sum. The last inequality is the summed-odds bound (EC.3) of Lemma EC.2, which bounds

the sum of expected odds over deterministic prefixes by $C\epsilon_i^{-2}\log(eT)$; here $\epsilon_i \geq \xi/4$. Summing over products bounds the expected number of failed repairs by $CN(K+1)\xi^{-2}\log(eT)$.

Step 3 (Integrating Revenue Gaps). Set $\Delta_t := R^* - R(S_t, v) \in [0, 1]$. For $g \in (0, 1]$, choose $\xi = v_{\min}g/4$. The coverage property used at the start of the proof now says that $\Delta_t > g$ requires a selected-coordinate error or a failed repair: in their absence, Lemma 1(ii) gives $\Delta_t \leq 2\xi/v_{\min} = g/2$. Adding the failed-repair count in (6) and the selected-error count in Lemma 2 yields

$$\mathbb{E}_v \left[\sum_{t=1}^T \mathbb{1}\{\Delta_t > g\} \right] \leq \frac{CN(K+1)\log(eNT)}{v_{\min}^2 g^2} + C. \quad (7)$$

Both event counts were bounded without conditioning on the revenue gap. The count is also at most T , which caps the g^{-2} bound near zero. Use $K+1 \leq 2K$ and absorb $v_{\min}^{-2}$ into $C \geq 1$. The tail-integral identity, followed by the change of variable $g = \sqrt{NK\log(eNT)/T}u$, gives

$$\begin{aligned} \text{Reg}(T, v) &= \int_0^1 \mathbb{E}_v \left[\sum_{t=1}^T \mathbb{1}\{\Delta_t > g\} \right] dg \\ &\leq C\sqrt{NKT\log(eNT)} \int_0^\infty \min\{1, u^{-2}\} du + C = 2C\sqrt{NKT\log(eNT)} + C. \end{aligned}$$

Extending the transformed integral to $(0, \infty)$ increases it. Its value is 2, which is absorbed into C in the stated bound. $\blacksquare$

REMARK 1 (EPOCH EXTENSION). The same regret bound holds when the independent conjugate policy retains its assortment until an outside-option choice, as established in Corollary EC.2 of the e-companion. In particular, the epoch guarantee covers the initialization $\alpha_0 = \beta_0 = 1$ of Algorithm 1 in Agrawal et al. (2026). Both proofs use the same one-coordinate repair and scalar posterior-odds bounds; only the accounting for the time between draws changes.

5. Sampling from the Complete MNL Posterior

The asynchronous-clock policy is simple to implement and admits an explicit regret bound, but it does not use all the posterior dependence generated by the MNL denominator. The next step is to study that dependence without reducing the history to product-level comparisons. The policy of this section retains all the likelihood information in the complete history and targets the posterior that the MNL likelihood induces. Section 5.1 constructs that posterior and the sampler that targets it, Section 5.2 states the Bayesian regret guarantee, and Section 5.3 proves it.

5.1. The Full-likelihood Thompson Sampling Policy

We write π_0 for the joint prior distribution of v , under which the attractions are independent across products with $v_i \sim \text{Gamma}_{[v_{\min}, 1]}(\alpha_0, \beta_0)$ for every $i \in \mathcal{N}$, where $v_{\min} \in (0, 1)$ and the subscript

indicates that each Gamma law is conditioned on the event $v_i \in [v_{\min}, 1]$ (Gamma laws use the shape–rate parameterization throughout). Given $\mathcal{H}_{t-1}$, the exact MNL likelihood is

$$\mathcal{L}(v \mid \mathcal{H}_{t-1}) = \prod_{\tau=1}^{t-1} \frac{v_{I_\tau}^{\mathbb{1}\{I_\tau \neq 0\}}}{1 + \sum_{j \in S_\tau} v_j}. \quad (8)$$

Write π_t for the posterior of v given $\mathcal{H}_{t-1}$. Conditional on the recorded history, the policy’s assortment-selection kernels do not depend on the candidate value of v , so they cancel from Bayes’ formula and $\pi_t(v) \propto \pi_0(v) \mathcal{L}(v \mid \mathcal{H}_{t-1})$. The realized assortments still enter the likelihood through the MNL denominators, which couple the attractions of products displayed together and prevent π_t from factorizing across coordinates. Adapting the Gamma–Exponential construction of Caron and Doucet (2012), the following representation turns the MNL denominator into product-wise independent Gamma conditionals.

PROPOSITION 2 (Factorization of the Augmented Posterior). *There is an augmented distribution whose v -marginal is π_t . Under this distribution, conditional on v and $\mathcal{H}_{t-1}$, the variables y_τ , $\tau = 1, \dots, t-1$, are mutually independent and satisfy $y_\tau \mid v, \mathcal{H}_{t-1} \sim \text{Exponential}\left(1 + \sum_{j \in S_\tau} v_j\right)$. Moreover, conditional on $\{y_\tau\}_{\tau=1}^{t-1}$, the attraction parameters are mutually independent, and for $i \in \mathcal{N}$,*

$$v_i \mid \{y_\tau\}_{\tau=1}^{t-1}, \mathcal{H}_{t-1} \sim \text{Gamma}_{[v_{\min}, 1]} \left(\alpha_0 + V_{i,t}, \beta_0 + \sum_{\tau=1}^{t-1} \mathbb{1}\{i \in S_\tau\} y_\tau \right).$$

The latent exposures turn the shared denominator into a linear one: a purchase of i increases its shape, and each display contributes to its rate. Integrating them out restores the dependence among the products offered together. Our second policy, *full-likelihood Thompson sampling*, depicted in algorithmic form in Algorithm 2, draws $\tilde{v}$ from this representation and offers $S^*(\tilde{v})$ in every period. Unlike the filtered-likelihood updates, every observation contributes a denominator exposure for all displayed products, including when the customer chooses the outside option, and a purchase additionally increases the shape of the chosen product. By retaining both sources of substitution information, the analysis moves from coordinate repair and scalar posterior odds to the comparison between two draws from the same posterior.

REMARK 2 (IMPLEMENTATION). Proposition 2 defines a Gibbs kernel targeting the complete posterior without updating all attractions as one joint block: with v fixed, the sampler draws the exposures, whose rates are the total attractions of the corresponding assortments plus the outside option; conditional on those exposures, the coordinates of v have independent Gamma laws. Repeated occurrences of the same assortment can be grouped without approximation. Let $\mathcal{A}_t$ be the set of distinct nonempty assortments among $S_1, \dots, S_{t-1}$, let $n_{A,t} := \sum_{\tau < t} \mathbb{1}\{S_\tau = A\}$, and aggregate the record-level exposures as $Y_A := \sum_{\tau < t: S_\tau = A} y_\tau$. Corollary EC.1 of Section EC.2

Algorithm 2 *M-Sweep Gibbs Thompson Sampling for the Complete MNL Posterior*

-
- 1: **Input:** horizon T , capacity K , revenues $\{r_i\}_{i=1}^N$, prior (α_0, β_0) , truncation $[v_{\min}, 1]$, sweeps $M \geq 1$.
 - 2: **Initialize:** set $V_i \leftarrow 0$, $\forall i \in \mathcal{N}$, and $n_A \leftarrow 0$ for every assortment A .
 - 3: **for** $t = 1, \dots, T$ **do**
 - 4: **for** $m = 1, \dots, M$ **do**
 - 5: Sample $Y_A \sim \text{Gamma}(n_A, 1 + \sum_{j \in A} \tilde{v}_j)$ for every A with $n_A > 0$, then $\tilde{v}_i \sim \text{Gamma}_{[v_{\min}, 1]}(\alpha_0 + V_i, \beta_0 + \sum_{A \ni i: n_A > 0} Y_A)$, $\forall i \in \mathcal{N}$.
 - 6: **end for**
 - 7: Offer $S_t \leftarrow S^*(\tilde{v})$, observe $I_t \in S_t \cup \{0\}$, and update $n_{S_t} \leftarrow n_{S_t} + 1$ if $S_t \neq \emptyset$, and $V_{I_t} \leftarrow V_{I_t} + 1$ if $I_t \neq 0$.
 - 8: **end for**
-

shows that drawing one Gamma exposure Y_A per $A \in \mathcal{A}_t$ leaves the Gibbs transition kernel on v unchanged; Algorithm 2 uses this grouped kernel.

At $t = 1$, the exposure block is empty and the attraction update draws from the prior. Subsequent periods start from the previous period's final draw.

5.2. Bayesian Regret Bound

The guarantee for this policy is Bayesian rather than worst case. In Section 4 the attraction vector was fixed and the loss controlled for every market compatible with the assumptions; here v is drawn once from π_0 , held fixed over the horizon, and the loss is averaged over the markets the prior deems plausible. The object we bound is the *Bayesian regret*

$$\text{Reg}_B(T) := \mathbb{E}_{v \sim \pi_0}[\text{Reg}(T, v)],$$

whose outer expectation averages (2) over the prior while retaining, for each realization of v , the same comparison with the full-information oracle. With this, the true attractions are compared directly with the sampled attractions that guide the next decision, so one-period regret is controlled without an optimism event relative to a fixed unknown parameter. This is the standard posterior-sampling comparison of Russo and Van Roy (2014), adapted to MNL assortment revenue.

Thompson sampling presumes an exact draw from the current posterior: in each period it draws $\tilde{v}_t \sim \pi_t$, independently of v given $\mathcal{F}_{t-1}$, and offers $S^*(\tilde{v}_t)$. Note, however, that Algorithm 2 does not, so we write *M-sweep Gibbs Thompson sampling* whenever that distinction matters and keep the unqualified name for the exact draw. Exact draws from the posterior are essential to our next result: conditional on $\mathcal{F}_{t-1}$, the true and the sampled attraction vectors are independent draws from the same law π_t .

ASSUMPTION 2. *The true attraction vector is drawn from the prior π_0 .*

THEOREM 2 (Bayesian Regret of Full-Likelihood Thompson Sampling). *Under Assumption 2, there is a constant $C < \infty$, depending only on $(v_{\min}, \beta_0)$, such that, uniformly for $1 \leq K \leq N$, full-likelihood Thompson sampling satisfies*

$$\text{Reg}_B(T) \leq C \sqrt{NT \log(eN) \log(1 + T/N)}.$$

In particular, the Bayesian regret is of order $\tilde{O}(\sqrt{NT})$.

The theorem shows that preserving the joint posterior allows the variation of revenue under the current belief to be controlled directly. To the best of our knowledge, it is the first $\tilde{O}(\sqrt{NT})$ Bayesian-regret guarantee, uniform over $1 \leq K \leq N$, for Thompson sampling with exact draws from the complete MNL posterior. The order itself is not new: Miao and Chao (2021) obtain it for Thompson sampling on a related decision problem with a different sampling construction, and Oh and Iyengar (2021) attain it with a frequentist UCB construction; nor does the bound dominate the frequentist guarantees of Lee and Oh (2025), which answer a different question through confidence-based optimism. What the posterior argument avoids, relative to the clock analysis, is reducing the history to product-wise comparisons and controlling the posterior odds of missing a useful product.

Theorem 2 says nothing about computation. Grouping compresses repeated assortments without changing the likelihood, but the M sweeps of Algorithm 2 only approximate a draw from π_t , and the theorem bounds neither that error nor how closely the chain tracks a posterior that changes every period. The next subsection develops a lighter approximation, which keeps part of the denominator correction but does not preserve the full posterior dependence.

5.3. Proof of Theorem 2

The proof measures posterior concentration and revenue variation in the same metric: the Fisher information of the choice model in log-coordinates. For an assortment S and attractions x , let $F_S(x)$ be the covariance of the choice indicator e_I , where $I \sim p(S, x)$ is the selected alternative and $e_0 := 0$, defined in Section EC.3.2, so that $h^\top F_S(x) h$ is the variance of h_I , the coordinate of h at the selected alternative, for a direction h with $h_0 = 0$. Every such matrix is comparable to its value at the all-ones point $\mathbf{1}$: by (EC.5), $v_{\min} F_S(\mathbf{1}) \preceq F_S(x) \preceq v_{\min}^{-1} F_S(\mathbf{1})$ for $x \in [v_{\min}, 1]^N$. We may therefore accumulate information at the single reference point $\mathbf{1}$: write $F_t := F_{S_t}(\mathbf{1})$ and $H_t := \kappa I_N + \sum_{\tau=1}^{t-1} F_\tau \succeq \kappa I_N$, with $\kappa := \beta_0$.

Throughout this proof, expectations are under the joint Bayesian law, and $\mathbb{E}_t$ conditions on $\mathcal{F}_{t-1}$. The proof has three steps. Step 1 uses posterior symmetry to reduce regret to revenue sensitivity and bounds that sensitivity by a Fisher width. Step 2 sums the widths through a log-determinant potential. Step 3 combines the two over probability and time.

Step 1 (Posterior Symmetry and Revenue Sensitivity). Conditional on $\mathcal{F}_{t-1}$, v and $\tilde{v}_t$ are independent with common law π_t , so they have the same expected optimal revenue. Since $S_t = S^*(\tilde{v}_t)$,

$$\mathbb{E}_t[R(S^*(v), v) - R(S_t, v)] = \mathbb{E}_t[R(S_t, \tilde{v}_t) - R(S_t, v)] \leq \mathbb{E}_t[|R(S_t, \tilde{v}_t) - R(S_t, v)|]. \quad (9)$$

This is the only posterior-sampling identity the proof uses, and it is what replaces an optimism event: regret becomes the sensitivity of the revenue of a single assortment to a change of attractions, with no probability of a favorable draw left to bound. It is also where the exactness of the draw is spent; an approximate sampler does not satisfy it.

To bound this sensitivity, set $h_t := \log \tilde{v}_t - \log v$. Along the log-linear path between the two attraction vectors, the revenue derivative is the covariance between the realized revenue and the corresponding coordinate of h_t . Cauchy–Schwarz bounds this covariance by one half of the Fisher width, since revenues lie in $[0, 1]$. Integrating along the path and applying the comparison with $\mathbf{1}$ gives the deterministic bound

$$|R(S_t, \tilde{v}_t) - R(S_t, v)| \leq \frac{1}{2\sqrt{v_{\min}}} \sqrt{h_t^\top F_t h_t}.$$

The supporting detail for Step 1 in Section EC.3.2 supplies the differentiation. This bound separates the revenue calculation from the posterior concentration needed to control h_t .

Measure the information in the chosen assortment relative to the information already accumulated by $a_t := \text{tr}(H_t^{-1} F_t)$. Lemma EC.3 supplies a nonnegative random multiplier Z_t with $\mathbb{E}_t[Z_t^2] \leq C v_{\min}^{-1} \log(eN)$, and its simultaneous bound (EC.6) gives $h_t^\top F_t h_t \leq Z_t^2 a_t$. The bound holds for every feasible assortment on the same sample path, so it applies to S_t even though the posterior draw selects that assortment. Combining the two width bounds yields

$$|R(S_t, \tilde{v}_t) - R(S_t, v)| \leq \frac{Z_t}{2\sqrt{v_{\min}}} \sqrt{a_t}. \quad (10)$$

The conditional second moment controls the multiplier Z_t ; it remains to control the cumulative information terms a_t .

Step 2 (Pathwise Fisher Accumulation). Each trace can be charged to the increase of a log-determinant potential. Put $B_t := H_t^{-1/2} F_t H_t^{-1/2}$, so that $a_t = \text{tr} B_t$. Since $\text{tr} F_t \leq 1$ and $H_t \succeq \kappa I_N$, every eigenvalue of B_t lies in $[0, \kappa^{-1}]$, where concavity of $\log(1+z)$ gives $z \leq C_\kappa \log(1+z)$ with $C_\kappa := \kappa^{-1} / \log(1 + \kappa^{-1})$. Summing over the eigenvalues yields $a_t \leq C_\kappa \log \det(I_N + B_t)$.

Because $H_{t+1} = H_t + F_t$, the determinant in this bound is $\det H_{t+1} / \det H_t$, so its logarithm telescopes over time. Using $H_1 = \kappa I_N$, then the arithmetic–geometric mean inequality on the eigenvalues together with $\sum_{t \leq T} \text{tr} F_t \leq T$,

$$\sum_{t=1}^T a_t \leq C_\kappa \log \frac{\det H_{T+1}}{\det H_1} = C_\kappa \log \det \left(I_N + \kappa^{-1} \sum_{t=1}^T F_t \right) \leq C_\kappa N \log \left(1 + \frac{T}{\kappa N} \right). \quad (11)$$

This is the log-determinant potential argument familiar from linear bandits, applied to the Fisher matrices at **1**. The calculation is deterministic given the realized assortment sequence, so adaptivity requires no further condition.

Step 3 (Regret Assembly). Both factors in the one-period bound can now be summed: Step 1 gives $\mathbb{E}[\sum_{t=1}^T Z_t^2] \leq C v_{\min}^{-1} T \log(eN)$, and Step 2 bounds $\sum_{t=1}^T a_t$ pathwise. Summing (9), applying (10), and using Cauchy–Schwarz for the factors Z_t and $\sqrt{a_t}$ over probability and time gives

$$\begin{aligned} \text{Reg}_B(T) &\leq \frac{1}{2\sqrt{v_{\min}}} \mathbb{E} \left[\sum_{t=1}^T Z_t \sqrt{a_t} \right] \\ &\leq \frac{1}{2\sqrt{v_{\min}}} \left(\mathbb{E} \left[\sum_{t=1}^T Z_t^2 \right] \right)^{1/2} \left(\mathbb{E} \left[\sum_{t=1}^T a_t \right] \right)^{1/2} \\ &\leq \frac{C}{v_{\min}} \sqrt{NT \log(eN) \log(1 + T/N)}. \end{aligned}$$

The last inequality uses the two cumulative bounds and $\log(1 + z/\kappa) \leq \max\{1, \kappa^{-1}\} \log(1 + z)$. The remaining constant depends only on $\kappa = \beta_0$, so it is independent of (α_0, N, K, T) , and all dependence on $v_{\min}$ appears in the explicit factor $v_{\min}^{-1}$. ■

6. Mean-Field Filtering

Our third policy, *mean-field filtering*, depicted in algorithmic form in Algorithm 3, sits between the two representations already considered. Like the clock policies, it keeps one independent law per product; unlike them, it does not discard the shared denominator, but replaces it by a correction computed from the current mean attractions. The coupling among the products displayed together is therefore approximated rather than dropped, and in a form that still separates across coordinates. Section 6.1 constructs the representation and the policy it induces; Section 6.2 states a regret bound for a covariance-corrected variant of it.

6.1. The Mean-Field Filtering Policy

Prior and Likelihood. A priori the attractions are independent across products, with $v_i \sim \text{Gamma}(\alpha_0, \beta_0)$ for every $i \in \mathcal{N}$. Unlike in Section 5, these laws are not truncated, which keeps the mean of each attraction equal to the ratio of its two parameters; that ratio is the point at which the policy linearizes, and the truncated laws admit no such expression. Taking logarithms in (8) and collecting the numerators by product, the log-likelihood is

$$\log \mathcal{L}(v \mid \mathcal{H}_{t-1}) = \sum_{i \in \mathcal{N}} V_{i,t} \log v_i - \sum_{\tau=1}^{t-1} \log \left(1 + \sum_{j \in S_\tau} v_j \right), \quad (12)$$

with $V_{i,t}$ the purchase count of Section 4. The first sum separates across products and is already in the form a Gamma prior absorbs. The second does not separate: each of its terms couples the

attractions of all the products displayed in that period, so it cannot be written as a sum of per-product terms and the posterior does not factorize. The approximation below acts on those $t - 1$ terms and leaves the first sum unchanged.

