# A Central Limit Theorem, Loss Aversion and Multi-Armed Bandits

Zengjing Chen [a] Larry G. Epstein [b,*] Guodong Zhang [c]

March 16, 2023

**Abstract**

This paper studies a multi-armed bandit problem where the decision-maker is loss averse, in particular she is risk averse in the domain of gains and risk loving in the domain of losses. The focus is on large horizons. Consequences of loss aversion for asymptotic (large horizon) properties are derived in a number of analytical results. The analysis is based on a new central limit theorem for a set of measures under which conditional variances can vary in a largely unstructured history-dependent way subject only to the restriction that they lie in a fixed interval. D81, D83, D91

Keywords: multi-armed bandit, loss aversion, sequential sampling, large-horizon approximations, central limit theorem, oscillating Brownian motion

*Corresponding author.

[a] Zhongtai Securities Inst. for Financial Studies, Shandong U., China; zjchen@sdu.edu.cn.

[b] Dept. of Economics, McGill U., Canada; larry.epstein@mcgill.ca.

[c] School of Mathematics, Shandong U., China; zhang_gd@mail.sdu.edu.cn.

# 1 Introduction

We study the following (multi-armed bandit) sequential choice problem.[1] There are finitely many arms (or actions), each yielding a random payoff. Probability distributions have a common mean but differ otherwise and may not be known to the decision-maker (DM). At each stage $i = 1, 2, ..., n$, DM chooses one arm, knowing the realized outcomes from previous choices. Ex ante she chooses a strategy to maximize expected utility, where the utility index is a function of the (suitably weighted) average payoff. Because we are interested in varying horizons, it is convenient to define a strategy for an infinite horizon, and then to use its truncation for any given finite horizon. Refer to a strategy as *asymptotically optimal* if the expected utility it implies in the limit as horizon $n \to \infty$ is at least as large as that implied by any other strategy. We study large-horizon approximations to the value (indirect utility) of the bandit problem and corresponding asymptotically optimal strategies.

A second novelty in our model is the assumption that DM is loss averse (global risk aversion is a limiting special case). Loss aversion was introduced via cumulative prospect theory by Tversky and Kahneman (1992), and has since been well-established empirically and widely applied in economics and finance (see for example, Kahneman and Tversky 2000, Kobberling and Wakker 2005, Barberis 2013, and the references therein). Its essential elements are (i) a reference point; (ii) utility depends only on gains and losses relative to that reference point rather than on the total payoff (or total wealth); (iii) risk aversion (concavity) for gains and risk loving (convexity) for losses; and (iv) greater sensitivity to losses than to gains. Our interest in this paper is the effect of loss aversion in the sequential context defined by a bandit problem. To our knowledge, this is the first study of loss aversion in bandit problems.[2]

We have two related reasons for studying asymptotics. First, it promotes

[1] Overviews and textbook treatments of the bandit model can be found in Berry and Fristadt (1985) and Slivkins (2019), for example. The first application to economics is Rothschild (1974). See Bergemann and Välimäki (2008) for references to a range of other economic applications.

[2] Xu and Zhou (2013) and Ebert and Strack (2015) study optimal stopping problems assuming prospect theory. Their focus is on the probability-weighting aspect of prospect theory and loss aversion plays no role in their analyses. Two studies of loss aversion in a sequential context are Easley and Yang (2015) and Shi et al. (2015). The former uses numerical analysis to study the wealth and price effects of loss aversion in the equilibrium of a dynamic heterogeneous-agent economy. The latter models dynamic portfolio choice with loss aversion, where the reference point varies endogenously in response to prior wealth outcomes. In both cases, analysis is largely numerical and there is little overlap with the bandit literature in general, and with our model, in particular. Guasoni et al. (2020) study shortfall aversion, which shares the spirit of loss aversion but which is more directly relevant to preference over deterministic consumption streams rather than over lotteries.

tractability and the derivation of analytical results. Though the literature on bandit problems is enormous, theoretical analysis of Bayesian models is, to the best of our knowledge, restricted to the special case of risk neutrality (see section 2.1.3 for elaboration and a qualification).[3] Besides its obvious limitations, risk neutrality also imposes the invariance of risk attitude as some outcomes are realized, and this invariance is key to well-known sequential properties of optimal strategies derived in the literature.[4] In contrast, endogenously varying risk attitude is inherent in loss aversion. Moreover, in our setting where means are known and common to all arms, risk neutrality would trivialize the problem.

Our second reason for studying asymptotics is that tractability is plausibly a concern not only for the modeler but also for the decision-maker within the model. We view her as struggling to comprehend an extremely complicated finite-horizon optimization problem, and adopting instead the simplifying assumption of an infinite horizon. She does so with the recognition that an asymptotically optimal strategy is approximately optimal if her horizon is sufficiently long.

Two settings that fit our model well are:
(i) A gambler chooses sequentially which of several given slot machines to play.
(ii) Each visitor to a news website decides whether to click depending on the news header presented to her. The website (DM) chooses the header (arm) with clicks being the payoffs. Visitors are drawn independently from a fixed (possibly unknown) distribution.
In both cases, outcomes are realized very soon after an arm is chosen and plausibly a large number of trials occur in a short period of time (arguing against discounting).

Here is an informal outline of some of our analytical results, which obtain as stated in the infinite-horizon limit and approximately for sufficiently large finite horizons.

1. Maximum ex ante expected utility depends on the distributions describing each arm *only through their variances*. Moreover, it depends only on the (asymptotically) largest and smallest variance. Arms with intermediate variances can be ignored.

2. Depending on the reference point, it is possible to achieve a level of ex ante expected utility that is equal to, or strictly greater than, the level when the payoff to each arm is riskless. In that sense, *risk may be desirable* in the

[3] Two studies of bandit problems that explicitly address risk are Sani et al. (2013) and Huo and Fu (2017). They assume regret minimization rather than expected utility maximization, and focus on computational algorithms rather than on qualitative theoretical results.

[4] For example, in an infinite-horizon setting where means can differ, and with one unknown arm and one arm whose distribution is known, then once the known arm is chosen it will continue to be chosen therafter (Rothschild 1974, pp. 190-191).

sequential context, even though "comparable" risks would be rejected in a one-shot choice setting.

3. Suppose that the distributions describing every arm are known. Then, in spite of the absence of learning, an asymptotically optimal strategy switches indefinitely between two fixed extreme arms (those with the smallest and largest variances) as the decision-maker moves between cumulative gains and cumulative losses. Given two arms that exhibit the two extreme variances, all other arms are redundant.

4. Suppose there are two arms and that the pair of variances is known, but there is prior uncertainty about which arm has which variance. Then it is asymptotically optimal to choose myopically at each stage, that is, as though there are no subsequent choices to be made.

5. The above results do not rely on assumptions about the nature of risk aversion in the domain of gains or about the nature of risk loving in the domain of losses. They depend only on preference over "mixed" lotteries.

Finally, we turn to the proofs of these and other results about bandits and loss aversion. It is not surprising that asymptotic results may be approached via limit theorems. However, classic limit results do not apply, and the key to our proofs is a new central limit theorem (CLT). The martingale version of the central limit theorem considers a sequence $(X_i)$ of random variables having zero conditional mean and constant conditional variance $\sigma^2$, and shows that (under suitable additional conditions) the distribution of $\Sigma_{i=1}^{n} X_i/\sqrt{n}$ converges to the normal $\mathbb{N}(0, \sigma^2)$ as $n \to \infty$. (The classic result for identically and independently distributed random variables is an immediate special case). This paper establishes a CLT under the relaxed assumption on variance according to which conditional variances can vary in a largely unstructured history-dependent way subject only to the restriction that they lie in a fixed interval $[\underline{\sigma}^2, \overline{\sigma}^2]$, in which case limits take a novel and tractable form. This CLT is the main technical contribution of the paper. One well-known motivation for generalizing from a single probability distribution (hence single variance) to a set of probability distributions (hence set of variances) is robustness to model uncertainty or ambiguity. However, model uncertainty plays no role in our bandit problem - DM is a Bayesian agent, perfectly confident in her understanding of the environment - thus highlighting the usefulness of sets of measures even for Bayesian models.