The Approximation. Let μ_i denote the current mean of the approximate law of product i , and $\mu_{S_\tau} := \sum_{j \in S_\tau} \mu_j$ its sum over the assortment offered in period τ , both fixed once S_τ has been selected and before the choice is observed. Since the aggregate attraction of S_τ is unknown, we linearize $\log(1 + x)$ at $x = \mu_{S_\tau}$,

$$\log\left(1 + \sum_{j \in S_\tau} v_j\right) \approx \log(1 + \mu_{S_\tau}) + \frac{1}{1 + \mu_{S_\tau}} \left(\sum_{j \in S_\tau} v_j - \mu_{S_\tau}\right),$$

and drop the terms free of v , which the normalization absorbs. Each period then contributes a separate linear penalty $-v_j/(1 + \mu_{S_\tau})$ for each product $j \in S_\tau$ that it displays. Summing over periods and regrouping by product replaces the second sum of (12) by $\sum_{i \in \mathcal{N}} \Upsilon_{i,t} v_i$, where

$$\Upsilon_{i,t} := \sum_{\tau=1}^{t-1} \frac{\mathbb{1}\{i \in S_\tau\}}{1 + \mu_{S_\tau}}$$

is a display count weighted by the linearization coefficients. Exponentiating what is left of (12) and multiplying by the prior, the approximate posterior $\tilde{\pi}_t$ satisfies

$$\tilde{\pi}_t(v) \propto \prod_{i \in \mathcal{N}} v_i^{\alpha_0 + V_{i,t} - 1} \exp(-(\beta_0 + \Upsilon_{i,t})v_i). \quad (13)$$

Every factor is an unnormalized Gamma density, so the approximation preserves both the independence across products and the Gamma family: purchases accumulate in the shape parameter and weighted displays in the rate parameter, and $\tilde{v}_{i,t} \mid \mathcal{F}_{t-1} \sim \text{Gamma}(\alpha_0 + V_{i,t}, \beta_0 + \Upsilon_{i,t})$ independently across products, with mean $\mu_i = (\alpha_0 + V_{i,t})/(\beta_0 + \Upsilon_{i,t})$, the quantity at which the next period linearizes. The pair $(V_{i,t}, \Upsilon_{i,t})$ therefore summarizes the entire history for product i , and the policy carries $O(1)$ state per product, as the clocks do.

The policy linearizes each new denominator at the means prevailing when the assortment was offered and never revisits it, so (13) is not the minimizer of a global divergence over a family of independent laws. The weight $(1 + \mu_{S_\tau})^{-1}$ is fixed once period τ has passed, and it sets how heavily that display weighs against each product shown. If the products on display were collectively attractive, μ_{S_τ} is large, the penalty is mild, and failing to sell i says little about v_i ; if they were not, the same failure says more. This is the substitution evidence the clocks discard, reduced to one scalar per period.

Algorithm 3 Thompson Sampling via Mean-Field Filtering

- 1: **Input:** horizon T , capacity K , revenues $\{r_i\}_{i=1}^N$, prior (α_0, β_0) .
 - 2: **Initialize:** set $V_i \leftarrow 0$, $\Upsilon_i \leftarrow 0$ and $\mu_i \leftarrow \alpha_0/\beta_0$, $\forall i \in \mathcal{N}$.
 - 3: **for** $t = 1, \dots, T$ **do**
 - 4: Sample $\tilde{v}_i \sim \text{Gamma}(\alpha_0 + V_i, \beta_0 + \Upsilon_i)$ independently, $\forall i \in \mathcal{N}$, offer $S_t \leftarrow S^*(\tilde{v})$, and set $\mu_{S_t} \leftarrow \sum_{j \in S_t} \mu_j$.
 - 5: Observe $I_t \in S_t \cup \{0\}$ and update $V_i \leftarrow V_i + \mathbb{1}\{I_t = i\}$, $\Upsilon_i \leftarrow \Upsilon_i + \mathbb{1}\{i \in S_t\}/(1 + \mu_{S_t})$ and $\mu_i \leftarrow (\alpha_0 + V_i)/(\beta_0 + \Upsilon_i)$, $\forall i \in \mathcal{N}$.
 - 6: **end for**
-

Algorithm 3 collects the procedure. Each period draws independently from the current Gamma laws, offers the optimal assortment under the draw, and evaluates μ_{S_t} at the means in force before the customer is observed; the counts, the exposures, and the means are then updated.

Once S_t is chosen, updating Υ_i and μ_i across the displayed products costs $O(|S_t|)$ time, so a period costs N independent gamma draws, one call to the oracle $S^*(\tilde{v})$, and an update confined to the assortment. Nothing beyond the two statistics per product is retained, in contrast with the sweeps of Algorithm 2, which revisit the record of past assortments before every decision.

Relative to the filtered-likelihood policies, a display updates i even when the sale goes to another displayed product, which is precisely the evidence the filtered likelihood discards; relative to the complete posterior, the correction reaches i only through the aggregate mean μ_{S_t} and leaves the coordinates independent.

6.2. Regret Bound for a Covariance-Corrected Variant

We prove no regret bound for Algorithm 3, which linearizes each denominator once and never revisits it, so that its approximate law minimizes no global divergence. The variant below admits one. It keeps the Gamma family, the mean recursion, and the $O(1)$ state per product, and corrects the sampling center and dispersion through two additions. The first certifies the sampling center against the product-versus-outside comparisons that the clocks of Section 4 already record: those counts yield an interval that covers v_i with high probability, and projecting $\mu_{i,t}$ onto it and shifting to its upper end gives a correction $m_{i,t}$ that lies above v_i and contracts onto it as the local clock advances. The second accumulates one Fisher statistic $J_{i,t}$ per product from the fitted choice probabilities of the assortments that displayed it, and, for a fixed $c \geq 1$, draws independently for $i \in \mathcal{N}$

$$\tilde{v}_{i,t} \sim \text{Gamma}\left(\frac{J_{i,t}}{c}, \frac{J_{i,t}}{cm_{i,t}}\right).$$

The covariance correction admits a fixed-assortment calibration, derived in Section EC.6.3 of the e-companion. Suppose a nonempty assortment S is offered in every period, and keep N , S , the prior, and the tuning parameters fixed. Assume $v_i \in (v_{\min}, 1)$ for $i \in S$; the raw-filter calculation also assumes convergence of its means. Write $v_S := (v_i)_{i \in S}$ and $D := 1 + \sum_{i \in S} v_i$. The leading conditional covariance terms after t periods of Algorithm 3, the corrected draw, and the complete posterior are, respectively,

$$\frac{D}{t} \text{diag}(v_S), \quad \frac{cD^2}{t} \text{diag}(v_S), \quad \frac{D}{t} \{\text{diag}(v_S) + v_S v_S^\top\}.$$

Under these conditions, the ratio of each corrected variance to its raw counterpart tends to cD . What the mean-field law omits is the rank-one term $v_S v_S^\top$, the positive coupling among products displayed together, which no product-wise law can carry. Since $v_S v_S^\top \preceq (\sum_{i \in S} v_i) \text{diag}(v_S)$, the corrected leading covariance dominates that of the complete posterior for $c \geq 1$. At $c = 1$, their leading variances of the aggregate attraction $\sum_{i \in S} v_i$ coincide.

The certificate covers the sampling center with finite-sample probability, so the guarantee below requires no concentration of the unprojected mean recursion $\mu_{i,t}$.

THEOREM 3 (Regret of Covariance-Corrected Mean-Field Sampling). *Suppose that $v \in [v_{\min}, 1]^N$ and $\alpha_0, \beta_0 > 0$, and run the covariance-corrected policy of Algorithm 4 in Section EC.6 with fixed $\lambda_0 \geq c \geq 1$ and initial Fisher statistics $J_{i,1} = \lambda_0$ for $i \in \mathcal{N}$. There exists $C < \infty$, depending only on $v_{\min}$, such that*

$$\text{Reg}(T, v) \leq CK \sqrt{cNT \log(2NT^2)} + C.$$

In particular, for fixed c , the regret is of order $\tilde{O}(K\sqrt{NT})$, uniformly over $1 \leq K \leq N$.

The rate is gap free. Section EC.6 of the e-companion states the policy in full, derives the certificate and the Fisher statistic, supplies the fixed-assortment calibration quoted above, and proves the theorem. The guarantee above covers the corrected variant, not Algorithm 3. Section 7 reports the practical performance of Algorithm 3.

7. Numerical Results

This section compares the numerical performance of the proposed policies and relevant benchmarks beyond the finite-time upper bound guarantees established above, when such guarantees exist. The experiments compare regret and computation across a baseline design, two structural stress tests, and variations in the number of products and capacity.

7.1. Experimental Design

Benchmarks and Performance Metric. In addition to our three policies, the asynchronous-clock policy, the Gibbs approximation to the MNL posterior, and mean-field filtering, we run the epoch-based independent-BetaPrime construction of Agrawal et al. (2026); these four are the focal policies below. Two external benchmarks complete the comparison: Algorithm 2 of Agrawal et al. (2026) and the outside-option-pivot AOA-RBPL policy of Saha and Gaillard (2025), referred to as the Saha–Gaillard UCB. For Algorithm 2 we use the authors’ empirical calibration $\alpha = \beta = 1$ and project negative raw attractions onto $\mathbb{R}_+$ before calling the MNL oracle, without resampling the Gaussian shocks.¹

Performance is measured in terms of the cumulative pseudo-regret (which we estimate by Monte Carlo),

$$\mathcal{R}(T) = \sum_{t=1}^T [R(S^*(v), v) - R(S_t, v)].$$

Instances and Priors. We use $\alpha_0 = 1$ and $\beta_0 = 3$ for the prior of every Thompson-sampling policy’s BetaPrime or Gamma law; the attraction generators and the full-likelihood implementation additionally use $v_{\min} = 0.05$. Baseline attractions are independent draws from the Gamma law truncated to $[0.05, 1]$, and revenues are independent uniform draws from $[0.05, 1]$. The baseline design uses $N = 1000$, $K = 10$, and $T = 200\,000$, with 30 instances and 10 replications per instance.

The scaling design varies $N \in \{100, 250, 500, 1000\}$ at $K = 10$ and $K \in \{1, 5, 10, 20, 50\}$ at $N = 100$, with 15 instances and 5 replications. Instances with fewer products are nested prefixes of one master instance, and capacities are compared on the same revenue and attraction vectors.

Uncertainty and Computation. Replications are averaged within each instance, and 95% intervals are the 2.5%–97.5% percentiles of 2000 bootstrap resamples of instances, with shared resampling indices preserving pairing across policies; tables report mean [lower, upper], except the scaling table, which reports means and standard errors. All algorithms are implemented in Julia 1.12.4 (Bezanson et al. 2017); reported times cover assortment selection and updating, in seconds per 1,000 periods. We calibrate the Gibbs implementation to $M = 50$ sweeps per decision, the smallest within 5% mean relative regret of a 1000-sweep reference, an 18.8-fold speedup, and use it throughout. Section 7.2 records the calibration, timing protocol, and execution provenance.

¹ This completion, unspecified in the original algorithm, affects approximately 0.10% of regular sampling epochs at $K = 10$, but 50.03% at $K = 1$; Section 7.2 gives the complete profile across capacities.

7.2. Implementation and Calibration

Instance Construction. The small-gap generator uses baseline attractions and revenues from the upper 15% of $[0.05, 1]$ for a random K -product target, the lower 5% otherwise, then tunes one fringe challenger. Substitution uses the upper 3% for a random 40-product pool, the lower 5% otherwise.

Execution and Timing. Reported times exclude metrics, checkpointing, and input/output. Simulation stops at T , and omitting an unfinished epoch’s terminal update affects neither the recorded actions nor the regret. We complete Algorithm 2 (Agrawal et al. 2026) by projecting negative raw attractions onto $\mathbb{R}_+$, which preserves feasible draws and Gaussian randomization; since each epoch applies the coordinate-wise maximum of K independent standard normals, one per unit of display capacity, the projection binds less often as capacity grows, affecting at least one coordinate in 50.03%, 3.14%, approximately 0.10%, and $5.1 \times 10^{-5}\%$ of regular sampling epochs at $K = 1, 5, 10, 20$, respectively, and never at $K = 50$.

Gibbs State and Calibration. After processing S_T , distinct nonempty assortments average 4178 (range 1696–9569), 8.4 (4–11), and 1645 (853–3209) over 300 Gibbs trajectories per family: baseline, small gaps, and substitution, respectively. Calibration uses ten paired instances, one trajectory per tabulated sweep count, sharing customer streams. For instance i ’s mean regret per period $a_{i,M}$, we average $2(a_{i,M} - a_{i,1000})/(a_{i,M} + a_{i,1000})$ across instances and select the smallest M whose average has magnitude at most 5%. (The 90% intervals only describe variability).

Table 1 Gibbs calibration ($N = 1000$, $K = 10$, $T = 10000$). Bold marks the selected sweep count.

M	Mean Regret per Period	Standard Error of Regret	Relative Difference from $M = 1000$ (%)	90% CI for the Difference (%)	Policy Time per 1,000 Periods (s)
20	0.008838	0.000905	9.02	[-1.80, 19.84]	0.73
50	0.007844	0.000522	-0.01	[-9.29, 9.27]	1.69
100	0.008584	0.000808	7.30	[-3.82, 18.43]	3.51
200	0.007858	0.000668	-1.18	[-13.99, 11.63]	6.44
500	0.007812	0.000723	-2.75	[-10.88, 5.38]	21.64
1000	0.008058	0.000829	0.00	[0.00, 0.00]	31.81

7.3. Baseline Comparison

Figure 1 and Table 2 show the regret–computation trade-off: Gibbs records the lowest regret, followed by mean-field filtering at a fraction of its cost. The two clocks attain nearly identical regret despite changing their decisions in 7.4% and 60.4% of periods, respectively. Both external benchmarks incur substantially more regret than the four focal policies.

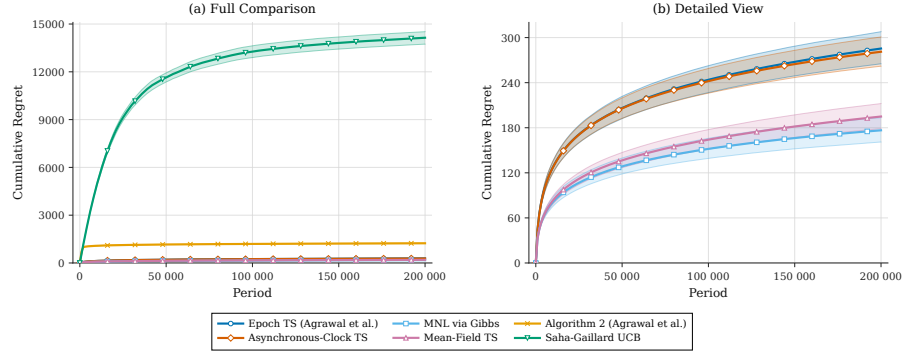


Figure 1 Cumulative-regret trajectories in the baseline design. Panel (a) presents the full comparison; panel (b) enlarges the four policies with the lowest regret.

Table 2 Cumulative regret at the end of the horizon, mean regret increment per period over the last 50,000 rounds, and terminal metrics in the baseline design. Times are seconds per 1,000 periods.

Policy	Regret		Optimal selection	Switch rate	Time (s)
	Cumulative	Mean per Period			
Epoch TS (Agrawal et al.)	285.4 [265.1, 307.9]	$3.54 \cdot 10^{-4}$ [$3.10 \cdot 10^{-4}$, $3.97 \cdot 10^{-4}$]	41.3% [34.9%, 47.9%]	7.4% [6.7%, 8.0%]	0.0141
Asynchronous-Clock TS	281.2 [261.9, 300.8]	$3.29 \cdot 10^{-4}$ [$2.98 \cdot 10^{-4}$, $3.61 \cdot 10^{-4}$]	42.0% [35.7%, 48.2%]	60.4% [55.4%, 65.1%]	0.111
MNL via Gibbs	176.6 [161.0, 193.4]	$2.06 \cdot 10^{-4}$ [$1.77 \cdot 10^{-4}$, $2.33 \cdot 10^{-4}$]	48.1% [41.0%, 55.3%]	51.2% [45.9%, 56.5%]	5.65
Mean-Field TS	194.9 [178.8, 212.2]	$2.66 \cdot 10^{-4}$ [$2.28 \cdot 10^{-4}$, $3.05 \cdot 10^{-4}$]	44.8% [36.8%, 52.7%]	53.5% [48.0%, 58.9%]	0.0692
Algorithm 2 (Agrawal et al.)	1 236.1 [1 222.4, 1 250.1]	$3.29 \cdot 10^{-4}$ [$3.02 \cdot 10^{-4}$, $3.57 \cdot 10^{-4}$]	47.8% [40.9%, 54.6%]	4.6% [4.3%, 4.9%]	0.0115
Saha-Gaillard UCB	14 138.4 [13 736.6, 14 518.9]	$6.55 \cdot 10^{-3}$ [$5.78 \cdot 10^{-3}$, $7.34 \cdot 10^{-3}$]	5.0% [4.0%, 6.0%]	12.7% [12.2%, 13.2%]	0.0268

7.4. Structural Stress Tests

The stress tests ask whether the baseline ordering persists when the two best assortments are nearly indifferent or when many nearby assortments are substitutes. We build each family directly: the small-gap generator tunes a one-product-swap challenger and certifies that it and the target are the two best assortments, with revenue gap in $[10^{-5}, 5 \times 10^{-4}]$; the substitution generator uses a 40-product high-revenue pool and certifies at least 20 distinct one-product swaps from the optimum with losses at most 0.01. Both designs use $N = 1000$, $K = 10$, and $T = 200\,000$, with 30 instances and 10 replications per instance; Section 7.2 gives the construction. Figures 2–3 show the trajectories; Table 3 reports both sets of metrics.

Small Gaps. Gibbs has the lowest regret, with paired differences of 2.26, 2.73, and 6.08 from the asynchronous policy, the epoch benchmark, and mean-field filtering, all with 95% intervals excluding zero. Mean-field filtering has the largest regret among the four, but its differences from the clocks are inconclusive. The external benchmarks select the optimum more often than the focal

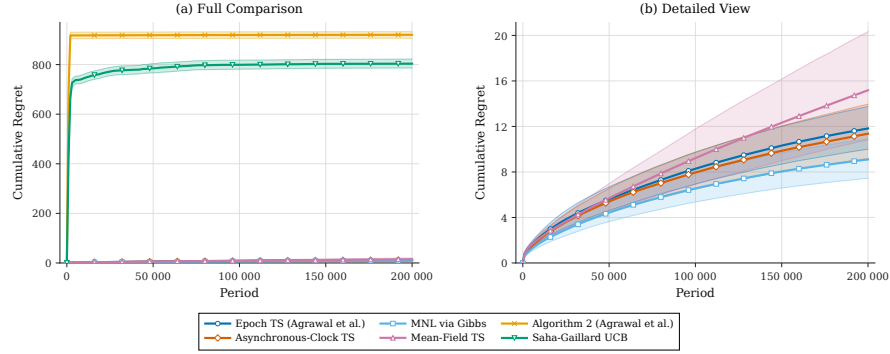


Figure 2 Cumulative regret when the gap between the two best assortments lies in $[10^{-5}, 5 \times 10^{-4}]$ ($N = 1000$, $K = 10$, $T = 200\,000$; 30 instances, 10 replications each).

policies yet incur much greater regret: their less frequent errors are more costly. Selection frequency alone therefore does not rank decision quality.

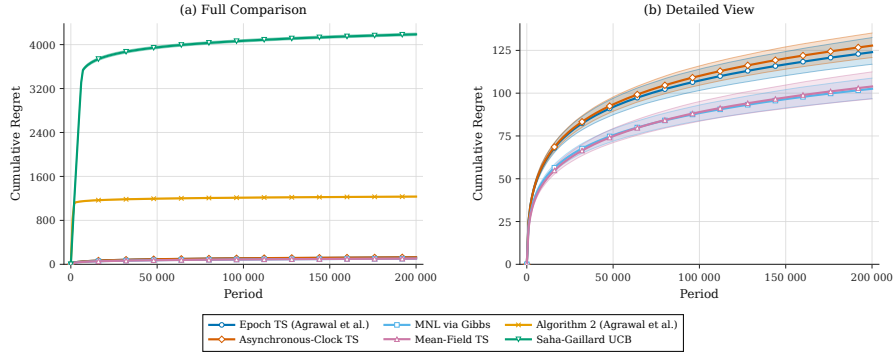


Figure 3 Cumulative regret in the substitution family, built around a structured 40-product high-revenue pool ($N = 1000$, $K = 10$, $T = 200\,000$; 30 instances, 10 replications each).

Substitution. Gibbs and mean-field filtering are not statistically separated: their paired regret difference is 1.35, with 95% interval $[-2.48, 5.77]$; their optimal-selection frequencies do not separate them either. Both outperform the clocks, whose difference remains inconclusive. Computation separates Gibbs from mean-field filtering more clearly: 3.67 against 0.0492 seconds per 1,000 periods. Thus the evidence favors Gibbs over mean-field filtering in the small-gap family, but not in the substitution family; it does not establish equivalence.

Table 3 Cumulative regret, optimal-assortment selection frequency, and policy time in the two stress families ($N = 1000$, $K = 10$, $T = 200\,000$; 30 instances, 10 replications each). Times are seconds per 1,000 periods.

Policy	(a) Small gaps			(b) Substitution		
	Regret	Optimal selection	Time (s)	Regret	Optimal selection	Time (s)
Epoch TS (Agrawal et al.)	11.8 [10.0, 13.7]	74.5% [69.9%, 78.8%]	0.0219	124.0 [116.9, 132.5]	57.5% [50.2%, 64.8%]	0.0116
Asynchronous-Clock TS	11.4 [9.2, 14.0]	75.2% [69.9%, 80.1%]	0.0893	127.8 [120.8, 135.1]	55.5% [49.1%, 62.4%]	0.0843
MNL via Gibbs	9.1 [7.4, 10.9]	79.9% [75.2%, 84.1%]	0.955	102.6 [96.8, 108.8]	59.0% [52.3%, 66.2%]	3.67
Mean-Field TS	15.2 [10.8, 20.3]	69.9% [61.8%, 77.5%]	0.0632	103.9 [96.6, 112.5]	58.2% [51.3%, 65.4%]	0.0492
Algorithm 2 (Agrawal et al.)	920.0 [906.5, 932.6]	93.6% [91.2%, 95.6%]	0.0156	1 233.0 [1 228.4, 1 238.2]	56.3% [49.0%, 64.1%]	0.0076
Saha-Gaillard UCB	803.6 [786.2, 822.0]	90.8% [87.1%, 93.8%]	0.0076	4 186.9 [4 163.7, 4 210.8]	19.7% [16.2%, 23.6%]	0.0107

7.5. Scaling with the Number of Products and Capacity

Table 4 shows how catalog size and capacity separately affect finite-horizon performance.

Table 4 Cumulative regret and policy time across combinations of N and K . Regret cells report the mean and, in parentheses, its standard error.

N	K	Cumulative Regret						Policy Time per 1,000 Periods (s)					
		Agrawal Epoch	Async MNL–Gibbs	Mean-Field	Algorithm 2 Agrawal et al.	Saha–Gaillard UCB		Agrawal Epoch	Async MNL–Gibbs	Mean-Field	Algorithm 2 Agrawal et al.	Saha–Gaillard UCB	
100	1	471.4	493.2	250.2	402.1	423.6		5 228.3					
		(19.7)	(26.7)	(29.2)	(64.8)	(21.8)		(117.9)					
100	5	122.6	125.1	102.3	98.1	195.7		2 845.2					
		(10.6)	(9.1)	(12.4)	(7.5)	(9.2)		(112.1)					
100	10	50.6	50.1	41.8	45.6	142.6		2 154.2					
		(4.1)	(5.2)	(4.3)	(5.6)	(5.6)		(94.0)					
100	20	5.4	5.6	4.5	13.4	101.4		1 925.6					
		(1.1)	(1.3)	(0.9)	(5.1)	(1.4)		(63.6)					
100	50	1.4	1.5	1.6	20.4	114.4		1 966.2					
		(0.2)	(0.2)	(0.3)	(7.5)	(3.2)		(66.6)					
250	10	115.1	113.5	83.5	88.8	342.5		4 547.1					
		(6.9)	(6.2)	(3.7)	(4.8)	(4.6)		(110.1)					
500	10	186.1	180.3	121.1	135.6	649.4		8 115.2					
		(12.4)	(13.4)	(8.3)	(9.4)	(9.7)		(209.2)					
1000	10	288.0	281.2	178.3	191.9	1 238.9		14 168.2					
		(18.1)	(16.1)	(12.3)	(15.5)	(13.4)		(274.0)					

As N grows from 100 to 1000, $\mathcal{R}(T)/\sqrt{NT}$ stays within $[0.0093, 0.0126]$ for Gibbs and $[0.0102, 0.0136]$ for mean-field filtering, descriptively consistent with $\sqrt{N}$ growth. Along the capacity axis, regret decreases for most policies as the constraint slackens: all 15 optimal assortments fill the capacity up to $K = 10$, 13 do at $K = 20$, and none at $K = 50$. Mean-field filtering has the largest mean regret among the focal policies at $K = 20$ and $K = 50$, and its mean regret increases between these capacities, though this does not isolate its aggregate denominator correction as the cause. Algorithm 2 (Agrawal et al. 2026) records the lowest policy time in every cell, but higher regret than all four focal policies in seven of eight; Gibbs remains the most computationally demanding.

8. Concluding Remarks

We have proposed three policies, differing in how much of the MNL likelihood each retains. The first is the asynchronous-clock policy with independent BetaPrime draws. The linear-fractional revenue structure identifies when correcting an omitted product would improve the sampled optimum; scalar posterior odds and the accumulated product-versus-outside comparisons control how long such a

product can remain unselected. Under the stated attraction conditions, this yields gap-free worst-case regret $\tilde{O}(\sqrt{NKT})$ without variance inflation. The contribution is to the sampling mechanism; UCB and correlated sampling retain sharper capacity dependence. The second concerns the value of retaining the complete likelihood that those clocks discard: a Gamma–Exponential augmentation turns the shared denominator into product-wise conditionals and yields a Gibbs sampler for the joint posterior, and ideal Thompson sampling from that posterior attains Bayesian regret $\tilde{O}(\sqrt{NT})$, also uniform in capacity, through posterior symmetry and pairwise Fisher concentration. The third places these two representations alongside mean-field filtering, which corrects the shared denominator to first order at $O(1)$ memory per product, and shows that a corrected form of that filter, with a certified sampling center and corrected dispersion, attains gap-free $\tilde{O}(K\sqrt{NT})$ regret.

Our numerical experiments describe a statistical–computational trade-off: the fifty-sweep Gibbs implementation attains the lowest regret in most designs and has the highest computational cost. Mean-field filtering is not statistically separated from Gibbs in the substitution design, but has higher regret in the small-gap design and the two largest-capacity cells. It is therefore a lower-cost, instance-dependent alternative; however, these comparisons do not identify the cause of the performance differences. Neither comparison is covered by a guarantee for the implementation tested: the Gibbs implementation approximates the posterior draw, and the mean-field bound covers a covariance-corrected variant rather than the filter as run.

Two questions remain open. The extra factor $\sqrt{K}$ in the asynchronous-clock regret bound is not required by the lower bound of Chen and Wang (2018) in its regime; whether it can be removed by sharpening the coordinate-repair analysis or is intrinsic to independent BetaPrime sampling remains unresolved. The second is whether the correction is necessary: a corrected form of mean-field filtering attains gap-free $\tilde{O}(K\sqrt{NT})$ regret, and we do not know whether the filter as implemented, which neither corrects its dispersion nor certifies its center, admits a bound of any order.

References

- Shipra Agrawal and Navin Goyal. Near-optimal regret bounds for thompson sampling. *Journal of the ACM*, 64(5):30:1–30:24, 2017. doi: 10.1145/3088510.
- Shipra Agrawal, Vashist Avadhanula, Vineet Goyal, and Assaf Zeevi. MNL-bandit: A dynamic learning approach to assortment selection. *Operations Research*, 67(5):1453–1485, 2019. doi: 10.1287/opre.2018.1832.
- Shipra Agrawal, Vashist Avadhanula, Vineet Goyal, and Assaf Zeevi. Thompson sampling for the multinomial logit bandit. *Mathematics of Operations Research*, 51(1):568–590, 2026. doi: 10.1287/moor.2020.0096.

- Peter Auer, Nicolò Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47(2–3):235–256, 2002. doi: 10.1023/A:1013689704352.
- Jeff Bezanson, Alan Edelman, Stefan Karpinski, and Viral B. Shah. Julia: A fresh approach to numerical computing. *SIAM Review*, 59(1):65–98, 2017. doi: 10.1137/141000671.
- Sergey G. Bobkov and Michel Ledoux. From Brunn–Minkowski to Brascamp–Lieb and to logarithmic Sobolev inequalities. *Geometric and Functional Analysis*, 10(5):1028–1052, 2000. doi: 10.1007/PL00001645.
- Felipe Caro and Jérémie Gallien. Dynamic assortment with demand learning for seasonal consumer goods. *Management Science*, 53(2):276–292, 2007. doi: 10.1287/mnsc.1060.0613.
- François Caron and Arnaud Doucet. Efficient Bayesian inference for generalized Bradley–Terry models. *Journal of Computational and Graphical Statistics*, 21(1):174–196, 2012. doi: 10.1080/10618600.2012.638220.
- A. Charnes and W. W. Cooper. Programming with linear fractional functionals. *Naval Research Logistics Quarterly*, 9(3–4):181–186, 1962. doi: 10.1002/nav.3800090303.
- Xi Chen and Yining Wang. A note on a tight lower bound for capacitated MNL-bandit assortment selection models. *Operations Research Letters*, 46(5):534–537, 2018. doi: 10.1016/j.orl.2018.08.006.
- Tianyuan Jin, Heyang Zhao, Vincent Y. F. Tan, and Quanquan Gu. Avoiding $\exp(k^*)$ scaling for thompson sampling in combinatorial semi-bandits: From multiple seeds to a single seed. In *Proceedings of the 39th Conference on Learning Theory (COLT)*, volume 336 of *Proceedings of Machine Learning Research*, pages 3825–3855, 2026. URL <https://proceedings.mlr.press/v336/jin26a.html>.
- Robert E. Kass, Luke Tierney, and Joseph B. Kadane. The validity of posterior expansions based on Laplace’s method. In Seymour Geisser, James S. Hodges, S. James Press, and Arnold Zellner, editors, *Bayesian and Likelihood Methods in Statistics and Econometrics*, pages 473–488. North-Holland, Amsterdam, 1990. URL <https://www.stat.cmu.edu/~kass/papers/validity.pdf>.
- A. Gürhan Kök, Marshall L. Fisher, and Ramnath Vaidyanathan. Assortment planning: Review of literature and industry practice. In *Retail Supply Chain Management*, pages 99–153. Springer, Boston, MA, 2008. doi: 10.1007/978-0-387-78902-6.6.
- Tor Lattimore and Csaba Szepesvári. *Bandit Algorithms*. Cambridge University Press, Cambridge, 2020. doi: 10.1017/9781108571401.
- Joongkyu Lee and Min-Hwan Oh. Improved online confidence bounds for multinomial logistic bandits. In *Proceedings of the 42nd International Conference on Machine Learning (ICML)*, volume 267 of *Proceedings of Machine Learning Research*, pages 33576–33615, 2025. URL <https://proceedings.mlr.press/v267/lee25z.html>.

- László Lovász and Santosh Vempala. The geometry of logconcave functions and sampling algorithms. *Random Structures & Algorithms*, 30(3):307–358, 2007. doi: 10.1002/rsa.20135. URL <https://faculty.cc.gatech.edu/~vempala/papers/logcon.pdf>.
- R. Duncan Luce. *Individual Choice Behavior: A Theoretical Analysis*. Wiley, New York, 1959.
- Eric Mazumdar, Aldo Pacchiano, Yi-An Ma, Michael I. Jordan, and Peter L. Bartlett. On approximate thompson sampling with Langevin algorithms. In *Proceedings of the 37th International Conference on Machine Learning (ICML)*, volume 119 of *Proceedings of Machine Learning Research*, 2020. URL <https://proceedings.mlr.press/v119/mazumdar20a.html>.
- Daniel McFadden. Modeling the choice of residential location. *Transportation Research Record*, 673:72–77, 1978.
- Sentao Miao and Xiuli Chao. Dynamic joint assortment and pricing optimization with demand learning. *Manufacturing & Service Operations Management*, 23(2):525–545, 2021. doi: 10.1287/msom.2019.0857.
- Min-hwan Oh and Garud Iyengar. Thompson sampling for multinomial logit contextual bandits. In *Advances in Neural Information Processing Systems 32*, 2019. URL https://papers.nips.cc/paper_files/paper/2019/hash/36d7534290610d9b7e9abed244dd2f28-Abstract.html.
- Min-hwan Oh and Garud Iyengar. Multinomial logit contextual bandits: Provable optimality and practicality. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(10):9205–9213, 2021. doi: 10.1609/aaai.v35i10.17111.
- My Phan, Yasin Abbasi-Yadkori, and Justin Domke. Thompson sampling and approximate inference. In *Advances in Neural Information Processing Systems 32*, 2019. URL https://papers.nips.cc/paper_files/paper/2019/hash/f3507289cfdc8c9ae93f4098111a13f9-Abstract.html.
- Paat Rusmevichientong, Zuo-Jun Max Shen, and David B. Shmoys. Dynamic assortment optimization with a multinomial logit choice model and capacity constraint. *Operations Research*, 58(6):1666–1680, 2010. doi: 10.1287/opre.1100.0866.
- Daniel Russo and Benjamin Van Roy. Learning to optimize via posterior sampling. *Mathematics of Operations Research*, 39(4):1221–1243, 2014. doi: 10.1287/moor.2014.0650.
- Daniel Russo, Benjamin Van Roy, Abbas Kazerouni, Ian Osband, and Zheng Wen. A tutorial on thompson sampling. *Foundations and Trends in Machine Learning*, 11(1):1–96, 2018. doi: 10.1561/22000000070.
- Aadirupa Saha and Pierre Gaillard. Finally rank-breaking conquers MNL bandits: Optimal and efficient algorithms for MNL assortment. In *Proceedings of the 13th International Conference on Learning Representations (ICLR)*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/file/b9a4d7b88a41652c63962ebcc21701b7-Paper-Conference.pdf.

- Denis Sauré and Assaf Zeevi. Optimal dynamic assortment planning with demand learning. *Manufacturing & Service Operations Management*, 15(3):387–404, 2013. doi: 10.1287/msom.2013.0429.
- Kalyan Talluri and Garrett van Ryzin. Revenue management under a general discrete choice model of consumer behavior. *Management Science*, 50(1):15–33, 2004. doi: 10.1287/mnsc.1030.0147.
- Martin A. Tanner and Wing Hung Wong. The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82(398):528–540, 1987. doi: 10.1080/01621459.1987.10478458.
- William R. Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3/4):285–294, 1933. doi: 10.2307/2332286.
- Kenneth E. Train. *Discrete Choice Methods with Simulation*. Cambridge University Press, Cambridge, 2003. doi: 10.1017/CBO9780511753930.
- Garrett van Ryzin and Siddharth Mahajan. On the relationship between inventory costs and variety benefits in retail assortments. *Management Science*, 45(11):1496–1509, 1999. doi: 10.1287/mnsc.45.11.1496.
- Siwei Wang and Wei Chen. Thompson sampling for combinatorial semi-bandits. In *Proceedings of the 35th International Conference on Machine Learning (ICML)*, volume 80 of *Proceedings of Machine Learning Research*, pages 5114–5122, 2018. URL <https://proceedings.mlr.press/v80/wang18a.html>.
- Jiaqi Yang. Fully gap-dependent bounds for multinomial logit bandit. In *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 130 of *Proceedings of Machine Learning Research*, pages 199–207, 2021. URL <https://proceedings.mlr.press/v130/yang21a.html>.
- Raymond Zhang and Richard Combes. Thompson sampling for combinatorial bandits: Polynomial regret and mismatched sampling paradox. In *Advances in Neural Information Processing Systems 37*, 2024. doi: 10.52202/079017-2839. URL https://papers.neurips.cc/paper_files/paper/2024/file/a2e707354da36956945dbb288efe82b3-Paper-Conference.pdf.
- Zixin Zhong, Wang Chi Cheung, and Vincent Y. F. Tan. Thompson sampling algorithms for cascading bandits. *Journal of Machine Learning Research*, 22(218):1–66, 2021. URL <https://www.jmlr.org/papers/v22/20-447.html>.

E-Companion: Constructions and Supporting Proofs

Sections EC.1 and EC.2 establish the two representations the policies sample from: the BetaPrime update that makes the asynchronous clocks conjugate, and the augmented MNL posterior underlying the Gibbs sampler. Section EC.3 collects the setup, auxiliary results, and supporting calculations that the proofs of Theorems 1 and 2 in the main text set aside, and Sections EC.4 and EC.5 prove those auxiliary results. Section EC.6 develops the covariance-corrected mean-field policy of Section 6 and proves Theorem 3. The probability space, the filtrations $\mathcal{F}_t$ and $\bar{\mathcal{F}}_t$, and the notational conventions are those fixed in Section EC.1.

EC.1. BetaPrime Conjugacy for Asynchronous Clocks

This section proves Proposition 1. A factorization of the one-period MNL kernel yields the required Bernoulli representation and the BetaPrime update. Section EC.3.1 uses the same representation to establish concentration uniformly over the possible clock counts.

Throughout the e-companion, $\mathcal{F}_t = \sigma(\mathcal{H}_t)$ is the history filtration of Section 3 and $\bar{\mathcal{F}}_t := \mathcal{F}_{t-1} \vee \sigma(S_t)$ is the pre-choice σ -field of period t , so that S_t is $\bar{\mathcal{F}}_t$ -measurable but need not be $\mathcal{F}_{t-1}$ -measurable.

Proof of Proposition 1. Let $i \in \mathcal{N}$ and set $\theta_i := v_i / (1 + v_i)$. The Bernoulli sequence is indexed by recorded comparisons. A zero records an outside-option choice while i is offered; periods without a comparison leave the next entry unrevealed. To construct this representation, we decide whether to record a comparison before revealing its outcome. Whenever $i \in S_t$, the MNL kernel factors as

$$\mathbb{P}_v(I_t \in \{i, 0\} \mid \bar{\mathcal{F}}_t) = p_i(S_t, v) + p_0(S_t, v), \quad \mathbb{P}_v(I_t = i \mid I_t \in \{i, 0\}, \bar{\mathcal{F}}_t) = \frac{p_i(S_t, v)}{p_i(S_t, v) + p_0(S_t, v)} = \theta_i.$$

When $S_t \setminus \{i\}$ is nonempty, the complementary conditional distribution assigns probability $v_j / \sum_{\ell \in S_t \setminus \{i\}} v_\ell$ to each $j \in S_t \setminus \{i\}$. If $S_t = \{i\}$, the complementary event has probability zero.

Conditional on v , take an infinite i.i.d. Bernoulli(θ_i) sequence $(Z_{i,k})_{k \geq 1}$, independent of all auxiliary randomization used to implement the policy and generate the remaining customer choices. Let $n_{i,t} := V_{i,t} + N_{i,t}$. We construct the history by induction, preserving three properties before each period t :

- (i) the constructed history has the original adaptive MNL law;
- (ii) exactly the prefix $Z_{i,1}, \dots, Z_{i,n_{i,t}}$ has been revealed, and $V_{i,t} = \sum_{k=1}^{n_{i,t}} Z_{i,k}$;

- (iii) conditional on everything revealed so far, the unused tail $(Z_{i,n_{i,t}+k})_{k \geq 1}$ is an i.i.d. Bernoulli(θ_i) sequence.

Base case. Before period 1, the history is empty and $V_{i,1} = n_{i,1} = 0$, so (i)–(iii) hold.

Induction hypothesis. Let $t \geq 1$ and suppose (i)–(iii) hold before period t .

Inductive step. Draw S_t from the policy's original conditional rule, using the constructed history and independent auxiliary randomization. If $i \notin S_t$, generate I_t from the MNL kernel using independent customer randomization and leave the sequence untouched. If $i \in S_t$, use randomization independent of the history and of the unused tail to accept a comparison with probability $p_i(S_t, v) + p_0(S_t, v)$. Upon acceptance, reveal $Z_{i,n_{i,t}+1}$ and set $I_t = i$ if it equals one and $I_t = 0$ otherwise. Upon rejection, draw from the complementary distribution without revealing a binary outcome.

The factorization above gives the original MNL kernel when $i \in S_t$; otherwise, that kernel is used directly. Together with the unchanged policy rule, this preserves (i) before period $t + 1$.

A comparison is recorded exactly upon acceptance, and its outcome is the revealed entry. Thus the count increments are

$$n_{i,t+1} - n_{i,t} = \mathbb{1}\{i \in S_t, I_t \in \{i, 0\}\}, \quad V_{i,t+1} - V_{i,t} = (n_{i,t+1} - n_{i,t})Z_{i,n_{i,t}+1}.$$

Since the clock advances by at most one, these identities give

$$V_{i,t+1} = \sum_{k=1}^{n_{i,t}} Z_{i,k} + (n_{i,t+1} - n_{i,t})Z_{i,n_{i,t}+1} = \sum_{k=1}^{n_{i,t+1}} Z_{i,k}.$$

Exactly the first $n_{i,t+1}$ entries have now been revealed, proving (ii).

Finally, selection and acceptance use randomization independent of the unused tail. A period without a comparison leaves that tail untouched; a comparison reveals only its first entry. By (iii) at period t , the remaining tail therefore retains its i.i.d. Bernoulli(θ_i) law conditional on everything revealed through period t , proving (iii) at period $t + 1$.