We proceed as follows. The bandit model and the results outlined above are described in detail in the next section. Proofs for these results must await the CLT which is presented next in section 3.2. Proofs of the CLT and most related

results are presented in Appendix A and proofs for the bandit application are in Appendix B.

# 2 Multi-Armed Bandits

## 2.1 Beliefs, utility and optimization

Let $\mathcal{A}$ be a finite set of arms (or actions). The outcome of any action lies in the finite set $\overline{\Omega} \subset \mathbb{R}$. Thus outcome sequences lie in $\Omega = \Pi_{i=1}^{\infty}\Omega_i$, where $\Omega_i = \overline{\Omega}$ for each $i$. The timing is as follows: At each $i \geq 1$, the history $\omega^{(i-1)} = (\omega_1, ..., \omega_{i-1})$ is known, $(\omega^{(0)} = \varnothing)$, an action $a_i \in \mathcal{A}$ is chosen, and then the resulting outcome $\omega_i$ is realized. Define $X_i(\omega) = \omega_i$, the outcome at stage $i$.

Let $\mathcal{G}_{i-1}$ be the $\sigma$-algebra representing information at stage $i$, $(\mathcal{G}_0 = \{\Omega, \varnothing\})$, and let $\mathcal{G} = \sigma(\cup_1^{\infty}\mathcal{G}_i)$ be the corresponding $\sigma$-algebra on $\Omega$.

The outcome resulting from any action is uncertain and the choice of a contingent plan, or strategy, is determined by expected utility maximization. The remaining primitives of the model - strategies, beliefs and the vNM utility index - are described next.

### 2.1.1 Strategies and beliefs

The contingent choice of action at stage $i$ depends on (conditional) beliefs about the next outcome, which generally depend on the arm being considered and also on what is learned from previous choices and their outcomes. Importantly, the inference to be drawn from the history $\omega^{(i-1)}$ of outcomes depends on which arms produced them. Thus, the choice of action at stage $i$ is expressed as

$$a_i = s_i(a^{(i-1)}, \omega^{(i-1)}), \tag{2.1}$$

where $a^{(i-1)} = (a_1, ..., a_{i-1})$ denotes the history of past actions $(a^0 = \varnothing)$. Refer to $s_i : \mathcal{A}^{i-1} \times \Pi_{j=1}^{i-1}\Omega_j \to \mathcal{A}$ as the *strategy at stage* $i$, and denote the set of all such $s_i$ by $\mathcal{S}_i$. The infinite sequence $s = (s_i)_1^{\infty}$ is called simply a *strategy*. The corresponding set of strategies is $\mathcal{S}$.[5]

Turn to beliefs. For the reasons noted above, beliefs about the next outcome depend on both the action being considered, hence on the strategy for the current stage, and on the history of past actions. Thus we model these beliefs for stage $i$ by the conditional probability measure

$$P_i^{s_i} = P_i^{s_i}\left(\cdot \mid a^{(i-1)}, \omega^{(i-1)}\right) \in \Delta(\Omega_i). \tag{2.2}$$

[5] For any given $n$, $s \in \mathcal{S}$ induces the contingent plan $(s_i)_1^n$, which is adequate if one is interested only in the $n$-horizon case. Because we will be interested in varying horizons, it is convenient to define a strategy to apply to all finite horizons.

The set of 1-step-ahead conditionals $\{P_i^{s_i}\}_{i\geq 1,\ s_i\in\mathcal{S}_i}$ is a *primitive* that represents beliefs (which may be taken to be subjective or objective).

Given a (fixed) strategy $s=(s_1,...,s_i,...)$, we describe how the primitive conditionals can be combined into a measure $P^s$ on $\Pi_{i=1}^{\infty}\Omega_i$. When conditionals do not depend on past actions, they can be pasted together to define a unique measure on the product state space (using the Ionescu-Tulcea extension theorem). This familiar procedure can be adapted to our setting. Suppose that DM is considering the strategy $s$. Then, given $s$, actions are determined by outcomes via repeated iteration of the relation

$$a_j = s_j\left(a^{(j-1)},\omega^{(j-1)}\right),\ j=1,2,... \tag{2.3}$$

In particular, for any stage $i$, one can infer past actions $a^{(i-1)}$ from past outcomes $\omega^{(i-1)}$ and the action history appearing in (2.2) can be substituted out, leaving only $\omega^{(i-1)}$ as the conditioning information. Therefore, the 1-step-ahead conditionals can be pasted together in the usual fashion to define a (unique) measure $P^s$,

$$P^s\in\Delta\left(\Pi_{i=1}^{\infty}\Omega_i,\mathcal{G}\right). \tag{2.4}$$

Moreover, its 1-step-ahead conditional $P_i^s\left(\cdot\mid\mathcal{G}_{i-1}\right)\in\Delta\left(\Omega_i\right)$ "agrees" with the primitive conditional $P_i^{s_i}$ in the sense that

$$P_i^s\left(\cdot\mid\mathcal{G}_{i-1}\right)\left(\omega^{(i-1)}\right)=P_i^{s_i}\left(\cdot\mid a^{(i-1)},\omega^{(i-1)}\right), \tag{2.5}$$

where $a^{(i-1)}$ is obtained from (2.3). (This is the counterpart of the familiar result in the standard framework that after pasting together a set of 1-step-ahead conditionals to obtain a measure $P$, the conditionals are recovered as the Bayesian updates of $P$. The mathematical proof of (2.5) is similar and elementary.)

We assume that each $P_i^s$ has *full support* on $\overline{\Omega}$, which implies that the measures $P^s$, $s\in\mathcal{S}$, are pairwise equivalent on each $\mathcal{G}_i$. We assume also that mean outcomes are common to all arms (hence also strategies) and fixed:

$$E_{P^s}\left[X_i\mid\mathcal{G}_{i-1}\right]=m=0\ \text{ for all } i\geq 1 \text{ and all } s\in\mathcal{S}, \tag{2.6}$$

where setting $m=0$ is without loss of generality.

A final assumption restricts conditional variances. Because the history relevant for conditioning includes not only past outcomes, but also past actions, and hence depends on past (stage) strategies via (2.3), we define, for any $i>1$ and $s\in\mathcal{S}$, the set $\mathcal{S}(s,i-1)$ consisting of all strategies $s'$ that agree with $s$ over the first $i-1$ stages. That is,

$$\mathcal{S}(s,i-1)=\{s'\in\mathcal{S}:s'_j=s_j,\ \text{for } 1\leq j\leq i-1\},\quad \mathcal{S}(s,0)=\mathcal{S}.$$

**Assumption-Variance**: (i) There exists $\overline{var} > 0$ such that, for any $s \in \mathcal{S}$,

$$\underset{s' \in \mathcal{S}(s,i-1)}{ess\sup} E_{P^{s'}}\left[X_i^2 | \mathcal{G}_{i-1}\right] \leq \overline{var}, \text{ for all } i \geq 1, \quad P^s\text{-a.s.}$$

(ii) There exist $0 < \underline{\sigma} < \overline{\sigma} < \infty$ such that, for every $s \in \mathcal{S}$, $P^s$-$a.s.$

$$\begin{aligned} \lim_{i \to \infty} \underset{s' \in \mathcal{S}(s,i-1)}{ess\sup} E_{P^{s'}}\left[X_i^2 | \mathcal{G}_{i-1}\right] &= \overline{\sigma}^2 \\ \text{and } \lim_{i \to \infty} \underset{s' \in \mathcal{S}(s,i-1)}{ess\inf} E_{P^{s'}}\left[X_i^2 | \mathcal{G}_{i-1}\right] &= \underline{\sigma}^2. \end{aligned} \tag{2.7}$$

In both parts, when computing $E_{P^{s'}}\left[X_i^2 | \mathcal{G}_{i-1}\right]$ as $s'$ varies over $\mathcal{S}(s, i-1)$ (rather than over $\mathcal{S}$), all conditioning information is represented by $\mathcal{G}_{i-1}$ while action history is fixed and determined by $s_1, ..., s_n$.