Induction now gives the claimed Bernoulli-prefix representation. The unused sequence remains available even if the policy stops offering i , so the clock need not advance indefinitely.

For the conjugate update, the product of the recorded Bernoulli factors is

$$\theta_i^{V_{i,t}} (1 - \theta_i)^{N_{i,t}} = \left(\frac{v_i}{1 + v_i} \right)^{V_{i,t}} \left(\frac{1}{1 + v_i} \right)^{N_{i,t}} = \frac{v_i^{V_{i,t}}}{(1 + v_i)^{V_{i,t} + N_{i,t}}}.$$

Multiplying by the BetaPrime(α_0, β_0) prior kernel yields

$$(v_i^{\alpha_0-1} (1 + v_i)^{-(\alpha_0+\beta_0)}) \cdot \frac{v_i^{V_{i,t}}}{(1 + v_i)^{V_{i,t} + N_{i,t}}} = v_i^{\alpha_0+V_{i,t}-1} (1 + v_i)^{-(\alpha_0+V_{i,t}+\beta_0+N_{i,t})},$$

the kernel of the prescribed BetaPrime($\alpha_0 + V_{i,t}, \beta_0 + N_{i,t}$) update, for every $t \leq T$. ■

EC.2. The Augmented MNL Posterior

This section proves Proposition 2 and then verifies that grouping repeated assortments preserves the complete Gibbs transition kernel, as stated in Corollary EC.1.

Proof of Proposition 2. Let $t \geq 1$ and condition on the realized history $\mathcal{H}_{t-1}$. Write $y = (y_1, \dots, y_{t-1})$. The augmentation uses the reciprocal identity $a^{-1} = \int_0^\infty e^{-ay} dy$, applied in period $\tau < t$ with $a = 1 + \sum_{j \in S_\tau} v_j$. Accordingly, define on $[v_{\min}, 1]^N \times \mathbb{R}_+^{t-1}$

$$\bar{\pi}_t(v, y) := \pi_t(v) \prod_{\tau=1}^{t-1} \left(1 + \sum_{j \in S_\tau} v_j \right) \exp \left(-y_\tau \left(1 + \sum_{j \in S_\tau} v_j \right) \right).$$

For fixed v , every factor after $\pi_t(v)$ is a normalized Exponential density. Tonelli's theorem therefore gives $\int_{\mathbb{R}_+^{t-1}} \bar{\pi}_t(v, y) dy = \pi_t(v)$, establishing the required marginal. Dividing the joint density by this marginal leaves a product of normalized one-dimensional densities, so the exposures are conditionally independent with the rates stated in the proposition.

It remains to derive the conditional distribution of v given y . For fixed y , Bayes' formula, the joint density constructed above, and (8) give

$$\begin{aligned} \bar{\pi}_t(v | y) &\stackrel{(a)}{\propto} \pi_0(v) \prod_{\tau=1}^{t-1} \frac{v_{I_\tau}^{\mathbb{1}\{I_\tau \neq \emptyset\}}}{1 + \sum_{j \in S_\tau} v_j} \left(1 + \sum_{j \in S_\tau} v_j \right) \exp \left(-y_\tau \left(1 + \sum_{j \in S_\tau} v_j \right) \right) \\ &\stackrel{(b)}{\propto} \prod_{i \in \mathcal{N}} v_i^{\alpha_0 + V_{i,t-1}} \exp \left(-v_i \left(\beta_0 + \sum_{\tau=1}^{t-1} \mathbb{1}\{i \in S_\tau\} y_\tau \right) \right) \mathbb{1}\{v_i \in [v_{\min}, 1]\}. \end{aligned}$$

Substituting the joint density and $\pi_t(v) \propto \pi_0(v) \mathcal{L}(v | \mathcal{H}_{t-1})$ gives (a), making the cancellation of each MNL denominator with its Exponential rate explicit. Relation (b) drops $\exp(-\sum_{\tau < t} y_\tau)$, which is constant in v , and groups the remaining terms by product: purchases add $V_{i,t}$ to the prior shape, while latent exposures from periods with $i \in S_\tau$ add to the prior rate.

Each factor is an integrable truncated Gamma kernel involving only v_i . Its shape and rate are those in Proposition 2, and the normalizing integral factorizes across products, proving the stated conditional independence. When $t = 1$, the products are empty, the latent block is absent, and the attraction distribution reduces to the prior.

The two full conditionals define a transition from any current state in $[v_{\min}, 1]^N$. To verify invariance, suppose that the current state has density π_t . Drawing y from its conditional density and then a new state v' from the attraction conditional completes one Gibbs sweep. For fixed $v' \in [v_{\min}, 1]^N$, Tonelli's theorem gives the following output density, with each integral taken over the corresponding support:

$$\begin{aligned} \int \int \pi_t(v) \bar{\pi}_t(y | v) \bar{\pi}_t(v' | y) dy dv &= \int \bar{\pi}_t(y) \bar{\pi}_t(v' | y) dy \\ &= \pi_t(v'). \end{aligned}$$

The first equality integrates out the current state, and the second uses the v' -marginal of the augmented density. Hence the output of the sweep has density π_t . $\blacksquare$

COROLLARY EC.1 (Exact Grouping of the Latent Variables). *Let $t \geq 1$ and condition on the realized history $\mathcal{H}_{t-1}$. Let $\mathcal{A}_t$ be the set of distinct nonempty assortments among $S_1, \dots, S_{t-1}$, and let $n_{A,t} := \sum_{\tau < t} \mathbb{1}\{S_\tau = A\}$ for $A \in \mathcal{A}_t$. For each $A \in \mathcal{A}_t$, aggregate the record-level exposures as $Y_A := \sum_{\tau < t: S_\tau = A} y_\tau$. Conditional on v and $\mathcal{H}_{t-1}$, these variables are mutually independent, with the following law for each $A \in \mathcal{A}_t$:*

$$Y_A \mid v, \mathcal{H}_{t-1} \sim \text{Gamma}\left(n_{A,t}, 1 + \sum_{j \in A} v_j\right).$$

Moreover, conditional on $\{Y_A\}_{A \in \mathcal{A}_t}$ and $\mathcal{H}_{t-1}$, the attraction parameters are mutually independent, and for each $i \in \mathcal{N}$,

$$v_i \mid \{Y_A\}_{A \in \mathcal{A}_t}, \mathcal{H}_{t-1} \sim \text{Gamma}_{[v_{\min}, 1]}\left(\alpha_0 + V_{i,t}, \beta_0 + \sum_{A \in \mathcal{A}_t: i \in A} Y_A\right).$$

Drawing these two blocks induces the same transition kernel on v as the record-level Gibbs update of Proposition 2 and therefore leaves π_t invariant.

Proof of Corollary EC.1. The attraction conditional in Proposition 2 depends on the record-level exposures only through their sums over repeated assortments. It therefore suffices to identify the law of these sums and verify that substituting them preserves the normalized attraction density. If $\mathcal{A}_t$ is empty, both updates draw directly from the prior, so assume $\mathcal{A}_t \neq \emptyset$.

Conditional on v and $\mathcal{H}_{t-1}$, the $n_{A,t}$ exposures associated with an assortment A are independent Exponential variables with common rate $1 + \sum_{j \in A} v_j$. Their sum Y_A thus has the Gamma law stated in the corollary. Distinct assortments use disjoint sets of these independent variables, so their sums are independent as well. Write $\bar{\pi}_t^G(Y \mid v)$ for the resulting product density. Exposures associated with an empty assortment enter no attraction conditional and integrate to one.

For each product i , partitioning its display periods by assortment gives $\sum_{\tau=1}^{t-1} \mathbb{1}\{i \in S_\tau\} y_\tau = \sum_{A \in \mathcal{A}_t: i \in A} Y_A$. Substituting this identity into the attraction conditional of Proposition 2 shows that it depends on y only through Y ; iterated conditioning therefore yields the conditional law stated in the corollary, with normalized product density $\bar{\pi}_t^G(v' \mid Y)$. The shapes, rates, and truncation intervals agree coordinate by coordinate, so their normalizing constants agree as well: $\bar{\pi}_t(v' \mid y) = \bar{\pi}_t^G(v' \mid Y)$ for every record-level vector with grouped sums Y .

It remains to integrate over the first block. For fixed $v, v' \in [v_{\min}, 1]^N$, substituting the equality of attraction densities and then using the law of the grouped sums gives

$$\begin{aligned} \int_{\mathbb{R}_+^{t-1}} \bar{\pi}_t(y | v) \bar{\pi}_t(v' | y) dy &= \int_{\mathbb{R}_+^{t-1}} \bar{\pi}_t(y | v) \bar{\pi}_t^G(v' | Y) dy \\ &= \int_{\mathbb{R}_+^{\mathcal{A}_t}} \bar{\pi}_t^G(Y | v) \bar{\pi}_t^G(v' | Y) dY. \end{aligned}$$

In the first two integrals, $Y_A = \sum_{\tau < t: S_\tau = A} y_\tau$; the last integral averages directly over those sums. The two sides are precisely the transition densities of the record-level and grouped updates. They therefore induce the same kernel for every current state v . Since the record-level update leaves π_t invariant by Proposition 2, the grouped update does as well. ■

Computational Cost. Computing the rates of the grouped latent variables and accumulating their contributions to the attraction rates visits each incidence pair (A, i) , with $i \in A$, a constant number of times. Together with the N attraction draws, one sweep therefore requires $O(N + \sum_{A \in \mathcal{A}_t} |A|)$ expected time under unit expected cost per univariate Gamma draw. The purchase counts, attraction values, assortment counts, and item-incidence lists occupy the same order of memory. Since $|A| \leq K$ for every feasible assortment and $|\mathcal{A}_t| \leq t - 1$,

$$M \sum_{\tau=1}^T \left(N + \sum_{A \in \mathcal{A}_\tau} |A| \right) \leq MNT + MK \sum_{\tau=1}^T |\mathcal{A}_\tau| \leq MNT + \frac{MKT(T-1)}{2}.$$

Thus the horizon cost is $O\{M(NT + K \sum_{\tau=1}^T |\mathcal{A}_\tau|)\}$, which is $O\{M(NT + KT^2)\}$ in the worst case.

EC.3. Regret Analysis

Theorems 1 and 2 are proved in Sections 4.3 and 5.3. This section collects their setup, auxiliary results, and supporting calculations. Every proposition and lemma stated in this section is proved in Section EC.4 for the asynchronous-clock results and Section EC.5 for the full-likelihood result.

EC.3.1. Asynchronous-Clock Policy

The revenue structure localizes an appreciable loss in either an inaccurate selected coordinate or an omitted product whose inclusion can be forced by correcting its sampled attraction. This one-coordinate repair converts independent sampling into a comparison of probabilities. The proof then charges the waiting cost to retained comparisons and controls it through deterministic Bernoulli prefixes, without assigning a binomial law to an adaptively observed count.

Setup and Notation. Let v be the true vector under Assumption 1. Probabilities $\mathbb{P}_v$ and expectations $\mathbb{E}_v$ include both customer and policy randomization; their versions $\mathbb{P}_t$ and $\mathbb{E}_t$ condition

on $\mathcal{F}_{t-1}$. Constants $C > 0$ may change from line to line and depend only on $(v_{\min}, \alpha_0, \beta_0)$ unless a smaller dependence is specified. Retain the inclusion-wise minimal selection rule of Section 3.

The number of comparisons retained before period t is $n_{i,t} := V_{i,t} + N_{i,t}$, and the conditional purchase probability at a comparison is $\theta_i := v_i/(1 + v_i) \in (0, 1/2]$. Algorithm 1 draws attractions according to

$$\tilde{v}_{i,t} \mid \mathcal{F}_{t-1} \sim \text{BetaPrime}(\alpha_0 + V_{i,t}, \beta_0 + N_{i,t}), \quad \frac{\tilde{v}_{i,t}}{1 + \tilde{v}_{i,t}} \mid \mathcal{F}_{t-1} \sim \text{Beta}(\alpha_0 + V_{i,t}, \beta_0 + N_{i,t}). \quad (\text{EC.1})$$

The second law follows by the change of variable $u = x/(1 + x)$ in the BetaPrime density: its inverse is $x = u/(1 - u)$ and its Jacobian is $(1 - u)^{-2}$. To count retained comparisons, condition also on the current draw, which determines S_t before the customer chooses. The MNL law gives

$$\mathbb{P}_v(i \in S_t, I_t \in \{i, 0\} \mid \mathcal{F}_{t-1}, \tilde{v}_t) = \mathbb{1}\{i \in S_t\} \frac{1 + v_i}{1 + \sum_{j \in S_t} v_j} \geq \frac{\mathbb{1}\{i \in S_t\}}{K + 1}. \quad (\text{EC.2})$$

The restrictions $v_j \leq 1$ and $|S_t| \leq K$ control the denominator. Conditioning on the draw makes selection and sampling-error events measurable before observing the customer, whose new randomization is independent of the policy's draw.

The retained comparisons remove the other products from relative choice probabilities, but their timing still depends on the policy and customer choices. The next proposition represents these comparisons by a fixed Bernoulli sequence while preserving the law of the complete adaptive history.

PROPOSITION EC.1 (Bernoulli Prefixes under Adaptive Selection). *For any product i and any admissible policy, there is a joint construction of its adaptive MNL history and an infinite i.i.d. sequence $(Z_{i,k})_{k \geq 1}$ of $\text{Bernoulli}(\theta_i)$ variables such that, simultaneously for all $t \geq 1$,*

$$V_{i,t} = \sum_{k=1}^{n_{i,t}} Z_{i,k}, \quad N_{i,t} = n_{i,t} - V_{i,t},$$

almost surely. In particular, $\sum_{k=1}^n Z_{i,k} \sim \text{Binomial}(n, \theta_i)$ for every deterministic $n \geq 0$, with empty sums equal to zero.

The construction treats one product at a time and does not assert independence between the sequences of different products. A binomial law at a deterministic prefix does not imply a binomial law conditional on the adaptive count $n_{i,t}$. The regret analysis first bounds the contribution at that adaptive count by a sum over deterministic prefixes, then uses their binomial distributions.

The repair argument uses three properties of MNL revenue: a fixed revenue threshold yields a linear objective, attraction errors control the revenue of a fixed assortment, and an inclusion-wise minimal optimum contains only products with positive objective coefficients.

LEMMA EC.1 (Revenue Thresholds and Linear Reformulation). *The MNL revenue and feasible family $\mathcal{S}$ satisfy the following properties.*

(i) Revenue threshold. For $S \in \mathcal{S}$, $x \in \mathbb{R}_{++}^N$, and $\lambda \in \mathbb{R}$,

$$R(S, x) - \lambda = \frac{\sum_{i \in S} (r_i - \lambda) x_i - \lambda}{1 + \sum_{j \in S} x_j}.$$

In particular, $R(S, x) \geq \lambda$ if and only if $\sum_{i \in S} (r_i - \lambda) x_i \geq \lambda$.

(ii) Revenue sensitivity. For $S \in \mathcal{S}$ and $x, y \in \mathbb{R}_{++}^N$,

$$|R(S, x) - R(S, y)| \leq \frac{\sum_{i \in S} |x_i - y_i|}{1 + \sum_{j \in S} y_j}.$$

(iii) Linear reformulation. For $x \in \mathbb{R}_{++}^N$, the revenue-maximizing assortments are precisely the maximizers of $\sum_{i \in S} (r_i - R^*(x)) x_i$, whose maximum is $R^*(x)$. Under the inclusion-wise minimal tie-breaking rule, $r_i > R^*(x)$ for every $i \in S^*(x)$.

The following bound holds for fixed positive prior parameters and, in particular, under Assumption 1.

LEMMA EC.2 (Summed Posterior Odds). *Let $\theta \in (0, 1/2]$ and $0 < \epsilon < \theta$. For each integer $n \geq 0$ and $s \in \{0, \dots, n\}$, define $q_n(s) := \mathbb{P}(\text{Beta}(\alpha_0 + s, \beta_0 + n - s) \geq \theta - \epsilon)$. There exists $C < \infty$, depending only on (α_0, β_0) , such that, for every $T \geq 1$,*

$$\sum_{n=0}^{T-1} \sum_{s=0}^n \binom{n}{s} \theta^s (1 - \theta)^{n-s} \frac{1 - q_n(s)}{q_n(s)} \leq \frac{C}{\epsilon^2} \log(eT). \quad (\text{EC.3})$$

For fixed n and s , $q_n(s)$ is the probability that a Beta draw reaches the threshold $\theta - \epsilon$. Thus, the ratio $(1 - q_n(s))/q_n(s)$ is the expected number of failed independent draws before the first threshold crossing, with the observed counts held fixed. The inner sum averages this waiting cost over the binomial distribution of the success count, and the outer sum accumulates these averages over prefix lengths. The lemma bounds this cumulative cost by $C\epsilon^{-2} \log(eT)$.

EC.3.2. Full-Likelihood Policy

Setup and Notation. Condition on the history $\mathcal{F}_{t-1}$. Throughout this subsection, a subscript t on probability or expectation denotes this conditioning under the joint Bayesian law. The true vector has posterior law $v \mid \mathcal{F}_{t-1} \sim \pi_t$. Full-likelihood Thompson sampling draws $\tilde{v}_t \sim \pi_t$ independently of v , conditional on $\mathcal{F}_{t-1}$, and offers $S_t = S^*(\tilde{v}_t)$. Set $U := \log v$, $\tilde{U}_t := \log \tilde{v}_t$, $\mathcal{U} := [\log v_{\min}, 0]^N$, and $\mathbf{1} := (1, \dots, 1)$.

For $x \in [v_{\min}, 1]^N$, set $x_0 := 1$. For $a \in \mathcal{N}$, let e_a be canonical and set $e_0 := 0$. Given S , let $I \in S \cup \{0\}$ have law $\mathbb{P}(I = a) = p_a(S, x)$. The Fisher information in log-coordinates is $F_S(x) := \text{Cov}(e_I)$. For $h_0 = 0$, the pairwise variance identity gives

$$h^\top F_S(x) h = \text{Var}(h_I) = \sum_{\substack{0 \leq a < b \leq N \\ a, b \in S \cup \{0\}}} p_a(S, x) p_b(S, x) (h_a - h_b)^2. \quad (\text{EC.4})$$

Thus $F_S(x) \succeq 0$ and $\text{tr } F_S(x) = \sum_{a \in S} p_a(S, x) - \sum_{a \in S} p_a(S, x)^2 \leq 1$.

The all-ones point $\mathbf{1}$ gives a Fisher matrix comparable to every other one in the box. Indeed, for $x \in [v_{\min}, 1]^N$ and $a \in S \cup \{0\}$, $p_a(S, x)/p_a(S, \mathbf{1}) = x_a(1 + |S|)/(1 + \sum_{j \in S} x_j)$, with $x_a = 1$ when $a = 0$. The first factor lies in $[v_{\min}, 1]$ and the second in $[1, v_{\min}^{-1}]$, because $1 + |S|v_{\min} \leq 1 + \sum_{j \in S} x_j \leq 1 + |S|$; hence $p_a(S, x)/p_a(S, \mathbf{1}) \in [v_{\min}, v_{\min}^{-1}]$.

The variance satisfies $\text{Var}_{p(S, x)}(h_I) = \inf_{c \in \mathbb{R}} \sum_{a \in S \cup \{0\}} p_a(S, x) (h_a - c)^2$, since $\mathbb{E}[(h_I - c)^2]$ is minimized at $c = \mathbb{E}[h_I]$. For every $c \in \mathbb{R}$, the probability-ratio bound gives

$$v_{\min} \sum_{a \in S \cup \{0\}} p_a(S, \mathbf{1}) (h_a - c)^2 \leq \sum_{a \in S \cup \{0\}} p_a(S, x) (h_a - c)^2 \leq v_{\min}^{-1} \sum_{a \in S \cup \{0\}} p_a(S, \mathbf{1}) (h_a - c)^2.$$

These inequalities hold pointwise in c , so taking infima in all three terms yields the Loewner comparison

$$v_{\min} F_S(\mathbf{1}) \preceq F_S(x) \preceq v_{\min}^{-1} F_S(\mathbf{1}). \quad (\text{EC.5})$$

Let $\kappa := \beta_0$ and let I_N denote the $N \times N$ identity matrix. Define the history-measurable information

$$H_t := \kappa I_N + \sum_{\tau=1}^{t-1} F_{S_\tau}(\mathbf{1}) \succeq \kappa I_N \succ 0.$$

Because S_t is selected by the posterior draw, its Fisher width cannot be controlled by treating the assortment as independent of that draw. The following lemma instead gives one pathwise bound simultaneously for every feasible assortment, together with the moment bound needed when regret is averaged.

LEMMA EC.3 (Pairwise Posterior Concentration). *There exists a random variable $Z_t \geq 0$ such that*

$$\mathbb{E}_t[Z_t^2] \leq C v_{\min}^{-1} \log(eN),$$

where C is a numerical constant. Moreover, pathwise and simultaneously for every feasible assortment S ,

$$(\tilde{U}_t - U)^\top F_S(\mathbf{1}) (\tilde{U}_t - U) \leq Z_t^2 \text{tr}(H_t^{-1} F_S(\mathbf{1})). \quad (\text{EC.6})$$

Supporting Detail for Step 1 of the Proof of Theorem 2. The sensitivity bound follows from the derivative of the revenue along a log-linear path. Let $S \in \mathcal{S}$ and $x, x' \in [v_{\min}, 1]^N$, set $h := \log x' - \log x$ with $h_0 = 0$, and let $u_s := (1 - s) \log x + s \log x'$ for $s \in [0, 1]$, with $u_{s,0} := 0$. Convexity of $\mathcal{U}$ keeps u_s inside the log-support, so $e^{u_s} \in [v_{\min}, 1]^N$ throughout.

We first identify the directional derivative of the revenue. Let $\mathbb{P}(I_s = a) = p_a(S, e^{u_s})$ for $a \in S \cup \{0\}$ and set $r_0 = 0$, so that $R(S, e^{u_s}) = \mathbb{E}[r_{I_s}]$. Write $p_a := p_a(S, e^{u_s})$ and let all sums over a and b range over $S \cup \{0\}$. The softmax derivative $\partial p_a / \partial u_b = p_a(\mathbb{1}\{a = b\} - p_b)$ and $\frac{d}{ds} u_{s,a} = h_a$ give, by the chain rule,

$$\frac{d}{ds} R(S, e^{u_s}) = \sum_a r_a \sum_b p_a (\mathbb{1}\{a = b\} - p_b) h_b = \sum_a r_a p_a h_a - \left(\sum_a r_a p_a \right) \left(\sum_b p_b h_b \right) = \text{Cov}(r_{I_s}, h_{I_s}).$$

We now compare this covariance with the Fisher width. Because $r_{I_s} \in [0, 1]$, $\text{Var}(r_{I_s}) \leq 1/4$. Hence, uniformly in s ,

$$\left| \frac{d}{ds} R(S, e^{u_s}) \right| \stackrel{(a)}{\leq} \sqrt{\text{Var}(r_{I_s}) \text{Var}(h_{I_s})} \stackrel{(b)}{\leq} \frac{1}{2} \sqrt{h^\top F_S(e^{u_s}) h} \stackrel{(c)}{\leq} \frac{1}{2\sqrt{v_{\min}}} \sqrt{h^\top F_S(\mathbf{1}) h},$$

where (a) is Cauchy–Schwarz for the covariance, (b) substitutes $\text{Var}(r_{I_s}) \leq 1/4$ and identifies $\text{Var}(h_{I_s})$ with $h^\top F_S(e^{u_s}) h$ through (EC.4), and (c) uses the upper bound in (EC.5), legitimate because $u_s \in \mathcal{U}$ and hence $e^{u_s} \in [v_{\min}, 1]^N$.

EC.4. Auxiliary Results for Asynchronous Clocks

This section proves the auxiliary results stated in Section EC.3.1. The Bernoulli-prefix representation handles the adaptive timing of the observations; the revenue identities establish the one-coordinate repair property; and the two scalar sampling bounds control failed repairs and errors on selected products. We retain the notation and constant conventions of that section.

Proof of Proposition EC.1. Apply the construction in the proof of Proposition 1. It preserves the complete adaptive MNL history while revealing exactly the first $n_{i,t}$ entries of an infinite i.i.d. Bernoulli(θ_i) sequence. The count identity in that construction gives $V_{i,t} = \sum_{k=1}^{n_{i,t}} Z_{i,k}$ simultaneously over periods, and $N_{i,t} = n_{i,t} - V_{i,t}$ gives the second identity. At each deterministic n , the sum of the first n entries has the asserted binomial law. $\blacksquare$

Proof of Lemma EC.1. Part (i). Subtracting λ from the revenue definition and collecting terms over the common denominator gives

$$R(S, x) - \lambda = \frac{\sum_{i \in S} r_i x_i - \lambda(1 + \sum_{j \in S} x_j)}{1 + \sum_{j \in S} x_j} = \frac{\sum_{i \in S} (r_i - \lambda) x_i - \lambda}{1 + \sum_{j \in S} x_j}.$$

The denominator is positive, so the revenue excess and the final numerator have the same sign. This proves the equivalence.

Part (ii). To isolate the coordinate errors $x_i - y_i$, first substitute the definition of $R(S, y)$ and put the difference over its positive denominator. The revenue identity $\sum_{i \in S} r_i x_i = R(S, x)(1 + \sum_{i \in S} x_i)$ then converts the numerator into a sum of coordinate errors:

$$\begin{aligned}
|R(S, x) - R(S, y)| &= \left| R(S, x) - \frac{\sum_{i \in S} r_i y_i}{1 + \sum_{j \in S} y_j} \right| \\
&= \frac{\left| R(S, x)(1 + \sum_{j \in S} y_j) - \sum_{i \in S} r_i y_i \right|}{1 + \sum_{j \in S} y_j} \\
&= \frac{\left| \sum_{i \in S} (r_i - R(S, x))(x_i - y_i) \right|}{1 + \sum_{j \in S} y_j} \\
&\stackrel{(a)}{\leq} \frac{\sum_{i \in S} |r_i - R(S, x)| |x_i - y_i|}{1 + \sum_{j \in S} y_j} \\
&\stackrel{(b)}{\leq} \frac{\sum_{i \in S} |x_i - y_i|}{1 + \sum_{j \in S} y_j}.
\end{aligned}$$

The triangle inequality gives (a), while the bounds $r_i, R(S, x) \in [0, 1]$ imply $|r_i - R(S, x)| \leq 1$ and hence (b).

Part (iii). At $\lambda = R^*(x)$, part (i) shows that every feasible assortment has linear objective value at most $R^*(x)$, with equality exactly when its revenue is optimal. Since $S^*(x)$ attains that revenue, the maximum of the linear objective equals $R^*(x)$ and the maximizers coincide.

Suppose a selected product $i \in S^*(x)$ satisfies $(r_i - R^*(x))x_i \leq 0$. Deleting it preserves feasibility. A negative objective coefficient would make deletion strictly improve the linear objective, contradicting optimality. A zero coefficient would preserve the objective and therefore revenue optimality, contradicting inclusion-wise minimality. Every selected product therefore has a positive coefficient, and $x_i > 0$ gives $r_i > R^*(x)$. ■

Proof of Lemma 1. *Part (i).* Let $\lambda = \max_{S \in \mathcal{S}: i \notin S} R(S, x)$. Since $x_{-i} \in \mathcal{D}_i(\xi)$, (3) implies that some feasible assortment S containing i has revenue above λ under $x^{(i)}$. Deleting i preserves feasibility, and the remaining products have the same attractions under x and $x^{(i)}$. Hence, by the definition of λ , $R(S \setminus \{i\}, x^{(i)}) = R(S \setminus \{i\}, x) \leq \lambda$. The threshold identity in Lemma EC.1(i), applied to $S \setminus \{i\}$ and S , therefore gives

$$\sum_{j \in S \setminus \{i\}} (r_j - \lambda) x_j^{(i)} \leq \lambda < \sum_{j \in S} (r_j - \lambda) x_j^{(i)}.$$

The two sums differ only in the term of product i , so $(r_i - \lambda)(v_i - \xi) > 0$. Since $v_i - \xi \geq v_{\min}/2 > 0$, we obtain $r_i > \lambda$.

Since $x_i \geq v_i - \xi$ and $r_i > \lambda$, increasing only coordinate i from $v_i - \xi$ to x_i cannot decrease the linear objective value of S . The threshold identity therefore gives $R(S, x) > \lambda$. Every assortment excluding i has revenue at most λ , so every optimum under x must contain i .

Part (ii). Let $\lambda = R^*(x)$. We increase the omitted coordinates to at least those of $v - \xi \mathbf{1}$ without changing the optimal value. By hypothesis, the selected coordinates lie within ξ of their true values, so the resulting vector will dominate $v - \xi \mathbf{1}$. To justify the simultaneous correction, use the linear objective at the fixed revenue level λ . Lemma EC.1(i),(iii) gives $r_j > \lambda$ for $j \in S^*(x)$ and

$$\max_{S \in \mathcal{S}} \sum_{j \in S} (r_j - \lambda) x_j = \sum_{j \in S^*(x)} (r_j - \lambda) x_j = \lambda.$$

This linear problem selects up to K products with the largest positive objective coefficients. It therefore suffices to show that correcting an omitted coordinate cannot raise its coefficient above the insertion or replacement threshold determined by $S^*(x)$.

Let $i \notin S^*(x)$. Since $S^*(x)$ is optimal and excludes i , $\max_{S \in \mathcal{S}: i \notin S} R(S, x) = \lambda$. Failure of the repair condition in (3) implies $R^*(x^{(i)}) \leq \lambda$, while $S^*(x)$ still attains λ under $x^{(i)}$. If $|S^*(x)| < K$, adding i cannot increase the linear objective, so $(r_i - \lambda)(v_i - \xi) \leq 0$. If $|S^*(x)| = K$, replacing any $j \in S^*(x)$ cannot increase it either: the coefficient of the added product is at most that of the removed product. Thus $(r_i - \lambda)(v_i - \xi) \leq \min_{j \in S^*(x)} (r_j - \lambda) x_j$.

Now increase each omitted coordinate to $z_i := \max\{x_i, v_i - \xi\}$, leaving $z_j = x_j$ for $j \in S^*(x)$. If $r_i - \lambda \leq 0$, increasing the attraction can only decrease its objective coefficient. Otherwise, the coefficient under z is the larger of the original and corrected coefficients, both of which satisfy the insertion or replacement bound. When capacity is full, each added product can be paired with a removed product whose coefficient is at least as large; any additional removal deletes a positive term. When capacity is not full, every omitted product has a nonpositive coefficient. Thus $S^*(x)$ continues to maximize the linear objective after all omitted coordinates have been increased. Its objective value remains λ , so the threshold identity gives $R^*(z) = \lambda$; ties may change the selected optimum but not its value.

It remains to compare this value with the true optimum. The optimized revenue $R^*(\cdot)$ is non-decreasing in each coordinate. Indeed, at an inclusion-wise minimal optimum under any positive vector, $r_j > R^*(\cdot)$ for every selected product j by Lemma EC.1(iii). Increasing attractions preserves its threshold inequality and hence cannot decrease the optimal value. This monotonicity concerns optimized revenue, rather than the revenue of an arbitrary fixed assortment. By construction and the error bound on selected coordinates, $z \geq v - \xi \mathbf{1}$.

To bound the revenue change when the true attractions decrease by ξ , keep the assortment $S^*(v)$ fixed. Apply Lemma EC.1(ii) with $v - \xi \mathbf{1}$ as the first vector and v as the second, so the denominator is evaluated at the true vector. Since $R(S^*(v), v) = R^*$, this gives

$$|R(S^*(v), v - \xi \mathbf{1}) - R^*| \leq \frac{\sum_{j \in S^*(v)} |(v_j - \xi) - v_j|}{1 + \sum_{j \in S^*(v)} v_j} = \frac{|S^*(v)| \xi}{1 + \sum_{j \in S^*(v)} v_j} \leq \frac{|S^*(v)| \xi}{1 + |S^*(v)| v_{\min}} \leq \frac{\xi}{v_{\min}}.$$

Each coordinate difference has absolute value ξ , and $v_j \geq v_{\min}$ bounds the denominator. We can therefore compare the true optimum with the unchanged value under z and then return to the true revenue of $S^*(x)$:

$$R^* - \frac{\xi}{v_{\min}} \stackrel{(a)}{\leq} R(S^*(v), v - \xi \mathbf{1}) \leq R^*(v - \xi \mathbf{1}) \stackrel{(b)}{\leq} R^*(z) = \lambda \stackrel{(c)}{\leq} R(S^*(x), v) + \frac{\xi}{v_{\min}}.$$

The bound for $S^*(v)$ gives (a), optimality gives the next comparison, and coordinate monotonicity gives (b), since $z \geq v - \xi \mathbf{1} > 0$. For (c), apply the same sensitivity bound to $S^*(x)$ under x and v . By hypothesis, $|x_i - v_i| \leq \xi$ for every $i \in S^*(x)$, so the same denominator estimate bounds the revenue error by $\xi/v_{\min}$; using $\lambda = R(S^*(x), x)$ gives (c). Rearranging proves part (ii). ■

Proof of Lemma EC.2. We first establish the finite-horizon bound for the uniform Beta prior and then compare a Beta prior with arbitrary positive shape parameters with that reference. Throughout this proof, $C > 0$ denotes a universal constant that may change between lines. It suffices to consider $\epsilon \leq \theta/4$: otherwise, replacing ϵ by $\theta/4 \geq \epsilon/4$ increases the posterior odds and introduces at most a factor of 16 in the required bound. Write $y := \theta - \epsilon$ and let $f_{n,z}(s) := \binom{n}{s} z^s (1-z)^{n-s}$ denote the binomial probability mass.

Step 1 (Uniform-Prior Reference). For $0 \leq s \leq m$, let $q_m^{\text{unif}}(s) := \mathbb{P}\{\text{Beta}(s+1, m-s+1) \geq y\}$ denote the threshold probability under the uniform prior. Lemma 2.9 of Agrawal and Goyal (2017), with $\mu_1 = \theta$, $y_i = y$, and $k = m$, bounds $\sum_{s=0}^m f_{m,\theta}(s)(1 - q_m^{\text{unif}}(s))/q_m^{\text{unif}}(s)$ by $3/\epsilon$ when $m < 8/\epsilon$ and otherwise by

$$C \left(e^{-m\epsilon^2/2} + \frac{e^{-mD(y\|\theta)}}{(m+1)\epsilon^2} + \frac{1}{e^{m\epsilon^2/4} - 1} \right), \quad (\text{EC.7})$$

where $D(p\|q) := p \log(p/q) + (1-p) \log((1-p)/(1-q))$ for $p, q \in (0, 1)$ is the Bernoulli Kullback–Leibler divergence. The constants are absolute, and their Appendix B expresses the expectation as this binomial sum.

Let $M \geq 0$ be an integer. We now sum the bounds from the cited lemma over prefix lengths $m = 0, \dots, M$. For each $0 \leq m < 8/\epsilon$, the binomial average over s is at most $3/\epsilon$ by the first case of the cited lemma. There are at most $8/\epsilon + 1$ such integers, so their contribution to the cumulative sum is at most $(8/\epsilon + 1)(3/\epsilon) \leq 27/\epsilon^2$, since $\epsilon < 1$.

For $m \geq 8/\epsilon$, we bound the contribution of these binomial averages by summing (EC.7) over m . Consider the three terms inside the parentheses, retaining the common factor C . Extending the sum of the first term to all $m \geq 1$ gives $\sum_{m \geq 1} e^{-m\epsilon^2/2} = (e^{\epsilon^2/2} - 1)^{-1} \leq 2/\epsilon^2$. The equality sums a geometric series, and the inequality uses $e^x - 1 \geq x$. Since $D(y\|\theta) \geq 0$, the second term

satisfies $e^{-mD(y\|\theta)}/((m+1)\epsilon^2) \leq 1/(m\epsilon^2)$. For the third term, $e^x - 1 \geq x$ with $x = m\epsilon^2/4$ gives $(e^{m\epsilon^2/4} - 1)^{-1} \leq 4/(m\epsilon^2)$. Combining the two regimes gives

$$\sum_{m=0}^M \sum_{s=0}^m f_{m,\theta}(s) \frac{1 - q_m^{\text{unif}}(s)}{q_m^{\text{unif}}(s)} \leq \frac{27}{\epsilon^2} + \frac{C}{\epsilon^2} \left(2 + 5 \sum_{m=1}^M \frac{1}{m} \right) \leq \frac{C}{\epsilon^2} \log(e(M+1)). \quad (\text{EC.8})$$

The last inequality uses $\sum_{m=1}^M m^{-1} \leq 1 + \log M$ for $M \geq 1$ and absorbs the numerical constants into C . For $M = 0$, the sum is empty and $\log(e(M+1)) = 1$.

Step 2 (Prior Comparison). To compare with the uniform prior, we decrease the prior's first shape to at most one and increase its second shape to an integer. This can only increase the odds of falling below y . Set $a := \min\{\alpha_0, 1\}$ and $b := \max\{1, \lceil \beta_0 \rceil\}$, and define $\bar{q}_n(s) := \mathbb{P}\{\text{Beta}(a+s, b+n-s) \geq y\}$. The restriction $a \leq 1$ allows a direct tail bound for zero successes, while the integer b permits comparison with a uniform-prior posterior after shifting the success and failure counts.

For the Gamma comparison, let $G_{a+s} \sim \text{Gamma}(a+s, 1)$ and $G_{\beta_0+n-s} \sim \text{Gamma}(\beta_0+n-s, 1)$ be independent, with unit rate throughout. Subscripts record shape parameters; variables specified as independent are distinct draws even when their shapes coincide. Independently of these variables and each other, take Gamma increments $G_{\alpha_0-a}, G_{b-\beta_0}$ with shapes $\alpha_0 - a$ and $b - \beta_0$, respectively; an increment is identically zero when its shape difference is zero. Both increments are nonnegative, so under this joint construction,

$$\frac{G_{a+s}}{G_{a+s} + G_{\beta_0+n-s} + G_{b-\beta_0}} \leq \frac{G_{a+s}}{G_{a+s} + G_{\beta_0+n-s}} \leq \frac{G_{a+s} + G_{\alpha_0-a}}{G_{a+s} + G_{\alpha_0-a} + G_{\beta_0+n-s}}.$$

Gamma additivity gives the leftmost ratio the $\text{Beta}(a+s, b+n-s)$ law defining $\bar{q}_n(s)$, and the rightmost ratio the $\text{Beta}(\alpha_0+s, \beta_0+n-s)$ law defining $q_n(s)$. Whenever the leftmost ratio reaches y , the rightmost ratio also reaches y . Taking probabilities therefore gives $\bar{q}_n(s) \leq q_n(s)$, which is the required stochastic comparison. Since $(1-q)/q = q^{-1} - 1$ decreases with $q > 0$, we also have $(1-q_n(s))/q_n(s) \leq (1-\bar{q}_n(s))/\bar{q}_n(s)$.

For positive counts, one more Gamma comparison connects both cases below to Step 1. Decreasing the first shape from $a+s$ to s lowers the Beta upper tail. The resulting law, $\text{Beta}(s, b+n-s)$, is the uniform-prior posterior after $m = n+b-2$ observations and $s-1$ successes: its parameters are $1+(s-1) = s$ and $1+m-(s-1) = b+n-s$. Since both shapes are positive integers, Fact 3 in Appendix A of Agrawal and Goyal (2017) also expresses this tail as a binomial probability. Thus, for $1 \leq s \leq n$,

$$\bar{q}_n(s) \geq q_{n+b-2}^{\text{unif}}(s-1) = \mathbb{P}\{\text{Binomial}(n+b-1, y) \leq s-1\}. \quad (\text{EC.9})$$

The comparison (EC.9) bounds the odds for each fixed (n, s) , but the desired sum weights them by $f_{n,\theta}(s)$. After the index shift, Step 1 instead weights the reference odds by $f_{n+b-2,\theta}(s-1)$.

Applying its summed bound therefore also requires controlling the ratio of these two binomial weights. This ratio introduces a factor proportional to $n\theta/s$, which need not be bounded for all positive counts. We separate the cases accordingly: $s = 0$ requires a direct bound because the shift removes one success; $s \geq 1$ with $s \geq n\theta/2$ gives $n\theta/s \leq 2$; and $1 \leq s < n\theta/2$ requires using the small probability of these counts to control their weighted odds.

- *Case $s = 0$.* The probability of this prefix is $f_{n,\theta}(0) = (1 - \theta)^n$. We show that its decay offsets the possible growth of the odds by first bounding the posterior probability $\bar{q}_n(0)$ from below.