(i) states that conditional variances are bounded uniformly across all measures $P^s$ and all stages $i$. (ii) states, roughly speaking, that extreme (largest and smallest) conditional variances are well-defined asymptotically and common to all $s$. On a technical note, though all measures $P^s$ are equivalent on finite histories (that is, on each $\mathcal{G}_{i-1}$), they are not necessarily equivalent on $\mathcal{G}$. This leads to the almost-sure requirement for all strategies $s$ and measures $P^s$.[6]

Assumption-Variance is readily verified in the special case where there is no learning (section 2.2.2). Then conditional distributions are history-independent and time-invariant. Consequently, conditional variances are constant at their unconditional values, implying that (2.7) is satisfied with $\underline{\sigma}^2$ and $\overline{\sigma}^2$ equal, respectively, to the smallest and largest unconditional variance across the (finite) set of arms. It is satisfied also in the model of learning described in section 2.2.3, though with $\underline{\sigma}^2$ and $\overline{\sigma}^2$ that differ from unconditional variances. (In both cases, condition (i) is satisfied with $\overline{var} = \overline{\sigma}^2$.)

**Remark**: For readers who find the strategy-dependence of probability measures unorthodox we add that it is readily understood in the following terms. Consider a generic static choice problem of the form $\sup_{a \in \mathcal{A}} E_\mu\left[u\left(X^a\right)\right]$, where $X^a$ is the random variable outcome associated with action $a$ and $\mu$ is a prior over the underlying state space $\Omega$.[7] Then each $X^a$ induces a probability distribution, denoted $p^a$, over $\Omega$, and the preceding optimization problem can be written as $\sup_{a \in \mathcal{A}} E_{p^a}\left[u\left(X\right)\right]$, where $X\left(\omega\right) = \omega$. Thus the choice between actions, modeled as the choice between random variables, can be expressed alternatively as the choice between action-dependent probability distributions over outcomes (that is,

[6] Peng (2019, Ch. 6) refers to "quasisure" convergence to describe such convergence with respect to an undominated set of measures.

[7] Here $\Omega$ is an abstract state space, not necessarily related to the product state space used in the bandit model. Similarly, for $\mathcal{A}$ and for $X$ below.

lotteries). The analogue of this reformulation for our sequential choice context leads to strategy-dependent probability measures.[8]

### 2.1.2 Utility

We assume that, at each stage $i$, outcomes for each action are evaluated according to whether they produce gains or losses relative to a reference point, which we take to be their common mean (taken to be zero for convenience). Then $X_i$ gives the gain/loss at stage $i$. Since gains/losses are incurred at each stage, they must be aggregated. We posit that, for any horizon length $n$, utility depends on their $\sqrt{n}$-weighted average. Consequently, given the strategy $s$, the implied stream of gains/losses has expected utility given by

$$U_n(s) = E_{P^s}\left[\varphi\left(\Sigma_1^n X_i/\sqrt{n}\right)\right], \tag{2.8}$$

where $\varphi$ is the vNM utility index, which will be described shortly.

The $\sqrt{n}$-weighted averaging calls for some discussion. Consider a setting (such as a casino, where trials correspond to playing one or another slot machine or gambling device) where the time between trials is so small as to preclude discounting, and where the monetary payoffs at different trials are perfect substitutes. We are not aware of any axiomatic (or empirical) guidance for how a decision-maker does or should aggregate or average money streams in this context given that arbitrarily large horizons are relevant. The unweighted arithmetic average might be slightly simpler to contemplate and calculate, but significantly, it also reflects a specific and possibly inappropriate weighting to finite sets of trials. Indeed, as is familiar from discussions of the classic law of large numbers (LLN) and CLT, one might argue that scaling by $\frac{1}{n}$ implies "too little" weight for finite sets of trials, particularly when considering volatility. This perspective is reflected in the prevalence of $\sqrt{n}$-weighting in statistical decision theory, particularly where the asymptotic properties of statistical decision rules are studied (Hirano and Porter 2020).[9]

**Remark**: To be perfectly clear, the utility functions $U_n$ rank strategies for any given horizon $n$. They do not rank horizons. That is, statements such as $U_n(s) \geq U_n(s')$ are meaningful, but statements such as $U_n(s) \geq U_{n'}(s)$ are not and do not play a role below.

[8] The use of action-dependent probabilities (or moral hazard) has been recognized in the decision theory literature (Dreze 1987, Kelsey and Milne 1999, and Karni 2011, for example). These papers are concerned primarily with axiomatic foundations, extending those for subjective expected utility, while our motivation in studying the bandit problem is more applied. We differ also in our focus on sequential choice.

[9] The Remark in section 2.2.1 contrasts the implications in our bandit setting of $\frac{1}{n}$-weighting as opposed to $\frac{1}{\sqrt{n}}$-weighting.

The utility index $\varphi$ appearing in (2.8) is defined by

$$\varphi(x) = \begin{cases} \varphi_1(x-c) & x \geq c \\ -\theta^{-1}\varphi_1\left(-\theta(x-c)\right) & x < c \end{cases} \tag{2.9}$$

where we assume:

**Assumption-Utility**: $\theta = \underline{\sigma}/\overline{\sigma} < 1$, $\varphi_1(0) = 0$, $\varphi_1 \in C_b^3(\mathbb{R}_+)$, and $\varphi_1$ is (strictly) increasing and (strictly) concave for $x > 0$.[10]

Then, $\varphi$ is increasing globally, concave for $x > c$ (corresponding to gains) and convex for $x < c$ (corresponding to losses), implying risk aversion for gains and risk seeking for losses. In addition,

$$x > y \geq 0 \implies \left(c+y, \tfrac{1}{2}; c-y, \tfrac{1}{2}\right) \succ \left(c+x, \tfrac{1}{2}; c-x, \tfrac{1}{2}\right), \tag{2.10}$$

indicating greater sensitivity to the increased loss ($-x < -y$) than to the increased gain ($x > y$). In differential form, it states that

$$\varphi'(c-x) > \varphi'(c+x), \text{ for all } x > 0. \tag{2.11}$$

We take these to be the defining properties of (strict) loss aversion, following Wakker and Tversky (1993, p. 164), for example. An implication is that $-\varphi(c-x) > \varphi(c+x)$, for all $x > 0$, that is, the lottery $\left(c+x, \frac{1}{2}; c-x, \frac{1}{2}\right)$ is strictly inferior to receiving 0 for sure.

The following example will be useful in the sequel (see (2.21)) because of its tractability.

**Example 1 (Exponential).** *Let* $\varphi_1(x) = 1 - \exp(-x)$*, so that*

$$\varphi(x) = \begin{cases} 1 - \exp(-(x-c)) & x \geq c \\ \theta^{-1}(\exp(\theta(x-c)) - 1) & x < c \end{cases} \tag{2.12}$$

*where* $c \in \mathbb{R}$ *and* $\theta = \underline{\sigma}/\overline{\sigma}$.