Let B denote the beta function. Since $b + n \geq 1$, bounding $(1 - x)^{b+n-1}$ by one in its defining integral gives $B(a, b + n) \leq \int_0^1 x^{a-1} dx = 1/a$. Also, $a \leq 1$ gives $x^{a-1} \geq 1$ on $(0, 1)$. Integrating this lower bound in the numerator yields

$$\bar{q}_n(0) = \frac{\int_y^1 x^{a-1} (1-x)^{b+n-1} dx}{B(a, b+n)} \geq \frac{a(1-y)^{b+n}}{b+n}. \quad (\text{EC.10})$$

We first use $1 - \bar{q}_n(0) \leq 1$, then substitute $f_{n,\theta}(0) = (1 - \theta)^n$ and the reciprocal bound from (EC.10). Finally, $y \leq 1/2$ and $(1 - \theta)/(1 - y) = 1 - \epsilon/(1 - y) \leq e^{-\epsilon}$ give

$$f_{n,\theta}(0) \frac{1 - \bar{q}_n(0)}{\bar{q}_n(0)} \leq \frac{f_{n,\theta}(0)}{\bar{q}_n(0)} \leq \frac{(b+n)(1-\theta)^n}{a(1-y)^{b+n}} = \frac{b+n}{a(1-y)^b} \left(\frac{1-\theta}{1-y} \right)^n \leq \frac{2^b(b+n)}{a} e^{-n\epsilon}.$$

Summing the last bound over $0 \leq n \leq T - 1$ and extending to all $n \geq 0$ gives a contribution of at most $(2^b/a) \sum_{n \geq 0} (b+n) e^{-n\epsilon}$. Separating the constant and linear terms in $b+n$ gives $\sum_{n \geq 0} (b+n) e^{-n\epsilon} = b \sum_{n \geq 0} e^{-n\epsilon} + \sum_{n \geq 1} n e^{-n\epsilon}$. The geometric bounds $\sum_{n \geq 0} e^{-n\epsilon} \leq 1 + \epsilon^{-1}$ and $\sum_{n \geq 1} n e^{-n\epsilon} \leq \epsilon^{-2}$ bound these two parts by $b(1 + \epsilon^{-1})$ and ϵ^{-2} , respectively. Restoring the factor $2^b/a$, the contribution from zero successes is at most $(2^b/a)(b(1 + \epsilon^{-1}) + \epsilon^{-2}) \leq C 2^b(b+1)/(a\epsilon^2)$, since $\epsilon < 1$. This includes $n = 0$.

- *Case $s \geq 1$ and $s \geq n\theta/2$.* Here we transfer the sum to Step 1 at a bounded cost. By (EC.9), $(1 - \bar{q}_n(s))/\bar{q}_n(s) \leq (1 - q_{n+b-2}^{\text{unif}}(s-1))/q_{n+b-2}^{\text{unif}}(s-1)$. We must also transfer its weight from $f_{n,\theta}(s)$ to $f_{n+b-2,\theta}(s-1)$.

Expanding the binomial masses, we obtain

$$\frac{f_{n,\theta}(s)}{f_{n+b-2,\theta}(s-1)} = \frac{\binom{n}{s}}{\binom{n+b-2}{s-1}} \theta(1-\theta)^{-(b-1)} \leq \frac{\binom{n}{s}}{\binom{n-1}{s-1}} \theta(1-\theta)^{-(b-1)} = \frac{n\theta}{s} (1-\theta)^{-(b-1)} \leq 2^b. \quad (\text{EC.11})$$

The first inequality replaces the denominator by the smaller coefficient $\binom{n-1}{s-1}$, since $b \geq 1$. The following equality uses $\binom{n}{s}/\binom{n-1}{s-1} = n/s$. The last inequality uses $n\theta/s \leq 2$ and $(1-\theta)^{-(b-1)} \leq 2^{b-1}$, which follow from the case condition and $\theta \leq 1/2$.

Combining the odds comparison with (EC.11) bounds each summand by 2^b times a summand from Step 1. Under the shift $(n, s) \mapsto (n + b - 2, s - 1)$, the new length is nonnegative, the success count lies between zero and that length, and distinct pairs remain distinct. We may therefore extend the shifted sum to all lengths through $T + b - 2$ and apply (EC.8):

$$\sum_{n=1}^{T-1} \sum_{\substack{1 \leq s \leq n \\ s \geq n\theta/2}} f_{n,\theta}(s) \frac{1 - \bar{q}_n(s)}{\bar{q}_n(s)} \leq 2^b \sum_{m=0}^{T+b-2} \sum_{s=0}^m f_{m,\theta}(s) \frac{1 - q_m^{\text{unif}}(s)}{q_m^{\text{unif}}(s)} \leq \frac{C2^b}{\epsilon^2} \log(e(T + b)).$$

- *Case $1 \leq s < n\theta/2$.* Here $n\theta/s$ may be large. We control the weighted odds by comparing the probability of this small success count under θ with its probability under $y = \theta - \epsilon$. This comparison will supply exponential decay in n .

The binomial probability in (EC.9) is the sum $\sum_{k=0}^{s-1} f_{n+b-1,y}(k)$ of nonnegative masses. Keeping only its term at $k = s - 1$ therefore gives $\bar{q}_n(s) \geq f_{n+b-1,y}(s - 1) > 0$. The weighted odds then satisfy

$$f_{n,\theta}(s) \frac{1 - \bar{q}_n(s)}{\bar{q}_n(s)} \leq \frac{f_{n,\theta}(s)}{\bar{q}_n(s)} \leq \frac{f_{n,\theta}(s)}{f_{n+b-1,y}(s - 1)} = \frac{f_{n,\theta}(s)}{f_{n,y}(s)} \frac{f_{n,y}(s)}{f_{n+b-1,y}(s - 1)}.$$

The first inequality uses $1 - \bar{q}_n(s) \leq 1$, and the second uses the lower bound on $\bar{q}_n(s)$ just obtained. The equality follows by multiplying and dividing by $f_{n,y}(s) > 0$. We bound the resulting two factors separately: the first compares θ and y at the same (n, s) ; the second compares (n, s) and $(n + b - 1, s - 1)$ at the same parameter y .

For the first factor, let $\lambda := \log(\theta(1 - y)/(y(1 - \theta))) > 0$. After canceling the binomial coefficients, the logarithm of $f_{n,\theta}(s)/f_{n,y}(s)$ is $s\lambda + n \log((1 - \theta)/(1 - y))$. The definition of D gives $\log((1 - \theta)/(1 - y)) = -D(y\|\theta) - \lambda y$, so

$$\frac{f_{n,\theta}(s)}{f_{n,y}(s)} = \left(\frac{\theta}{y}\right)^s \left(\frac{1 - \theta}{1 - y}\right)^{n-s} = \exp(-nD(y\|\theta) - \lambda(ny - s)). \quad (\text{EC.12})$$

To control the exponent, we bound $\lambda(ny - s)$ from below. Since $\epsilon \leq \theta/4$, we have $y \geq 3\theta/4$; the case condition then gives $s < n\theta/2 \leq 2ny/3$, so $ny - s \geq ny/3$. Also, $\lambda \geq \log(\theta/y) > 0$ because $(1 - y)/(1 - \theta) > 1$. These bounds yield

$$\lambda(ny - s) \geq \frac{n\lambda y}{3} \geq \frac{ny}{3} \log\left(\frac{\theta}{y}\right) \geq \frac{ny\epsilon}{3\theta} \geq \frac{n\epsilon}{4}.$$

The third inequality applies $\log u \geq 1 - 1/u$ with $u = \theta/y$, using $\theta - y = \epsilon$; the last uses $y \geq 3\theta/4$. Substituting this bound into (EC.12) and using $D(y\|\theta) \geq 0$ shows that the first factor is at most $e^{-n\epsilon/4}$.

For the second factor, expanding the masses at parameter y gives $f_{n,y}(s)/f_{n+b-1,y}(s-1) = ((n)/\binom{n+b-1}{s-1})y(1-y)^{-b}$. Here $\binom{n+b-1}{s-1} \geq \binom{n-1}{s-1}$ again bounds the coefficient ratio by n/s . Substituting the bounds on both factors into the preceding bound for the weighted odds gives

$$f_{n,\theta}(s) \frac{1 - \bar{q}_n(s)}{\bar{q}_n(s)} \leq e^{-n\epsilon/4} \frac{ny}{s} (1-y)^{-b} \leq 2^b y \frac{n}{s} e^{-n\epsilon/4}.$$

The last inequality uses $y \leq 1/2$. The factor n/s is now accompanied by exponential decay in n , which makes the contribution summable.

It remains to sum these weighted odds. The geometric-sum bound from the case $s=0$, applied with $\epsilon/4$, gives $\sum_{n \geq 1} n e^{-n\epsilon/4} \leq 16/\epsilon^2$. Also, $\sum_{s=1}^n s^{-1} \leq 1 + \log n \leq \log(eT)$ for $1 \leq n \leq T-1$. Extending each inner sum to all $1 \leq s \leq n$ and using $y \leq 1/2$ therefore gives

$$\sum_{n=1}^{T-1} \sum_{1 \leq s < n\theta/2} f_{n,\theta}(s) \frac{1 - \bar{q}_n(s)}{\bar{q}_n(s)} \leq 2^b y \sum_{n=1}^{T-1} n e^{-n\epsilon/4} \sum_{s=1}^n \frac{1}{s} \leq C 2^b \epsilon^{-2} \log(eT).$$

Combining the three contributions proves the bound, since $\log(e(T+b)) \leq \log(eT) + \log(b+1)$ for $T \geq 1$. The resulting constant is at most a universal multiple of $2^b(b+1)/a$, including the initial reduction of ϵ , and therefore depends only on (α_0, β_0) , not on θ , ϵ , or T . $\blacksquare$

Proof of Lemma 2. The lemma counts periods with an inaccurate selected coordinate, whereas concentration controls the number of retained comparisons. A product with few ticks may be selected repeatedly without advancing its clock; the lower bound on its tick probability controls the expected number of such selections. Once the clock has enough ticks, an inaccurate draw requires either an atypical empirical prefix or a sampling deviation from an accurate one. We first obtain the two concentration bounds, then count these three sources of error in calendar time.

Step 1 (Concentration at Deterministic Prefixes). Choose the cutoff $n_0 := \lceil C_0 \xi^{-2} \log(8N^2 T^3) \rceil$, where $C_0 := \max\{1600, 40(\alpha_0 + \beta_0)\}$. Beyond this length, both deviations will be rare and the prior's effect on the Beta mean will be small.

We begin with the sampling deviation. For $Y \sim \text{Beta}(a, b)$ with $a, b > 0$, let $m = a/(a+b)$ and represent $Y = G_a/(G_a + G_b)$, where $G_a \sim \text{Gamma}(a, 1)$ and $G_b \sim \text{Gamma}(b, 1)$ are independent with unit rate.

We first bound the upper tail. Let $x \in (m, 1)$. The event $Y \geq x$ is equivalent to $(1-x)G_a - xG_b \geq 0$. Choose $s = (x-m)/(x(1-x)) > 0$. The identities $1-s(1-x) = m/x > 0$ and $1+sx = (1-m/x)$

$m)/(1-x) > 0$ ensure that both Gamma moment generating functions are finite at the arguments used below. On the event $Y \geq x$, we have $e^{s((1-x)G_a - xG_b)} \geq 1$, so

$$\begin{aligned} \mathbb{P}(Y \geq x) &\stackrel{(a)}{\leq} \mathbb{E}[e^{s((1-x)G_a - xG_b)}] \\ &= \mathbb{E}[e^{s(1-x)G_a}] \mathbb{E}[e^{-sxG_b}] \\ &= (1 - s(1-x))^{-a} (1 + sx)^{-b} \\ &= \left(\frac{x}{m}\right)^a \left(\frac{1-x}{1-m}\right)^b \\ &= \exp(-(a+b)D(m\|x)). \end{aligned}$$

Inequality (a) follows from Chernoff's bound. Independence gives the product of expectations, and the Gamma moment generating functions give the next equality. The penultimate equality substitutes the identities for s above. The last uses $a = (a+b)m$ and $b = (a+b)(1-m)$ to identify the Bernoulli Kullback–Leibler divergence D defined in the proof of Lemma EC.2.

For $x \in (0, m)$, the same choice has $s < 0$; applying Chernoff's bound on the lower-tail event $Y \leq x$, equivalently $(1-x)G_a - xG_b \leq 0$, gives the corresponding bound.

For Bernoulli distributions, Pinsker's inequality gives $D(m\|x) \geq 2(m-x)^2$ for $x \in (0, 1)$. With $x = m + u$ or $x = m - u$, each nonempty tail is therefore bounded by $\exp(-2(a+b)u^2)$. Combining the two tails, for every $u > 0$,

$$\mathbb{P}(|Y - m| > u) \leq \mathbb{P}(Y > m + u) + \mathbb{P}(Y < m - u) \leq 2\exp(-2(a+b)u^2). \quad (\text{EC.13})$$

The first inequality is subadditivity, and the second uses the two one-sided bounds. The upper tail is empty when $m + u \geq 1$, and the lower tail when $m - u \leq 0$.

We next control empirical deviations at every possible long prefix. Proposition EC.1 represents each deterministic prefix by independent Bernoulli(θ_i) variables $Z_{i,k} \in \{0, 1\} \subset [0, 1]$. Thus,

$$\mathbb{P}_v \left(\exists n \in \{n_0, \dots, T-1\} : \left| \frac{1}{n} \sum_{k=1}^n Z_{i,k} - \theta_i \right| > \xi/40 \right) \stackrel{(a)}{\leq} 2T \exp(-n_0 \xi^2/800) \stackrel{(b)}{\leq} \frac{1}{4NT^2}. \quad (\text{EC.14})$$

Inequality (a) applies Hoeffding's inequality at each deterministic prefix and a union bound over at most T lengths, using $n \geq n_0$. The definition of n_0 , with $C_0 \geq 1600$, gives (b). If $n_0 \geq T$, the index set is empty. Each product's bound refers to its own marginal construction; summing these probabilities does not require joint independence.

The mean of the Beta law in (EC.1) is the center of the concentration bound (EC.13). At a prefix of length $n \geq n_0$, its difference from the empirical frequency satisfies

$$\left| \frac{\alpha_0 + \sum_{k=1}^n Z_{i,k}}{\alpha_0 + \beta_0 + n} - \frac{1}{n} \sum_{k=1}^n Z_{i,k} \right| = \frac{|\alpha_0 n - (\alpha_0 + \beta_0) \sum_{k=1}^n Z_{i,k}|}{n(\alpha_0 + \beta_0 + n)} \leq \frac{\alpha_0 + \beta_0}{n} \leq \frac{\xi}{40}. \quad (\text{EC.15})$$

For the first inequality, $0 \leq \sum_{k=1}^n Z_{i,k} \leq n$ places the expression inside the numerator's absolute value between $-\beta_0 n$ and $\alpha_0 n$. Hence, the numerator is at most $(\alpha_0 + \beta_0)n$, whereas the denominator $n(\alpha_0 + \beta_0 + n)$ is at least n^2 . The last inequality follows from $C_0 \geq 40(\alpha_0 + \beta_0)$ and $\xi \leq 1/2$.

On an accurate prefix, $|\frac{1}{n} \sum_{k=1}^n Z_{i,k} - \theta_i| \leq \xi/40$. To compare the Beta mean with θ_i , insert the empirical frequency:

$$\left| \frac{\alpha_0 + \sum_{k=1}^n Z_{i,k}}{\alpha_0 + \beta_0 + n} - \theta_i \right| \leq \left| \frac{\alpha_0 + \sum_{k=1}^n Z_{i,k}}{\alpha_0 + \beta_0 + n} - \frac{1}{n} \sum_{k=1}^n Z_{i,k} \right| + \left| \frac{1}{n} \sum_{k=1}^n Z_{i,k} - \theta_i \right| \leq \frac{\xi}{40} + \frac{\xi}{40} = \frac{\xi}{20}.$$

The first inequality is the triangle inequality. For the second, (EC.15) bounds the first term by $\xi/40$, and prefix accuracy bounds the second by $\xi/40$.

We now relate attraction errors to deviations of the transformed draw. Since $v_i \leq 1$ and $\xi \leq 1/2$,

$$|\tilde{v}_i - v_i| > \xi \implies \left| \frac{\tilde{v}_i}{1 + \tilde{v}_i} - \theta_i \right| > \xi/5.$$

Indeed, the endpoint differences are $\xi/\{(1 + v_i)(1 + v_i \pm \xi)\}$, whose denominators are at most five.

For an accurate prefix with $n_{i,t} \geq n_0$, an attraction error places the transformed draw more than $\xi/5 - \xi/20 > \xi/10$ from its Beta mean. At these realized counts, its conditional Beta law has total shape $\alpha_0 + \beta_0 + n_{i,t} \geq n_{i,t}$, so (EC.13) gives

$$\mathbb{P}_t(|\tilde{v}_{i,t} - v_i| > \xi) \leq 2e^{-n_{i,t}\xi^2/50} \leq 2e^{-n_0\xi^2/50} \leq \frac{1}{8N^2T^3} \quad (\text{EC.16})$$

The last inequality follows from the definition of n_0 .

Step 2 (Counting Errors on Selected Coordinates). First consider selections made before a product's clock reaches n_0 . Conditional on $\mathcal{F}_{t-1}$ and $\tilde{v}_t$, the count and selected assortment are fixed. Multiplying (EC.2) by $\mathbb{1}\{n_{i,t} < n_0\}$ therefore bounds the conditional expectation of $\mathbb{1}\{i \in S_t, n_{i,t} < n_0, I_t \in \{i, 0\}\}$ from below by $\mathbb{1}\{i \in S_t, n_{i,t} < n_0\}/(K + 1)$. Taking unconditional expectations by the law of iterated expectations, summing over t , and rearranging gives

$$\mathbb{E}_v \left[\sum_{t=1}^T \mathbb{1}\{i \in S_t, n_{i,t} < n_0\} \right] \leq (K + 1) \mathbb{E}_v \left[\sum_{t=1}^T \mathbb{1}\{i \in S_t, n_{i,t} < n_0, I_t \in \{i, 0\}\} \right] \leq (K + 1)n_0. \quad (\text{EC.17})$$

The last inequality holds pathwise: each term inside the second sum advances the local count, so at most n_0 such terms can occur before it reaches n_0 .

For a fixed product, the event in (EC.14) includes every visit to a long, atypical empirical prefix. On each path, at most T periods can be charged to this event. Its contribution is therefore at most $T/(4NT^2)$ per product, or $1/(4T)$ after summing over products.

On histories with a long, accurate prefix, (EC.16) bounds the conditional probability of sampling error. Dropping the selection restriction, applying iterated expectations, and summing over the NT product-period pairs gives a contribution of at most $NT/(8N^2T^3) = 1/(8NT^2)$.

Every error on a selected coordinate belongs to one of these three categories. A union bound and the estimates above give

$$\mathbb{E}_v \left[\sum_{t=1}^T \mathbf{1} \{ \exists i \in S_t : |\tilde{v}_{i,t} - v_i| > \xi \} \right] \leq N(K+1)n_0 + \frac{1}{4T} + \frac{1}{8NT^2}.$$

Since $n_0 \leq C\xi^{-2} \log(eNT)$ and $N, T \geq 1$, this proves (4). $\blacksquare$

EC.4.1. Extension to Epoch Decisions

The independent conjugate epoch policy draws from (EC.1) at each epoch boundary and retains the resulting assortment until an outside-option choice ends the epoch, or the horizon ends. The next draw incorporates all purchases and the terminating outside choice. With $\alpha_0 = \beta_0 = 1$, this is Algorithm 1 of Agrawal et al. (2026), using the assortment oracle of Section 3.

COROLLARY EC.2 (Regret under the Epoch Clock). *For $v \in [v_{\min}, 1]^N$ and fixed $\alpha_0, \beta_0 > 0$, the independent conjugate epoch policy satisfies the bound of Theorem 1, with C depending only on $(v_{\min}, \alpha_0, \beta_0)$.*

The repair argument applies at each fresh draw. Its cost must be weighted by the time for which the assortment is retained; only the accounting for durations and starting prefixes changes from the proof of Theorem 1.

Proof of Corollary EC.2. Use the failed-repair events and odds from the proof of Theorem 1.

Let τ_ℓ be the start of epoch ℓ and L_ℓ its number of periods observed through T . All sums below range over epochs with $\tau_\ell \leq T$, and $\sum_\ell L_\ell = T$. An epoch start is determined by past customer choices: $\{\tau_\ell = t\} \in \mathcal{F}_{t-1}$. We apply the deterministic-time conditional sampling law and failed-repair comparison (5) on each event $\{\tau_\ell = t\}$, then sum over t to obtain the same conditional statements for the fresh draw at τ_ℓ .

An error at an epoch start persists for the time that the assortment is retained. To bound this cost, condition on the boundary history and its draw. The assortment is then fixed, and the untruncated duration is geometric with outside-choice probability $(1 + \sum_{j \in S_{\tau_\ell}} v_j)^{-1}$. Truncating at T can only reduce its mean, so

$$\mathbb{E}_v[L_\ell \mid \mathcal{F}_{\tau_\ell-1}, \tilde{v}_{\tau_\ell}] \leq 1 + \sum_{j \in S_{\tau_\ell}} v_j \leq K+1. \quad (\text{EC.18})$$

Let $0 < \xi \leq v_{\min}/2$. The failed-repair event $\mathcal{B}_{i,\tau_\ell}(\xi)$ means that product i is repairable but omitted, as defined in Step 1 of the theorem proof. Its indicator is fixed by the boundary history and draw, so it can be taken outside the conditional expectation in (EC.18). Averaging over the draw by the law of iterated expectations gives $\mathbb{E}_v[L_\ell \mathbb{1}\{\mathcal{B}_{i,\tau_\ell}(\xi)\} \mid \mathcal{F}_{\tau_\ell-1}] \leq (K+1)\mathbb{P}_v(\mathcal{B}_{i,\tau_\ell}(\xi) \mid \mathcal{F}_{\tau_\ell-1})$.

Two epochs selecting product i cannot start at the same prefix. Before the later one begins, the earlier epoch has ended with an outside comparison for i . All starting prefixes have length at most $T-1$; the final incomplete epoch uses its starting prefix only once. For $n = n_{i,\tau_\ell}$, Proposition EC.1 and the scalar odds from Step 2 of the theorem proof give $\omega_{i,\tau_\ell}(\xi) = \omega_n(\sum_{k=1}^n Z_{i,k}, \epsilon_i)$. Consequently,

$$\begin{aligned} \mathbb{E}_v \left[\sum_{\ell} L_\ell \mathbb{1}\{\mathcal{B}_{i,\tau_\ell}(\xi)\} \right] &\stackrel{(a)}{\leq} (K+1) \mathbb{E}_v \left[\sum_{\ell} \mathbb{1}\{\mathcal{B}_{i,\tau_\ell}(\xi)\} \right] \\ &\stackrel{(b)}{\leq} (K+1) \mathbb{E}_v \left[\sum_{\ell} \omega_{i,\tau_\ell}(\xi) \mathbb{1}\{i \in S_{\tau_\ell}\} \right] \\ &\stackrel{(c)}{\leq} (K+1) \sum_{n=0}^{T-1} \mathbb{E}_v \left[\omega_n \left(\sum_{k=1}^n Z_{i,k}, \epsilon_i \right) \right] \\ &\stackrel{(d)}{\leq} \frac{C(K+1)}{\xi^2} \log(eT). \end{aligned}$$

Inequality (a) averages the conditional estimate above over boundary histories and sums over epochs; (b) applies the failed-repair comparison (5). For (c), extend the nonnegative sum from the distinct starting prefixes to all deterministic prefixes on each path, then take expectations. Inequality (d) applies (EC.3) with $\epsilon_i \geq \xi/4$. Thus the factor $K+1$ accounts for duration, and no additional factor is needed to pass from selected epochs to prefixes.

For selected-coordinate errors, retain the cutoff $n_0 \leq C\xi^{-2} \log(eNT)$ from the proof of Lemma 2. A prefix is short when it has fewer than n_0 comparisons; a longer prefix is accurate when $|\frac{1}{n} \sum_{k=1}^n Z_{i,k} - \theta_i| \leq \xi/40$. These distinctions give the same three contributions as in that lemma.

First, a product can be selected at most n_0 epochs starting at a short prefix, because those prefixes are distinct. By (EC.18), their total expected duration is at most $(K+1)n_0$ per product.

Second, (EC.14) bounds the probability that any deterministic prefix of length $n \in \{n_0, \dots, T-1\}$ is atypical by $1/(4NT^2)$ per product. Outside this event, every adaptively visited long boundary prefix is accurate. On any path, all epochs starting at long, atypical prefixes together last at most T observed periods. Their contribution, summed over products, is therefore at most $1/(4T)$. This uses the pathwise duration bound, without conditioning epoch lengths on a future concentration event.

Third, at a long, accurate boundary prefix, (EC.16) bounds the conditional sampling-error probability by $1/(8N^2T^3)$. The conditioning argument used in (a) bounds the expected time spent in this error by $(K+1)/(8N^2T^3)$ per product and epoch start. Summing over at most NT such pairs gives $(K+1)/(8NT^2) \leq 1/(4T^2)$, since $K \leq N$. Adding these three contributions yields

$$\mathbb{E}_v \left[\sum_{\ell} L_{\ell} \mathbb{1}\{\exists i \in S_{\tau_{\ell}} : |\tilde{v}_{i,\tau_{\ell}} - v_i| > \xi\} \right] \leq N(K+1)n_0 + \frac{1}{4T} + \frac{1}{4T^2} \leq \frac{CN(K+1)}{\xi^2} \log(eNT) + C.$$

Each epoch keeps the revenue gap $\Delta_{\tau_{\ell}}$ fixed for its observed duration. For $g \in (0, 1]$, take $\xi = v_{\min}g/4$. If neither a selected-coordinate error nor a failed repair occurs at the epoch start, Lemma 1(ii) gives $\Delta_{\tau_{\ell}} \leq 2\xi/v_{\min} = g/2$. Hence a gap above g requires at least one of these errors. Summing the failed-repair bound over the N products and adding the selected-error bound therefore gives (7) with $\sum_{\ell} L_{\ell} \mathbb{1}\{\Delta_{\tau_{\ell}} > g\}$ in place of the period sum. This duration-weighted count is also at most $\sum_{\ell} L_{\ell} = T$, so the tail integration in Step 3 of the theorem proof completes the argument. $\blacksquare$

EC.5. Bayesian Regret of the Exact-Posterior Policy

This section proves Lemma EC.3. It uses the notation of Section EC.3.2; throughout, Assumption 2 holds, and $C > 0$ may change from line to line and is numerical.

Proof of Lemma EC.3. Condition on the realized history through $t-1$. We first bound the curvature of the log-coordinate posterior; this will control its fluctuations in each pairwise direction. Set $u_0 := 0$ for the outside option. Under $v_i = e^{u_i}$, the prior-Jacobian factor is $\exp(\alpha_0 u_i - \beta_0 e^{u_i})$, while observation τ in (8) contributes $\exp(\mathbb{1}\{I_{\tau} \neq 0\}u_{I_{\tau}})/(1 + \sum_{j \in S_{\tau}} e^{u_j})$. Hence the negative log-density ϕ_t of the posterior law of U on $\mathcal{U}$ satisfies

$$\nabla^2 \phi_t(u) \stackrel{(a)}{=} \beta_0 \text{diag}(e^{u_1}, \dots, e^{u_N}) + \sum_{\tau=1}^{t-1} F_{S_{\tau}}(e^u) \stackrel{(b)}{\succeq} \beta_0 v_{\min} I_N + v_{\min} \sum_{\tau=1}^{t-1} F_{S_{\tau}}(\mathbf{1}) \stackrel{(c)}{=} v_{\min} H_t. \quad (\text{EC.19})$$

The rate and log-denominator terms produce the diagonal and Fisher matrices in (a); the terms involving α_0 and the observed purchases are affine and disappear upon differentiation. The bound $e^{u_i} \geq v_{\min}$ and the comparison (EC.5) give (b), while (c) uses the definition of H_t with $\kappa = \beta_0$. In particular, the curvature bound is independent of α_0 .

Since (EC.4) expresses every Fisher width through pairwise differences, we first obtain concentration for a fixed pair $0 \leq a < b \leq N$. Set $\ell_{ab} := e_a - e_b$ and $q_{ab,t} := \ell_{ab}^{\top} H_t^{-1} \ell_{ab}$. Since $H_t \succeq \beta_0 I_N \succ 0$ and $\ell_{ab} \neq 0$, we have $q_{ab,t} > 0$. The scale $\sqrt{q_{ab,t}}$ will normalize the error for this pair.

For $s \in \mathbb{R}$, define the centered log moment generating function $\Lambda_{ab}(s) := \log \mathbb{E}_t[\exp(s \ell_{ab}^{\top} (U - \mathbb{E}_t[U]))]$. Differentiating this function introduces the exponentially tilted posterior, whose density

is proportional to $\exp(s\ell_{ab}^\top u - \phi_t(u))$; write $\mathbb{E}_{t,s}$ and $\text{Var}_{t,s}$ for expectation and variance under this law. Boundedness of U makes the normalizing constant positive and finite for every $s \in \mathbb{R}$ and permits differentiation under the expectation, giving $\Lambda'_{ab}(s) = \mathbb{E}_{t,s}[\ell_{ab}^\top U] - \mathbb{E}_t[\ell_{ab}^\top U]$. A second derivative gives the tilted variance in the display below.

The tilt adds only a linear term to the potential, so its Hessian remains $\nabla^2 \phi_t$. On the open convex box $\mathcal{U}^\circ = (\log v_{\min}, 0)^N$, the tilted potential is C^2 and strictly convex by (EC.19). The boundary has zero probability under the posterior and its tilted laws. The Brascamp–Lieb variance inequality in Proposition 2.1 of Bobkov and Ledoux (2000) holds on open convex domains. With the linear function $\ell_{ab}^\top u$, it therefore gives

$$\Lambda''_{ab}(s) = \text{Var}_{t,s}(\ell_{ab}^\top U) \leq \mathbb{E}_{t,s}[\ell_{ab}^\top (\nabla^2 \phi_t(U))^{-1} \ell_{ab}] \leq v_{\min}^{-1} q_{ab,t}.$$

Inverting (EC.19) reverses the Loewner order, so $(\nabla^2 \phi_t(u))^{-1} \preceq v_{\min}^{-1} H_t^{-1}$; the quadratic form is $q_{ab,t}$ by definition, giving the last inequality. Since $\Lambda_{ab}(0) = \Lambda'_{ab}(0) = 0$, Taylor's theorem at zero and the uniform bound on Λ''_{ab} yield $\Lambda_{ab}(s) \leq s^2 v_{\min}^{-1} q_{ab,t} / 2$ for all $s \in \mathbb{R}$.

Conditional on $\mathcal{F}_{t-1}$, U and $\tilde{U}_t$ are independent draws from the same log-coordinate posterior. Their centering terms cancel in the difference, and factoring their moment generating functions gives

$$\mathbb{E}_t \left[\exp \left(s \frac{\ell_{ab}^\top (\tilde{U}_t - U)}{\sqrt{q_{ab,t}}} \right) \right] = \exp \left(\Lambda_{ab} \left(\frac{s}{\sqrt{q_{ab,t}}} \right) + \Lambda_{ab} \left(-\frac{s}{\sqrt{q_{ab,t}}} \right) \right) \leq \exp(s^2 v_{\min}^{-1}).$$

Dividing by $\sqrt{q_{ab,t}}$ has made the bound independent of the pair. To control all these normalized errors simultaneously, set $h_t := \tilde{U}_t - U$ and define

$$Z_t := \max_{0 \leq a < b \leq N} \frac{|\ell_{ab}^\top h_t|}{\sqrt{q_{ab,t}}}.$$

For $u > 0$, Chernoff's inequality with $s = v_{\min} \sqrt{u} / 2$ bounds the upper-tail probability at level $\sqrt{u}$ by $\exp(-v_{\min} u / 4)$; using $-s$ gives the same lower-tail bound. Since Z_t maximizes over the $N(N+1)/2$ unordered pairs, including the outside option, $\{Z_t^2 > u\}$ is the union of these tails. Subadditivity gives

$$\mathbb{P}_t(Z_t^2 > u) \leq \sum_{0 \leq a < b \leq N} 2e^{-v_{\min} u / 4} = N(N+1) \exp(-v_{\min} u / 4).$$

To obtain the second moment, set $z_0^2 := 4v_{\min}^{-1} \log(N(N+1))$, the threshold at which this exponential bound equals one. The tail-integration formula, using the probability bound one below z_0^2 and the exponential bound above it, yields

$$\mathbb{E}_t[Z_t^2] \leq z_0^2 + \int_{z_0^2}^{\infty} N(N+1) e^{-v_{\min} u / 4} du = 4v_{\min}^{-1} (1 + \log(N(N+1))) \leq C v_{\min}^{-1} \log(eN).$$

The last inequality uses $\log(N(N+1)) \leq 2\log(eN)$ and $\log(eN) \geq 1$, proving the moment bound in the statement.

It remains to translate the simultaneous pairwise control into the pathwise Fisher-width bound. By definition, $(\ell_{ab}^\top h_t)^2 \leq Z_t^2 q_{ab,t}$ for every pair. Thus, for any feasible assortment S ,

$$h_t^\top F_S(\mathbf{1}) h_t = \sum_{\substack{0 \leq a < b \leq N \\ a, b \in S \cup \{0\}}} p_a(S, \mathbf{1}) p_b(S, \mathbf{1}) (\ell_{ab}^\top h_t)^2 \leq Z_t^2 \sum_{\substack{0 \leq a < b \leq N \\ a, b \in S \cup \{0\}}} p_a(S, \mathbf{1}) p_b(S, \mathbf{1}) q_{ab,t} = Z_t^2 \text{tr}(H_t^{-1} F_S(\mathbf{1})).$$

The first equality is (EC.4), and the nonnegative pair weights preserve the bound on each squared displacement. The same pairwise identity gives the matrix representation $F_S(\mathbf{1}) = \sum_{a < b: a, b \in S \cup \{0\}} p_a(S, \mathbf{1}) p_b(S, \mathbf{1}) \ell_{ab} \ell_{ab}^\top$. Substituting $q_{ab,t} = \text{tr}(H_t^{-1} \ell_{ab} \ell_{ab}^\top)$ and using linearity of the trace therefore gives the last equality. Since the same Z_t controls all pairs, (EC.6) holds simultaneously for all feasible S , including the posterior-selected S_t , without an independence assumption. ■

EC.6. Covariance-Corrected Gamma Mean-Field Sampling

The corrected policy retains the mean-field recursion of Section 6 and adds two statistics: an outside-comparison count to certify the sampling center, and a diagonal Fisher statistic to set its variance. Its regret guarantee uses this certificate without requiring convergence of the raw means. Section EC.6.3 then calibrates the corrected dispersion against the complete posterior on a fixed assortment.

EC.6.1. Policy Construction

Mean-Field Recursion. The augmentation in Proposition 2 applies here with the truncation indicators removed. Under a product approximation, taking the expected log density over the attraction factors gives an Exponential exposure factor with rate $1 + \sum_{j \in S_\tau} \mu_j$. Its mean, $(1 + \sum_{j \in S_\tau} \mu_j)^{-1}$, is the exposure used by the filter. The online recursion freezes it at the means prevailing when the assortment was offered, rather than revisiting past exposures as in batch variational updates.

Retain $a_{i,t} = \alpha_0 + V_{i,t}$, $b_{i,t} = \beta_0 + \Upsilon_{i,t}$, and $\mu_{i,t} = a_{i,t}/b_{i,t}$. The updates in Algorithm 3 imply $\mu_{i,t+1} - \mu_{i,t} = (\mathbb{1}\{I_t = i\} - p_i(S_t, \mu_t))/(b_{i,t} + p_0(S_t, \mu_t))$ for $i \in S_t$, using $p_i(S_t, \mu_t) = \mu_{i,t} p_0(S_t, \mu_t)$. Its numerator is the MNL log-likelihood score in coordinate $\log v_i$. The certificate below controls the sampling center derived from this recursion.

Certified Sampling Center. Use the counts $V_{i,t}$ and $N_{i,t}$ from the clock construction, where $N_{i,t}$ counts outside choices while i was offered. Thus $n_{i,t} = V_{i,t} + N_{i,t}$ is the retained-prefix length and $\theta_i = v_i/(1 + v_i)$. Let $\theta_{\min} := v_{\min}/(1 + v_{\min})$ and $h_n := \sqrt{\log(2NT^2)/(2n)}$ for $n \geq 1$. When $n_{i,t} \geq 1$, intersect $[V_{i,t}/n_{i,t} - h_{n_{i,t}}, V_{i,t}/n_{i,t} + h_{n_{i,t}}]$ with $[\theta_{\min}, 1/2]$, and transform each endpoint x to

$x/(1-x)$. Denote the resulting interval by $C_{i,t} = [\ell_{i,t}, u_{i,t}]$; if the intersection is empty or $n_{i,t} = 0$, set $C_{i,t} = [v_{\min}, 1]$. Define

$$m_{i,t} := \min\{1, \text{proj}_{C_{i,t}}(\mu_{i,t}) + u_{i,t} - \ell_{i,t}\}. \quad (\text{EC.20})$$

The projection belongs to $[\ell_{i,t}, u_{i,t}]$, so adding the width gives a value between $u_{i,t}$ and $2u_{i,t} - \ell_{i,t}$. Consequently, whenever $v_i \in C_{i,t}$,

$$v_i \leq m_{i,t} \leq v_i + 2(u_{i,t} - \ell_{i,t}). \quad (\text{EC.21})$$

Capping at one preserves both bounds because $v_i \leq 1$.

Fisher Statistic and Sampling Law. For a nonempty assortment $S \in \mathcal{S}$ and attraction vector $x \in \mathbb{R}_{++}^N$, let $p_S = (p_i(S, x))_{i \in S}$. The $S \times S$ block of the log-coordinate Fisher information is $\text{diag}(p_S) - p_S p_S^\top$. Since $\sum_{i \in S} p_i(S, x) = 1 - p_0(S, x)$, Cauchy–Schwarz gives $(p_S^\top z)^2 \leq (1 - p_0(S, x)) z^\top \text{diag}(p_S) z$ for every $z \in \mathbb{R}^{|S|}$, and hence

$$p_0(S, x) \text{diag}(p_S) \preceq \text{diag}(p_S) - p_S p_S^\top \preceq \text{diag}(p_S).$$

The inverse of the lower matrix dominates the inverse Fisher information on the offered coordinates. This motivates accumulating $p_0(S, x)p_i(S, x)$ as the information available for an independent draw of product i .

Evaluate these probabilities at the certified center m_t , so that $p_i(S_t, m_t) = m_{i,t}p_0(S_t, m_t)$ for $i \in S_t$. For fixed $\lambda_0 \geq c \geq 1$, set

$$J_{i,t} := \lambda_0 + \sum_{\tau < t: i \in S_\tau} p_0(S_\tau, m_\tau) p_i(S_\tau, m_\tau), \quad \tilde{v}_{i,t} \mid \mathcal{F}_{t-1} \sim \text{Gamma}\left(\frac{J_{i,t}}{c}, \frac{J_{i,t}}{cm_{i,t}}\right) \quad (\text{EC.22})$$

independently across products, using shape and rate. The conditional mean is $m_{i,t}$ and the variance is $cm_{i,t}^2/J_{i,t}$.

The four stored statistics are (a_i, b_i, N_i, J_i) ; the mean $\mu_i = a_i/b_i$ is derived from them. Both updates use probabilities evaluated before the choice. Updating these statistics costs $O(|S_t|)$ and does not revisit the history.

EC.6.2. Regret Bound

Proof of Theorem 3. We use the one-coordinate repair argument of Theorem 1. An appreciable revenue loss requires either an inaccurate selected draw or a repairable product that remains unselected. Both failures can be charged to selections before the clock or Fisher statistic reaches a concentration threshold, apart from rare sampling errors.

Algorithm 4 Covariance-Corrected Gamma Mean-Field Thompson Sampling

-
- 1: **Input:** horizon T , capacity K , revenues $\{r_i\}_{i=1}^N$, prior (α_0, β_0) , lower bound $v_{\min}$, and constants $\lambda_0 \geq c \geq 1$.
 - 2: **Initialize:** $a_i \leftarrow \alpha_0$, $b_i \leftarrow \beta_0$, $N_i \leftarrow 0$, $J_i \leftarrow \lambda_0$, and $\mu_i \leftarrow a_i/b_i$, $\forall i \in \mathcal{N}$.
 - 3: **for** $t = 1, \dots, T$ **do**
 - 4: Construct $C_i = [\ell_i, u_i]$ from $V_i = a_i - \alpha_0$ and $n_i = V_i + N_i$, and compute m_i from (EC.20), $\forall i \in \mathcal{N}$.
 - 5: Draw $\tilde{v}_i \sim \text{Gamma}(J_i/c, J_i/(cm_i))$ independently, $\forall i \in \mathcal{N}$, and offer $S_t \leftarrow S^*(\tilde{v})$.
 - 6: Set $\hat{p}_0 \leftarrow p_0(S_t, \mu)$ and evaluate $p_0(S_t, m)$ once; then observe $I_t \in S_t \cup \{0\}$.
 - 7: For each $i \in S_t$, update $a_i \leftarrow a_i + \mathbf{1}\{I_t = i\}$, $b_i \leftarrow b_i + \hat{p}_0$, $N_i \leftarrow N_i + \mathbf{1}\{I_t = 0\}$, $J_i \leftarrow J_i + p_0(S_t, m)p_i(S_t, m)$, and $\mu_i \leftarrow a_i/b_i$.
 - 8: **end for**
-

Throughout, $\mathbb{P}_t$ and $\mathbb{E}_t$ condition on $\mathcal{F}_{t-1}$ under the fixed true vector v . Constants C may change from line to line and depend only on $v_{\min}$. Let $0 < \xi \leq v_{\min}/2$ be an attraction-error tolerance.