Because of its origins in prospect theory, loss aversion is often viewed as tied to probability weighting or distortion, (which is absent in our expected utility model), and also to a kink in the utility index at the reference point (which is also absent here because $\varphi$ defined above is continuously differentiable everywhere). However, neither is necessary mathematically or conceptually for the above behavioral properties that define loss aversion.[11] Accordingly, consistent with common practice,

[10] $C_b^3(\mathbb{R}_+)$ is the set of functions on the non-negative real line with continuous and bounded third order derivatives.

[11] Kobberling and Wakker (2005) argue explicitly for a conceptual separation between loss aversion and probability weighting. They write (p. 124): "We have introduced utility, probability weighting and loss aversion as logically independent factors of risk attitude ... their (in)dependence empirically is more intricate."

we exclude probability distortions, hence Allais-type behavior, in order to isolate the effects of loss aversion on sequential decision-making. As for a kink, it has limited empirical content; for example, a finite set of pairwise rankings of lotteries, as is common in experimental investigations of loss aversion, cannot refute differentiability. Moreover, the theoretical connection of a kink to loss aversion is very much dependent on the choice of functional form. For example, suppose that, instead of (2.9), one posits that

$$\varphi(x) = \begin{cases} \varphi_1(x-c) & x \geq c \\ -\lambda\varphi_1(-(x-c)) & x < c \end{cases} \tag{2.13}$$

where $\lambda \geq 1$. Then (2.10) is satisfied if and only if $\lambda > 1$, which renders $\varphi$ nondifferentiable at $c$. Thus a kink is necessary for loss aversion given (2.13), but not given (2.9).

We add some interpretation of the functional form (2.9). Take $c = 0$ for simplicity. Then, as observed above, loss aversion implies

$$\left(x, \tfrac{1}{2}; -x, \tfrac{1}{2}\right) \prec 0.$$

How might one measure the degree of loss aversion expressed thereby? One possibility is to use the reduction in the loss needed to imply indifference, but then the new lottery would have nonzero mean which would obfuscate the determination of "greater sensitivity to losses". Similarly if one were to increase the odds of a gain with prizes unchanged. Thus we adjust both so as to keep the zero mean. Specifically, we look for $\lambda > 1$ such that

$$(x, \lambda p; -\lambda x, p) \sim 0 \text{ for all } x > 0 \text{ and } 0 < p < 1. \tag{2.14}$$

(For probabilities to sum to 1, one needs $p(1+\lambda) = 1$, but that can be safely ignored for present purposes given expected utility theory.) The above condition states that when both the odds of a gain and the size of the loss are increased by the factor $\lambda$, then (the zero mean condition is satisfied and) indifference with 0 is restored. In contrast, when $\lambda = 1$, then the strictly inferior $\frac{1}{2}/\frac{1}{2}$ lottery is obtained. This suggests using $\lambda - 1$ to measure loss aversion. Such a measure is well-defined for our model, using (2.9), since (2.14) is satisfied (uniquely) with $\lambda = \theta^{-1}$. Thus $\theta^{-1} - 1$ gives a measure of loss aversion that is *behavioral* (defined by the preference condition (2.14)), and *global* (the same $\lambda$ works for all $x$ and $p$ as indicated).[12] Alternatively, in our model $(\alpha x, p; -x, \alpha p) \sim 0$ is satisfied (uniquely)

[12] In fact, existence of $\lambda$ satisfying (2.14) is *equivalent* to our specification with $\lambda = \theta^{-1}$. More generally, one might weaken (2.14) by allowing $\lambda$ to depend on $x$ and/or $p$. From that perspective, our model yields a constant measure of loss aversion, perhaps suggesting a partial analogue to CARA utility functions.

by $\alpha = \theta < 1$, suggesting $1 - \theta$ as a measure of loss aversion. In either case, the parameter $\theta$ admits a simple behavioral interpretation.

The results below, and the CLT underlying them, are limited to the case $\theta = \underline{\sigma}/\overline{\sigma}$. However, they are robust to the specification of $\varphi_1$, which is unrestricted except for nonparametric monotonicity and concavity assumptions and technical (smoothness and boundedness) conditions. In particular, what follows makes no assumption about the nature of either risk aversion in the domain of gains or of risk loving in the domain of losses. The only relevant restriction, imposed by (2.9) and expected utility theory, is on preference over "mixed" lotteries.

### 2.1.3 Optimization

The preceding leads finally to the optimization problem (for each $n$)

$$V_n \equiv \sup_{s \in \mathcal{S}} E_{P^s} \left[ \varphi \left( \Sigma_1^n X_i / \sqrt{n} \right) \right]. \tag{2.15}$$

(Recall that $X_i(\omega) = \omega_i$ gives the outcome at stage $i$.)

The finite horizon problem is not tractable (for us). For reasons of tractability, Bayesian models in the literature typically take $\varphi$ to be linear. In that case, given the fixed finite horizon $n$, the $\frac{1}{\sqrt{n}}$ factor is irrelevant and the objective is to maximize the expected value of the sum $\Sigma_1^n X_i$. If outcomes are monetary prizes that are perfect substitutes, which is the way we think of our model, then a linear $\varphi$ implies risk neutrality as remarked in the introduction. An alternative is that outcomes are measured in utils, as in the common expected-additive-utility model of preference over risky consumption streams. Then the underlying prizes (consumption levels, for example) at different stages are not perfect substitutes, and also the ranking of the risky consumption at stage $i$ is independent of the risks involved at other stages (implying indifference to correlation in consumption risks). In applications where these features are appropriate, indifference to risk in consumption (or other underlying prizes) is not implied by a linear $\varphi$. However, for the settings we have in mind, *tractability comes at the cost of assuming risk neutrality.*

Consider briefly a common approach to solving bandit problems analytically which is to establish the optimality of index-based strategies, most commonly using the Gittins index (Gittins and Jones 1974). When arms can be valued separately, then at each stage and history an index summarizes each arm and comparison of these indices determines which arm to pull. This approach does not work in our model because arms cannot be delinked for at least two reasons: (i) outcomes from one arm may be informative about the distribution describing other arms because of common unknown parameters (see section 2.2.3); (ii) because of loss aversion risk attitude depends on the sign of the sum of past payoffs from *all* arms.

Our approach to analysing (2.15) for the loss averse utility index (2.9) is to study large-horizon approximations to the value (indirect utility) of the bandit problem and corresponding approximately optimal strategies. More precisely, define, conditional on showing below that the following limit exists,

$$V \equiv \lim_{n\to\infty} V_n. \tag{2.16}$$

Below we derive results for $V$, which therefore imply approximate results for $V_n$ when $n$ is sufficiently large. Secondly, say that the strategy $s^*$ is *asymptotically optimal* if

$$\lim_{n\to\infty} U_n(s^*) = \lim_{n\to\infty} V_n; \tag{2.17}$$

or, equivalently, if, for every $\epsilon > 0$, there exists $n^*$ such that

$$\mid U_n(s^*) - V_n \mid < \epsilon \text{ if } n > n^*.$$

Thus asymptotic optimality of $s^*$ is a more concise way to say that "$s^*$ is approximately optimal for problems with sufficiently long horizon."[13]

## 2.2 Results

In all our results for the bandit model, the assumptions specified above are adopted: conditional beliefs satisfy full support, (2.6) and Assumption-Variance, and the utility index $\varphi$ is given by (2.9) and satisfies Assumption-Utility. Though the latter requires $\overline{\sigma} > \underline{\sigma}$, all the results that follow are trivially valid, by the classic martingale CLT, also when $\overline{\sigma} = \underline{\sigma}$. Then all arms have a common variance and are equivalent in the large horizon limit, making the (asymptotic) choice between arms trivial. It simplifies discussions below to exclude that case.

### 2.2.1 Value

Our first result concerns the limiting value $V$. We emphasize the surprising (to us) degree to which this result is *robust* to specifications of $\varphi_1$ and the primitives $\{P_i^{s_i}\}_{i\geq 1,\ s_i\in\mathcal{S}_i}$, and therefore also to assumptions about the nature of learning .

**Theorem 2.1.** *(i) Let $V_n$ be the value of the n-horizon problem (2.15). Then $\lim_{n\longrightarrow\infty} V_n$ exists. Moreover,*

$$V = \lim_{n\longrightarrow\infty} V_n = \int_{-\infty}^{\infty} \varphi(y)\, q(y)\, dy, \tag{2.18}$$

[13] An implication is that, for any $s$, $\lim_{n\to\infty} U_n(s) \leq \lim_{n\to\infty} U_n(s^*)$. This follows from (2.17) and $U_n(s) \leq V_n$ for all $n$.

*where $q$ is the pdf in (B.1)-(B.2), which for $c = 0$ has the simple form*

$$q(y) = \begin{cases} q^*(y; \underline{\sigma}) \left[\frac{2\overline{\sigma}}{\underline{\sigma}+\overline{\sigma}}\right] & y \geq 0 \\ \\ q^*(y; \overline{\sigma}) \left[\frac{2\underline{\sigma}}{\underline{\sigma}+\overline{\sigma}}\right] & y < 0 \end{cases} \tag{2.19}$$

*Here $q^*(y;\sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-(y/\sigma)^2/2)$ is the pdf for $\mathbb{N}(0, \sigma^2)$.*
*(ii) Let primitive beliefs be modified to $\left\{\widehat{P_i^{s_i}}\right\}_{i\geq 1,\ s_i \in \mathcal{S}_i}$, another set satisfying our assumptions, including counterparts of (2.6) and (2.7), and where the latter is satisfied by the identical variance extremes $\underline{\sigma}^2$ and $\overline{\sigma}^2$. Then $\widehat{V} = V$.*
*(iii) The limiting value $V$ satisfies*

$$V = \begin{cases} = & \varphi(0) & c = 0 \\ > & \varphi(0) & c > 0 \\ < & \varphi(0) & c < 0 \end{cases} \tag{2.20}$$

(i) not only proves that the large-horizon limit $V$ is well-defined, but also gives an explicit description of $V$. Moreover, for some functions $\varphi$ the integral in (2.18) can be expressed in closed form yielding a closed form expression for $V$ for each $c$. For example, if $\varphi$ is taken to be the exponential example (2.12), then, using the density in (B.1)-(B.2),

$$V = \begin{cases} \Phi(-\frac{c}{\underline{\sigma}}) - \Phi(\frac{c}{\underline{\sigma}}) + e^{\frac{\sigma^2}{2}} \left(e^{-c}\Phi(-\underline{\sigma} + \frac{c}{\underline{\sigma}}) - e^{c}\Phi(-\underline{\sigma} - \frac{c}{\underline{\sigma}})\right) & c \leq 0 \\ \frac{\overline{\sigma}}{\underline{\sigma}} \left[\Phi(-\frac{c}{\overline{\sigma}}) - \Phi(\frac{c}{\overline{\sigma}}) + e^{\frac{\sigma^2}{2}} \left(e^{-\frac{\sigma c}{\overline{\sigma}}}\Phi(-\underline{\sigma} + \frac{c}{\overline{\sigma}}) - e^{\frac{\sigma c}{\overline{\sigma}}}\Phi(-\underline{\sigma} - \frac{c}{\overline{\sigma}})\right)\right] & c > 0, \end{cases} \tag{2.21}$$

where $\Phi$ is the standard normal cdf.

The density $q$ in (2.19) yields a zero mean and variance equal to $\underline{\sigma}\overline{\sigma}$, the geometric average of the two extreme variances. Incorporation of the low (high) variance normal density for positive (negative) arguments reflects risk aversion and loving on the two subdomains respectively. Evidently, $q$ reduces to the normal density if $\underline{\sigma} = \overline{\sigma}$, for example, there is a single arm. Then (2.18) is an immediate implication of the classic CLT. In the same way, (i) follows directly from the new CLT in section 3.2. Moreover, (i) is the main content of the theorem - the other parts follow immediately from it. (ii) follows by inspection of the density and (iii) follows from a simple calculation (see details in Appendix B).

Part (ii) supports our hypothesis that the long-horizon heuristic reduces the cognitive burden of the decision-maker. She need only know the variances of arms, and even then, only for arms that have extreme variances. Here is some rough intuition: Let the horizon be $n$ and consider the choice of arm at the last

stage given past realizations $x_i$ of $X_i$, $i < n$. It can be thought of as maximizing $E_{P_n^s}\left[\varphi\left(((\Sigma_1^{n-1}x_i + X_n)/\sqrt{n}\right)\right]$ by choice of $s_n$ ($P_n^s$ is the 1-step-ahead conditional in (2.5)). The incremental payoff $X_n/\sqrt{n}$ is small if $n$ is large. Thus a second-order Taylor series expansion in $X_n$ can be used to approximate the objective function, implying that the latter can be approximated (for each $s$) by a linear function of both the mean (equal to zero by (2.6)) and the (conditional) variance. Finally, the maximum value of a linear function of variance is achievable at an arm associated with either $\underline{\sigma}$ or $\overline{\sigma}$.

To interpret (iii), consider first the case $c = 0$. Thus, for large $n$, maximum expected utility is approximately equal to that achievable when the payoff to each action is riskless, hence identically equal to the common mean, implying zero gains and losses for sure. In other words, *risk is a matter of indifference in the limit.* The freedom to switch between arms in response to experience is critical. If one arm must be chosen ex ante for all trials, then maximum expected utility is negative, hence less than $\varphi(0) = 0$. (The classic CLT applies to each arm separately and, by loss aversion, $\varphi(-x) < -\varphi(x)$ for all $x > 0$; hence $\varphi(\cdot)$ has negative expected value under the normal $\mathbb{N}(0, \sigma^2)$ for any positive variance.) For further perspective, consider the following lottery: Toss a fair coin. If Heads, then receive a positive prize according to $\mathbb{N}(0, \underline{\sigma}^2)$ conditioned on $\mathbb{R}_+$ and if Tails receive a negative prize according to $\mathbb{N}(0, \overline{\sigma}^2)$ conditioned on $\mathbb{R}_-$. This lottery has negative expected utility using $\varphi$. It is less attractive because while the ability to choose actions sequentially affords some influence over positive versus negative outcomes, in the lottery that influence belongs to nature alone.

Finally, (iii) implies that, *in the limit $n \to \infty$, a decision-maker with a positive reference point ($c > 0$) strictly prefers the risky sequential choice problem to receiving zero gain/loss for sure.* The intuition is that zero for sure is a certain loss relative to a positive reference point, which makes it unattractive. A positive reference point $c$ also reduces the limit value $V$, because it reduces all gains and increases all losses ($\varphi(x) \overset{c}{\searrow}$ for all $x$), but to a lesser degree because of the flexibility afforded by switching actions. Similarly, a negative reference point implies the preference for the certain zero outcome. In this sense, a higher benchmark or aspiration level leads to more participation in risky endeavors.