Step 1 (Coverage and Short Clocks). Recall that $h_n = \sqrt{\log(2NT^2)/(2n)}$, for $n \geq 1$, is the confidence radius for the empirical comparison frequency. Let $\mathcal{C}_t$ be the history event that $|V_{i,t}/n_{i,t} - \theta_i| \leq h_{n_{i,t}}$ for every product with $n_{i,t} \geq 1$. Proposition EC.1 represents each count by a prefix of independent Bernoulli variables in $[0, 1]$. Hoeffding's inequality gives $\mathbb{P}_v(|n^{-1} \sum_{k=1}^n Z_{i,k} - \theta_i| > h_n) \leq 2\exp(-2nh_n^2) = 1/(NT^2)$ for every deterministic $1 \leq n \leq T$. Failure of $\mathcal{C}_t$ at any period implies failure at one of these NT product-prefix pairs. A union bound gives

$$\mathbb{P}_v \left(\bigcup_{t=1}^T \mathcal{C}_t^c \right) \leq \sum_{i=1}^N \sum_{n=1}^T \mathbb{P}_v \left(\left| n^{-1} \sum_{k=1}^n Z_{i,k} - \theta_i \right| > h_n \right) \leq \frac{NT}{NT^2} = \frac{1}{T}. \quad (\text{EC.23})$$

Consequently, $\sum_{t=1}^T \mathbb{P}_v(\mathcal{C}_t^c) \leq T\mathbb{P}_v(\bigcup_{t=1}^T \mathcal{C}_t^c) \leq 1$. The union uses deterministic prefixes from each product's marginal construction, without conditioning their binomial law on an adaptive count.

On $\mathcal{C}_t$, a product with $n_{i,t} \geq 1$ has θ_i in both its empirical interval and $[\theta_{\min}, 1/2]$. The construction of $C_{i,t}$ therefore gives $v_i = \theta_i/(1 - \theta_i) \in C_{i,t}$. At zero comparisons, the fallback interval $[v_{\min}, 1]$ also contains v_i . Hence (EC.21) gives $m_{i,t} \geq v_i$.

Still on $\mathcal{C}_t$, consider the transformation $x/(1-x)$ from comparison probabilities to attractions. Its derivative is $1/(1-x)^2 \leq 4$ on $[0, 1/2]$. For $n_{i,t} \geq 1$, the intersected probability interval has width at most $2h_{n_{i,t}}$. The mean value theorem therefore gives $u_{i,t} - \ell_{i,t} \leq 4(2h_{n_{i,t}}) = 8h_{n_{i,t}}$, and the center bound yields $m_{i,t} - v_i \leq 2(u_{i,t} - \ell_{i,t}) \leq 16h_{n_{i,t}}$. Thus, with $n_\xi := \lceil 512 \log(2NT^2)/\xi^2 \rceil$, reaching $n_{i,t} \geq n_\xi$ ensures $0 \leq m_{i,t} - v_i \leq \xi/2$ on $\mathcal{C}_t$.

The counting argument for (EC.17) uses only the tick probability in (EC.2) and the fact that at most n_0 ticks occur before the cutoff. It therefore applies to this policy with any positive integer cutoff; taking $n_0 = n_\xi$ gives

$$\mathbb{E}_v \left[\sum_{t=1}^T \mathbf{1}\{i \in S_t, n_{i,t} < n_\xi\} \right] \leq (K+1)n_\xi. \quad (\text{EC.24})$$

Step 2 (Gamma Deviations and Repair Probabilities). We first control relative deviations of $G_k \sim \text{Gamma}(k, 1)$ from its mean k , where $k > 0$ and the rate is one. Its moment generating function is $\mathbb{E}[e^{sG_k}] = (1-s)^{-k}$ for $s < 1$. Let $0 < \delta \leq 1/4$. For the upper tail, choose $s = \delta/(1+\delta) \in (0, 1)$. On $G_k \geq k(1+\delta)$, the exponential satisfies $e^{sG_k} \geq e^{sk(1+\delta)}$, so

$$\mathbb{P}(G_k \geq k(1+\delta)) \stackrel{(a)}{\leq} e^{-sk(1+\delta)} \mathbb{E}[e^{sG_k}] = \exp(-sk(1+\delta) - k \log(1-s)) = \exp(-k(\delta - \log(1+\delta))).$$

Inequality (a) is Chernoff's bound. The first equality uses the Gamma moment generating function, and the last substitutes $s(1+\delta) = \delta$ and $1-s = 1/(1+\delta)$.

For the lower tail, choose $s = -\delta/(1-\delta) < 0$. The sign of s ensures that $G_k \leq k(1-\delta)$ implies $e^{sG_k} \geq e^{sk(1-\delta)}$. The same exponential comparison therefore gives $\mathbb{P}(G_k \leq k(1-\delta)) \leq \exp(-k(-\delta - \log(1-\delta)))$.

We next bound the exponent terms below by quadratic functions of δ . For the upper tail,

$$\delta - \log(1+\delta) = \int_0^\delta \frac{u}{1+u} du \geq \frac{1}{1+\delta} \int_0^\delta u du = \frac{\delta^2}{2(1+\delta)} \geq \frac{\delta^2}{3}.$$

The first inequality uses $1+u \leq 1+\delta$ for $u \in [0, \delta]$, and the last uses $\delta \leq 1/4$. For the lower tail, $-\delta - \log(1-\delta) = \int_0^\delta u/(1-u) du \geq \int_0^\delta u du = \delta^2/2$, since $1-u \leq 1$. Combining the two tail bounds by subadditivity yields

$$\mathbb{P}(|G_k - k| > k\delta) \leq e^{-k\delta^2/3} + e^{-k\delta^2/2} \leq 2e^{-k\delta^2/3}.$$

We now apply this relative-deviation bound to the policy's draw. Conditional on $\mathcal{F}_{t-1}$, the policy draws $\tilde{v}_{i,t} = m_{i,t}G_k/k$ with $k = J_{i,t}/c$. Since $|\tilde{v}_{i,t} - m_{i,t}| = (m_{i,t}/k)|G_k - k|$, the event $|\tilde{v}_{i,t} - m_{i,t}| > \xi/2$ is equivalent to $|G_k - k| > k\delta$ with $\delta = \xi/(2m_{i,t})$. The bounds $m_{i,t} \in [v_{\min}, 1]$ and $\xi \leq v_{\min}/2$ ensure $\delta \in [\xi/2, 1/4]$, as required.

To make this error probability at most $1/(NT^2)$, set $j_\xi := 12c \log(2NT^2)/\xi^2$. On a history with $J_{i,t} \geq j_\xi$, the relative-deviation bound gives

$$\mathbb{P}_t(|\tilde{v}_{i,t} - m_{i,t}| > \xi/2) \leq 2e^{-k\delta^2/3} \leq 2 \exp\left(-\frac{J_{i,t}\xi^2}{12c}\right) \leq \frac{1}{NT^2}. \quad (\text{EC.25})$$

The second inequality uses $k = J_{i,t}/c$ and $\delta \geq \xi/2$; the last follows from the definition of j_ξ .

We now count selections before this Fisher threshold is reached. By the definition of $J_{i,t}$ in (EC.22), offering product i adds $p_0(S_t, m_t)p_i(S_t, m_t)$ to its Fisher statistic. Substituting the fitted choice probabilities gives

$$J_{i,t+1} - J_{i,t} = p_0(S_t, m_t)p_i(S_t, m_t) = \frac{m_{i,t}}{(1 + \sum_{j \in S_t} m_{j,t})^2} \geq \frac{v_{\min}}{(K+1)^2},$$

using $m_{i,t} \geq v_{\min}$, $m_{j,t} \leq 1$, and $|S_t| \leq K$. Dividing the threshold by this minimum increment, with one additional selection allowed to cross it, gives the pathwise count

$$\sum_{t=1}^T \mathbb{1}\{i \in S_t, J_{i,t} < j_\xi\} \leq 1 + \frac{(K+1)^2 j_\xi}{v_{\min}}. \quad (\text{EC.26})$$

We next bound the probability that a repairable product remains unselected. As in the proof of Theorem 1, let $\mathcal{B}_{i,t}(\xi) := \{\tilde{v}_{-i,t} \in \mathcal{D}_i(\xi), i \notin S_t\}$ denote this event. For the Gamma draws, let $q_{i,t}(\xi) := \mathbb{P}_t(\tilde{v}_{i,t} \geq v_i - \xi)$. The shape $J_{i,t}/c \geq \lambda_0/c \geq 1$ makes the Gamma density log-concave. Lemma 5.4 in the authors' version of Lovász and Vempala (2007) gives probability at least e^{-1} above its mean. Since $m_{i,t} \geq v_i$ on $\mathcal{C}_t$, we have $q_{i,t}(\xi) \geq e^{-1}$ there. The comparison (5) uses only conditional independence and Lemma 1(i). Applying it to these draws on a history in $\mathcal{C}_t$ gives

$$\mathbb{P}_t(\mathcal{B}_{i,t}(\xi)) \leq \frac{1 - q_{i,t}(\xi)}{q_{i,t}(\xi)} \mathbb{P}_t(i \in S_t) \leq (e-1) \mathbb{P}_t(i \in S_t). \quad (\text{EC.27})$$

Step 3 (Counting and Integrating Revenue Gaps). Let $\Delta_t := R^* - R(S_t, v) \in [0, 1]$. For $g \in (0, 1]$, choose $\xi = v_{\min}g/4$. Lemma 1(ii) implies that $\Delta_t > g$ requires a selected-coordinate error larger than ξ or a failed repair: in their absence, $\Delta_t \leq 2\xi/v_{\min} = g/2$.

Both failures can be controlled outside a common exceptional event. Let $\mathcal{E}_t(\xi) := \{\exists i \in \mathcal{N} : J_{i,t} \geq j_\xi, |\tilde{v}_{i,t} - m_{i,t}| > \xi/2\}$. By (EC.25) and a union bound, $\mathbb{P}_t(\mathcal{E}_t(\xi)) \leq 1/T^2$.

On a history in $\mathcal{C}_t$, consider draws outside $\mathcal{E}_t(\xi)$. A selected-coordinate error requires $n_{i,t} < n_\xi$ or $J_{i,t} < j_\xi$: otherwise, the center bound from Step 1 and the triangle inequality give $|\tilde{v}_{i,t} - v_i| \leq |\tilde{v}_{i,t} - m_{i,t}| + |m_{i,t} - v_i| \leq \xi/2 + \xi/2 = \xi$. A failed repair must also have $J_{i,t} < j_\xi$. Indeed, Lemma 1(i) implies $\tilde{v}_{i,t} < v_i - \xi \leq m_{i,t} - \xi$, which would put the draw in $\mathcal{E}_t(\xi)$ if $J_{i,t} \geq j_\xi$.

The selected-error contribution is therefore bounded by the probabilities of selecting products below either threshold. For each product with $J_{i,t} < j_\xi$, (EC.27) bounds the failed-repair probability by $e-1$ times its selection probability. For a history in $\mathcal{C}_t$, adding these contributions and the common exceptional probability gives

$$\mathbb{P}_t(\Delta_t > g) \leq \sum_{i \in \mathcal{N}: n_{i,t} < n_\xi} \mathbb{P}_t(i \in S_t) + e \sum_{i \in \mathcal{N}: J_{i,t} < j_\xi} \mathbb{P}_t(i \in S_t) + \frac{1}{T^2}.$$

The coefficient $e = 1 + (e - 1)$ combines the selected-error contribution and the repair comparison. We now count large gaps over time, separating histories in $\mathcal{C}_t$ from those outside it. The coverage event and both threshold conditions are measurable before the draw, so the conditional selection probabilities can be converted into expected counts by the law of iterated expectations. Summing over periods gives

$$\begin{aligned}
\mathbb{E}_v \left[\sum_{t=1}^T \mathbb{1}\{\Delta_t > g\} \right] &\stackrel{(a)}{\leq} \mathbb{E}_v \left[\sum_{t=1}^T \sum_{i \in S_t} (\mathbb{1}\{n_{i,t} < n_\xi\} + e \mathbb{1}\{J_{i,t} < j_\xi\}) \right] + \frac{1}{T} + 1 \\
&\stackrel{(b)}{\leq} N(K+1)n_\xi + eN \left(1 + \frac{(K+1)^2 j_\xi}{v_{\min}} \right) + \frac{1}{T} + 1 \\
&\stackrel{(c)}{\leq} \frac{N \log(2NT^2)}{\xi^2} \left(512(K+1) + \frac{12ec(K+1)^2}{v_{\min}} \right) + N(K+1+e) + 2 \\
&\stackrel{(d)}{\leq} \frac{CcNK^2 \log(2NT^2)}{g^2} + C.
\end{aligned}$$

In (a), removing $\mathbb{1}\{\mathcal{C}_t\}$ increases the nonnegative selection counts. The exceptional draws contribute at most $T/T^2 = 1/T$, and (EC.23) gives $\sum_{t=1}^T \mathbb{P}_v(\mathcal{C}_t^c) \leq 1$. Inequality (b) applies (EC.24) and (EC.26) to each product. Inequality (c) substitutes the thresholds, using $\lceil x \rceil \leq x + 1$ and $1/T \leq 1$. Inequality (d) uses $\xi = v_{\min}g/4$ and $K+1 \leq 2K$. The residual $N(K+1+e)$ is absorbed into the leading term because $c \geq 1$, $g \leq 1$, and $\log(2NT^2) \geq \log 2$; the constant depends only on $v_{\min}$.

Combining this estimate with the trivial count bound T gives $\mathbb{E}_v[\sum_{t=1}^T \mathbb{1}\{\Delta_t > g\}] \leq \min\{T, CcNK^2 \log(2NT^2)/g^2\} + C$. Integrate this bound as in Step 3 of the proof of Theorem 1. The change of variable $g = K\sqrt{cN \log(2NT^2)/T}u$ and $\int_0^\infty \min\{1, u^{-2}\} du = 2$ give

$$\text{Reg}(T, v) = \int_0^1 \mathbb{E}_v \left[\sum_{t=1}^T \mathbb{1}\{\Delta_t > g\} \right] dg \leq CK\sqrt{cNT \log(2NT^2)} + C.$$

■

EC.6.3. Fixed-Assortment Calibration

The regret bound is now proved. This subsection asks a separate question: what the covariance correction does to the sampling law itself. It derives the three leading covariance expressions quoted in Section 6.2 of the main text, for Algorithm 3 as it stands, for the corrected draw, and for the complete posterior. A single assortment offered in every period, as T grows, is the one regime where all three admit a closed form and can be set side by side. The comparison shows what the correction buys, namely a dispersion that dominates the posterior's once $c \geq 1$, and what no product-wise law can reproduce, namely the rank-one coupling $v_S v_S^\top$ among the products displayed together. None of it enters the proof of Theorem 3.

Let S be a nonempty assortment offered in every period, keep $N, S, c, \lambda_0, \alpha_0, \beta_0$ fixed, and let $T \rightarrow \infty$. Write $v_S = (v_i)_{i \in S}$ and assume $v_i \in (v_{\min}, 1)$ on S for the comparison with the complete

posterior. The raw-filter calculation also assumes convergence of its means. All covariances below condition on the observed history.

First, any finite limits $\bar{\mu}_i$ of the raw means must equal v_i . Indeed, the law of large numbers for purchases and the frozen-exposure update give $\bar{\mu}_i = p_i(S, v)(1 + \sum_{j \in S} \bar{\mu}_j)$. Summing over $i \in S$ yields $\sum_{i \in S} \bar{\mu}_i = \sum_{i \in S} v_i$, and substitution gives $\bar{\mu}_i = v_i$. Since $a_{i,T}/T \rightarrow p_i(S, v)$, the raw sampling variance is $\mu_{i,T}^2/a_{i,T} = v_i(1 + \sum_{j \in S} v_j)/T + o_{\mathbb{P}}(T^{-1})$.

The corrected law does not need this convergence assumption. The clock law gives $n_{i,t}/t \rightarrow (1 + v_i)/(1 + \sum_{j \in S} v_j)$ and $V_{i,t}/n_{i,t} \rightarrow \theta_i$ almost surely. For every fixed $\varepsilon \in (0, 1)$, the radii satisfy $h_{n_{i,t}} = O(\sqrt{\log T/T})$ uniformly over $\varepsilon T \leq t \leq T$. Interiority makes the intersections nonempty eventually, so their endpoints and the certified centers converge uniformly to v_i on these periods.

By continuity of the fitted choice probabilities, the increments $p_0(S, m_t)p_i(S, m_t)$ also converge uniformly to $p_0(S, v)p_i(S, v)$ over these periods. Thus, for each fixed $\varepsilon \in (0, 1)$,

$$\frac{1}{T} \sum_{t=\lceil \varepsilon T \rceil}^{T-1} p_0(S, m_t)p_i(S, m_t) \rightarrow (1 - \varepsilon)p_0(S, v)p_i(S, v).$$

The remaining fewer than εT increments are each at most $1/4$, so their contribution to $J_{i,T}/T$ lies between zero and $\varepsilon/4$; the initial term λ_0/T vanishes. Taking $T \rightarrow \infty$ at fixed ε , and then $\varepsilon \downarrow 0$, gives $J_{i,T}/T \rightarrow p_0(S, v)p_i(S, v)$. Consequently,

$$\frac{cm_{i,T}^2}{J_{i,T}} = \frac{cv_i(1 + \sum_{j \in S} v_j)^2}{T} + o_{\mathbb{P}}(T^{-1}).$$

Independence makes both sampling covariance matrices diagonal. Their variance ratio tends to $c(1 + \sum_{j \in S} v_j)$ whenever the raw means converge.

For the complete posterior, interiority places v_S inside the parameter box, so the leading covariance follows from an expansion of the posterior moments in the fixed-dimensional multinomial model. Work in the coordinates $u_i = \log v_i$, $i \in S$. Three conditions support the expansion. First, the negative log-likelihood Hessian per observation is bounded below by $v_{\min}(1 + |S|)^{-2}I$ on the box, so the posterior has uniform curvature there. Second, the likelihood maximizer on S converges to v_S , and the transformed Gamma prior is smooth and positive in a neighborhood of that limit. Third, the same curvature bound controls the posterior mass outside that neighborhood. Under these conditions, Kass et al. (1990, Theorem 4, equation (2.6)), applied to the functions e^{u_i} and $e^{u_i+u_j}$, expands the first two posterior moments, and subtracting the products of the posterior means leaves the inverse Fisher information at v_S . In attraction coordinates that information is $\mathcal{I}_S(v_S) = p_0(S, v) \text{diag}(v_S)^{-1} - p_0(S, v)^2 \mathbf{1}\mathbf{1}^\top$. Inverting this matrix yields

$$\text{Cov}_{\pi_T}(v_S) = \frac{\mathcal{I}_S(v_S)^{-1}}{T} + o_{\mathbb{P}}(T^{-1}) = \frac{1 + \sum_{i \in S} v_i}{T} (\text{diag}(v_S) + v_S v_S^\top) + o_{\mathbb{P}}(T^{-1}).$$

Using only the $T - 1$ observations available before period T changes none of these first-order terms.

Two comparisons follow. First, Cauchy–Schwarz gives $v_S v_S^\top \preceq (\sum_{i \in S} v_i) \text{diag}(v_S)$, so for $c \geq 1$ the corrected covariance dominates that of the complete posterior in the Loewner order. Second, the leading variance of the aggregate attraction $\sum_{i \in S} v_i$ is $\mathbf{1}^\top \text{Cov} \mathbf{1}$, which is the sum of the diagonal entries for the diagonal corrected law and the sum of all entries for the complete posterior. At $c = 1$ both equal $T^{-1}(1 + \sum_{i \in S} v_i)^2 \sum_{i \in S} v_i$.