**Remark**: Suppose that DM uses the unweighted arithmetic average and maximizes $E_{P^s}[\varphi((\Sigma_1^n X_i)/n)]$. Then a LLN would replace the CLT underlying (2.18) and would yield, by the LLN in Peng (2019, Theorem 2.4.1),

$$\lim_{n\to\infty} V_n = \varphi(0) = 0. \tag{2.22}$$

To reflect, consider the special case where there is independence across trials of a single arm and across arms. Then by the classic LLN, the expected utility of

playing any $a \in \mathcal{A}$ at every stage and history converges to 0 as $n \to \infty$. Consequently, for large $n$, DM is approximately indifferent between repeated plays of $a$ and repeated plays of any other $a'$, because their means are identical. The implication of (2.22) is that all such single-arm strategies are asymptotically optimal, from which we conclude that, (in our setting, where only variances differ), the LLN cannot serve as the basis for usefully approximating optimal strategies for finite horizon problems. Furthermore, under the LLN, (2.22) is valid not only for the loss averse functions $\varphi$ that we assume throughout, but also for all (suitably bounded and continuous) $\varphi$ satisfying $\varphi(0) = 0$. In contrast, in our model using the $\sqrt{n}$-weighted average, such asymptotic risk neutrality is satisfied only in the knife-edge case $c = 0$, and risk is even strictly desirable for $c > 0$.

### 2.2.2 Strategies and the absence of learning

We describe an asymptotically optimal strategy for the special case where there is no learning. The latter corresponds to the following restriction on the primitive conditionals $\{P_i^{s_i}\}_{i\geq 1,\ s_i \in \mathcal{S}_i}$: For all $i \geq 1$, $s_i \in \mathcal{S}_i$ and histories $\left(a^{(i-1)}, \omega^{(i-1)}\right)$,

$$P_i^{s_i}\left(\cdot \mid a^{(i-1)}, \omega^{(i-1)}\right) = P_1^{s_1} \text{ if } s_i\left(a^{(i-1)}, \omega^{(i-1)}\right) = s_1. \tag{2.23}$$

Recall that at stage 1, history is null. Hence $s_1$ is simply an action and $P_1^{s_1}$ gives (unconditional or) prior beliefs about the outcome of action $s_1$. Thus (2.23) stipulates that for each given action ($s_1$ above), subsequent *beliefs about the next outcome of that action do not change with history* (where history includes past outcomes associated with any, possibly different, action). An implication is that for each fixed arm $a$, the joint probability distribution over outcomes given repeated choice of $a$ is i.i.d. However, for other strategies $s$, the induced measure $P^s$ (recall (2.4)) need not be a product measure. (For example, if $\omega_1$ and $\omega_1'$ are distinct outcomes, and if $s$ specifies different actions at the histories $(a_1, \omega_1)$ and $(a_1, \omega_1')$, then the two conditional probability distributions for stage 2 outcomes generally differ. This reflects a difference in the choice of action at stage 2 rather than updating or learning.)

Define

$$\sigma_a^2 = E_{P_1^{s_1}}\left[X_1^2\right], \text{ if } s_1 = a \in \mathcal{A}.$$

Then (2.7) is satisfied with

$$\overline{\sigma} = \max_{a\in\mathcal{A}} \sigma_a \text{ and } \underline{\sigma} = \min_{a\in\mathcal{A}} \sigma_a.$$

For simplicity, we focus first on $c = 0$ and then indicate at the end of this subsection how to accommodate $c \neq 0$.

**Theorem 2.2.** *Let $c = 0$. Define strategy $s^*$ by $s_1^* = \overline{a}$ and, for $n > 1$,*

$$s_n^* = \begin{cases} \overline{a} & \text{if } \Sigma_1^{n-1} X_i \leq 0 \\ \underline{a} & \text{if } \Sigma_1^{n-1} X_i > 0 \end{cases} \tag{2.24}$$

*where $\sigma_{\overline{a}} = \overline{\sigma}$ and $\sigma_{\underline{a}} = \underline{\sigma}$. Then: (i) $s^*$ is asymptotically optimal.*
*(ii) For every $N > 0$,*

$$P^{s^*}\left(\cap_{n=N}^{\infty}\{\Sigma_1^n X_i \leq 0\}\right) \leq \frac{\overline{\sigma}}{\overline{\sigma} + \underline{\sigma}} < 1 \quad \text{and}$$

$$P^{s^*}\left(\cap_{n=N}^{\infty}\{\Sigma_1^n X_i > 0\}\right) \leq \frac{\overline{\sigma}}{\overline{\sigma} + \underline{\sigma}} < 1.$$

*(iii) The high variance action is chosen less frequently in the limit. In fact,*

$$\lim_{n\to\infty} \frac{P^{s^*}\left(\sigma_{s_n^*} = \overline{\sigma}\right)}{P^{s^*}\left(\sigma_{s_n^*} = \underline{\sigma}\right)} = \frac{\underline{\sigma}}{\overline{\sigma}} < 1. \tag{2.25}$$

(i) identifies an asymptotically optimal $s^*$, while (ii) states that $s^*$ exhibits switching between actions indefinitely with positive probability according to the measure $P^{s^*}$ induced by $s^*$. The latter fact indicates a difference between our model with loss aversion and many bandit models. Commonly in the bandit literature, learning (or exploration) provides the reason for switching, and eventually it is decided that one arm is superior and experimentation ceases. Here, in contrast, switching is optimal even in the absence of learning and (with positive probability) persists indefinitely. This is because loss aversion implies that the identity of the more attractive action or arm depends on whether one is in a region of cumulative gains ($\Sigma_1^n X_i > 0$) or cumulative losses ($\Sigma_1^n X_i < 0$).[14] Finally, (iii) gives explicitly the limiting relative frequencies induced by $s^*$.

Finally, we describe how the theorem can be extended to accommodate $c \neq 0$. For that purpose, instead of using a single strategy to approximate finite-horizon problems, consider a *sequence* $s^n = (s_i^n)$ *of strategies*, where, for each $n$, $s^n \in \mathcal{S}$ is thought of as a strategy used in the $n$-horizon problem (2.15). (Accordingly, components $s_i^n$ with $i > n$ are irrelevant.) The counterpart of (2.17) is

$$\lim_{n\to\infty} E_{P^{s^n}}\left[\varphi\left(\Sigma_1^n X_i/\sqrt{n}\right)\right] = \lim_{n\to\infty} V_n = V \tag{2.26}$$

Then, arguing as in the proof of Theorem 2.2, one can show that (2.26) is satisfied by $s^n$, where, for each $n \geq 1$ and $1 \leq i \leq n$,

$$s_i^n = \begin{cases} \overline{a} & \text{if } \Sigma_1^{i-1} X_j/\sqrt{n} \leq c \\ \underline{a} & \text{if } \Sigma_1^{i-1} X_j/\sqrt{n} > c. \end{cases}$$

$s_i^n$ can be defined arbitrarily if either $n = 1$ or $i > n$.

[14] A global risk averter would choose the low variance action $\underline{a}$ at every stage.

### 2.2.3 A classic two-armed bandit problem revisited

There are two arms, $a$ and $b$, hence $\mathcal{A} = \{a, b\}$. The set of possible outcomes for each arm and stage is $\overline{\Omega} = \{1, -1, 0\}$, and outcomes are governed, both ex ante and *for any history*, by the following probabilities:

$$\begin{aligned} \text{arm } a\text{: } & \Pr(1) = \Pr(-1) = p_a/2 \\ \text{arm } b\text{: } & \Pr(1) = \Pr(-1) = p_b/2. \end{aligned}$$

For each arm, outcomes follow a random walk with zero mean and with variance equal to the appropriate value of $p$. It is known that

$$\{p_a, p_b\} = \{\underline{p}, \overline{p}\}, \tag{2.27}$$

where $0 < \underline{p} < \overline{p} < 1$ are known. However, there is uncertainty about which of $\underline{p}$ and $\overline{p}$ describes arm $a$ and which describes arm $b$, that is, there is uncertainty about which arm has the higher variance. DM has prior beliefs about which arm is which, and forms Bayesian posteriors as experience accumulates. At each stage, she chooses which arm to pull taking into account what she has learned about the arms from past experience.

**Remark**: Uncertainty about "which arm is which" in a 2-arm setting is a classic version of the bandit problem (Bradt et el. 1956; Feldman 1962); indeed, the former refer to it (p. 1060) as "the Two-armed Bandit." These and subsequent papers typically assume a finite horizon and maximization of the expected value of the sum of payoffs, (in particular, means rather than variances are the focus).

Our framework accommodates the above learning process. The set of primitive conditionals $\{P_n^{s_n}\}_{n \geq 1, s_n \in \mathcal{S}_n}$ is defined as follows. DM's prior beliefs about which arm is which are completely specified by $\mu_1$, the probability she assigns initially to $p_a = \underline{p}$. Thus, prior probabilities of the outcomes from choosing arm $\alpha$, $\alpha = a, b$, are given by

$$\begin{aligned} P_1^a(1) &= \mu_1 \underline{p}/2 + (1-\mu_1)\overline{p}/2 = P_1^a(-1) \\ P_1^b(1) &= (1-\mu_1)\underline{p}/2 + \mu_1 \overline{p}/2 = P_1^b(-1), \end{aligned}$$

which can be expressed in terms of our formalism by

$$\begin{aligned} P_1^{s_1}(\omega_1) = & I_{\{s_1=a, \omega_1 \neq 0\}} \left[\mu_1 \underline{p}/2 + (1-\mu_1)\overline{p}/2\right] \\ & + I_{\{s_1=a, \omega_1=0\}} \left[\mu_1(1-\underline{p}) + (1-\mu_1)(1-\overline{p})\right] \\ & + I_{\{s_1=b, \omega_1 \neq 0\}} \left[(1-\mu_1)\underline{p}/2 + \mu_1 \overline{p}/2\right] \\ & + I_{\{s_1=b, \omega_1=0\}} \left[(1-\mu_1)(1-\underline{p}) + \mu_1(1-\overline{p})\right]. \end{aligned}$$

For later stages, DM updates her prior probability that $p_a = \underline{p}$ to the Bayesian posterior $\mu_n$, $n > 1$, defined inductively by

$$\begin{aligned} &\log\left(\frac{\mu_{n+1}/(1-\mu_{n+1})}{\mu_n/(1-\mu_n)}\right) \\ &= \left[I_a(a_n) - I_b(a_n)\right]\left((1 - I_0(\omega_n))\log\left(\frac{\underline{p}}{\overline{p}}\right) + I_0(\omega_n)\log\left(\frac{1-\underline{p}}{1-\overline{p}}\right)\right). \end{aligned} \tag{2.28}$$

Then the conditional probability $P_n^{s_n}$, for each $n > 1$ and stage strategy $s_n$, is given by

$$\begin{aligned} P_n^{s_n}\left(\omega_n | a^{(n-1)}, \omega^{(n-1)}\right) &= I_{\{s_n=a,\omega_n\neq 0\}}\left[\mu_n\underline{p}/2 + (1-\mu_n)\overline{p}/2\right] \\ &+ I_{\{s_n=a,\omega_n=0\}}\left[\mu_n(1-\underline{p}) + (1-\mu_n)(1-\overline{p})\right] \\ &+ I_{\{s_n=b,\omega_n\neq 0\}}\left[(1-\mu_n)\underline{p}/2 + \mu_n\overline{p}/2\right] \\ &+ I_{\{s_n=b,\omega_n=0\}}\left[(1-\mu_n)(1-\underline{p}) + \mu_n(1-\overline{p})\right]. \end{aligned} \tag{2.29}$$

Consider also the probability measure $P^s$, for $s \in \mathcal{S}$, constructed as in (2.4) by pasting the above conditionals. It is completely described by its restriction to finite dimensional cylinders, and thus we may view $P^s$ also as a measure on $\Pi_{i=1}^n \Omega_i$. For any $\omega^{(n)} = (\omega_1, ..., \omega_n)$, the outcomes of the first $n$ trials, and the given $s$, define the induced frequency vector $f^s\left(\omega^{(n)}\right)$,

$$f^s\left(\omega^{(n)}\right) = \left(f_a^s\left(\omega^{(n)}\right), f_b^s\left(\omega^{(n)}\right), f_{a,0}^s(\omega^{(n)}), f_{b,0}^s\left(\omega^{(n)}\right)\right), \tag{2.30}$$

where: for $\alpha \in \{a, b\}$, $f_\alpha^s\left(\omega^{(n)}\right)$ and $f_{\alpha,0}^s\left(\omega^{(n)}\right)$ give, respectively, the number of trials of arm $\alpha$ and the number of those that yield the outcome 0. Then the ex ante probability of the above outcomes are given by[15]

$$\begin{aligned} P^s(\omega_1, ..., \omega_n) &= \mu_1\left[(\underline{p}/2)^{f_a^s - f_{a,0}^s}(\overline{p}/2)^{f_b^s - f_{b,0}^s}\left(1-\underline{p}\right)^{f_{a,0}^s}(1-\overline{p})^{f_{b,0}^s}\right] \\ &+ (1-\mu_1)\left[(\overline{p}/2)^{f_a^s - f_{a,0}^s}\left(\underline{p}/2\right)^{f_b^s - f_{b,0}^s}(1-\overline{p})^{f_{a,0}^s}\left(1-\underline{p}\right)^{f_{b,0}^s}\right]. \end{aligned} \tag{2.31}$$

The two terms on the right correspond to the two possible scenarios, $p_a = \underline{p}$ or $\overline{p}$, weighted by their prior probabilities. Conditional on each scenario the expression reflects two assumptions: (i) independence between distinct trials, whether conducted with the same arm or with different arms; and (ii) all trials with a given arm are viewed as similar (or interchangeable) so that the probability of any (finite) sequence of outcomes for that arm is invariant to any reordering (accordingly, for each arm, the probability of a set of outcomes depends only on the number

[15] The proof is elementary and is omitted.

of occurrences of 0 and $\{1,-1\}$). This latter assumption of "symmetry" within each arm is known as *partial exchangeability,* a property introduced by de Finetti (1938), who also showed that it implies conditional independence as in (i), and, in fact, that it characterizes a representation such as in (2.31).[16]

The preceding satisfies Assumption-Variance. To verify (2.7), we observe first that the process of posteriors $\{\mu_n\}$ satisfies properties familiar from Bayesian learning theory.

**Lemma 2.3.** *Let $s \in \mathcal{S}$ be any strategy. Then:*
*(i) Posteriors converge to certainty, that is, for any prior $\mu_1$,*

$$\lim_{n\to\infty} \mu_n \in \{0,1\} \quad P^s\text{-a.s. for every } s \in \mathcal{S}. \tag{2.32}$$

*(ii) Suppose that, unknown to the decision-maker, the truth is that $p_a = \underline{p}$. Consequently, given any strategy $s \in \mathcal{S}$, outcomes are governed by the probability law $Q^s \in \Delta\left(\Pi_1^\infty \Omega_i, \mathcal{G}\right)$, whose 1-step-ahead conditionals are $Q_i^s$, $i \geq 1$, given by*

$$Q_i^s(1) = Q_i^s(-1) = \begin{cases} \underline{p}/2 & \text{if} \quad s_i = a \\ \overline{p}/2 & \text{if} \quad s_i = b \end{cases}$$

*Then, for every $\mu_1 > 0$,*

$$\lim_{n\to\infty} \mu_n = 1 \quad Q^s\text{-a.s.} \tag{2.33}$$

*(iii) Assumption-Variance is satisfied with $\underline{\sigma}^2 = \underline{p}$ and $\overline{\sigma}^2 = \overline{p}$.*

Viewing $\{\mu_n\}$ as representing subjective beliefs, (2.32) expresses the decision-maker's ex ante complete confidence that asymptotically she will be certain "which arm is which." In (ii), $Q^s$ is the true probability law over outcome sequences when strategy $s$ is adopted, and hence (2.33) is an expression of "Bayesian consistency". Both results are valid for any strategy, and thus reflect Bayesian updating alone and not asymptotic optimality. (iii) is implied by (i). (See details in Appendix B.)

Conclude that Theorem 2.1 applies. Moreover, with the added structure assumed herein we can also identify an asymptotically optimal strategy.[17]

[16] The stronger property of exchangeability, which is better known, assumes interchangeability also across distinct arms and thus views the two arms as being identical, which is excluded in our case because of (2.27) and $\underline{p} \neq \overline{p}$. See Link (1980) and Diaconis and Freedman (1982) for more on partial exchangeability and Kallenberg (2005) for a comprehensive treatment of probabilistic symmetries.

[17] When $\mu_1 \in \{0,1\}$, we are back in the no-learning case of the last section and Theorem 2.2 applies.

**Theorem 2.4.** *Let $c = 0$ and $\mu_1 \in [0,1]$. Let $s^*$ be the strategy given by $s_1^* = a$ and, for $n > 1$,*

$$s_n^* = \begin{cases} a \quad \textit{if} & \Sigma_1^{n-1} X_j \leq 0,\ \mu_n < \frac{1}{2}\ \textit{OR} \\ & \Sigma_1^{n-1} X_j > 0,\ \mu_n > \frac{1}{2} \\ b \quad \textit{if} & \textit{otherwise} \end{cases}$$

*Then $s^*$ is asymptotically optimal.*

According to $s^*$, arm $a$ is used at stage $n > 1$ if (and only if) there are cumulative losses and it is more likely that $a$ has higher variance ($\mu_n < \frac{1}{2}$), or there are cumulative gains and it is more likely that $a$ has lower variance ($\mu_n > \frac{1}{2}$). Intuition argues for this choice of arm at stage $n$ if there are no later trials remaining, but may seem myopic more generally. Nevertheless, $s^*$ is approximately optimal for large horizons. (For other instances where myopic strategies are optimal in bandit problems see, for example, Banks and Sundaram (1992) and the papers cited in the preceding Remark.)

# 3 A Central Limit Theorem

## 3.1 Preliminaries

The mathematical basis for our analysis of the bandit problem is a central limit theorem about sets of measures that will be provided here. To smooth the transition for the reader, we begin with a few remarks about connecting the bandit model to sets of measures.

In section 2.1.1, we introduced the primitive set of one-step-ahead conditionals $\{P_i^{s_i}\}_{i \geq 1,\ s_i \in \mathcal{S}_i}$, and then pointed out that, for each $s = (s_1, ..., s_i, ...)$, these conditionals can be pasted together to obtain a measure $P^s \in \Delta\left(\Pi_{i=1}^{\infty} \Omega_i, \mathcal{G}\right)$. Now we collect all these measures and define the set $\mathcal{P} \subset \Delta\left(\Pi_{i=1}^{\infty} \Omega_i, \mathcal{G}\right)$ by

$$\mathcal{P} = \{P^s : s \in \mathcal{S}\}. \tag{3.1}$$

Our CLT will be applied to this set. However, in order to better reveal its underlying structure and to facilitate other potential applications, (for example, to models concerned with robustness to model uncertainty), the CLT will be formulated and proven more generally.

One more observation is helpful for the transition. For the set $\mathcal{P}$ defined by (3.1), it is immediate that, for each $n$,

$$V_n = \sup_{s \in \mathcal{S}} E_{P^s}[\varphi\left(\Sigma_1^n X_i / \sqrt{n}\right)] = \sup_{Q \in \mathcal{P}} E_Q[\varphi\left(\Sigma_1^n X_i / \sqrt{n}\right)]. \tag{3.2}$$

The CLT will involve expressions such as that on the right in (3.2). However, we can think of the supremum over measures as equivalent to optimization over strategies.

To proceed, adopt the mathematical primitives $(\Pi_{i=1}^{\infty}\Omega_i, \{\mathcal{G}_i\}_{i=1}^{\infty})$ and $\mathcal{G}$, though with possibly different interpretations.[18] For each $i \geq 1$, $X_i : \Pi_1^{\infty}\Omega_j \longrightarrow \mathbb{R}$ is $\mathcal{G}_i$-measurable. Another primitive is a set $\mathcal{P} \subset \Delta(\Pi_{i=1}^{\infty}\Omega_i, \mathcal{G})$, not to be confused with the set in (3.1). $\mathcal{H}$ denotes the set of all random variables $X$ on $(\Pi_{i=1}^{\infty}\Omega_i, \mathcal{G})$ satisfying $\sup_{Q\in\mathcal{P}} E_Q[|X|] < \infty$.

We assume throughout that $(X_i)$ and $\mathcal{P}$ satisfy: all measures in $\mathcal{P}$ are equivalent on each $\mathcal{G}_i$, and counterparts of (2.6) and Assumption-Variance. That is, assume

$$E_Q[X_i|\mathcal{G}_{i-1}] = 0 \text{ for all } Q \in \mathcal{P} \text{ and all } i \geq 1, \tag{3.3}$$

and, defining

$$\mathcal{P}(Q,i) = \{Q' \in \mathcal{P} : \ Q'_{|\mathcal{G}_i} = Q_{|\mathcal{G}_i}\}, \ \ Q \in \mathcal{P},$$

assume there exist $\overline{var} > 0$ and $0 < \underline{\sigma} \leq \overline{\sigma} < \infty$ such that, for all $i \geq 1$ and $Q \in \mathcal{P}$, $Q$-$a.s.$

$$\begin{aligned} \overline{var} &\geq \operatorname*{ess\,sup}_{Q'\in\mathcal{P}(Q,i-1)} E_Q\left[X_i^2|\mathcal{G}_{i-1}\right] \\ \overline{\sigma}^2 &= \lim_{i\to\infty} \operatorname*{ess,sup}_{Q'\in\mathcal{P}(Q,i-1)} E_{Q'}\left[X_i^2|\mathcal{G}_{i-1}\right] \\ \text{and } \ \underline{\sigma}^2 &= \lim_{i\to\infty} \operatorname*{ess,sup}_{Q'\in\mathcal{P}(Q,i-1)} E_{Q'}\left[X_i^2|\mathcal{G}_{i-1}\right]. \end{aligned} \tag{3.4}$$

Assume also that $(X_i)$ satisfies the *Lindeberg condition*,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \sup_{Q\in\mathcal{P}} E_Q\left[|X_i|^2 I_{\{|X_i|>\sqrt{n}\epsilon\}}\right] = 0, \quad \forall \epsilon > 0. \tag{3.5}$$

The final assumption, called *quadratic-consistency*, is the following: For any continuous bounded functions $f$ and $h$, and $i \geq 1$,

$$\begin{aligned} &\sup_{Q\in\mathcal{P}} E_Q\left[f\left(\Sigma_{j=1}^{i-1}X_j\right) + h\left(\Sigma_{j=1}^{i-1}X_j\right)X_i^2\right] \\ &= \sup_{Q\in\mathcal{P}} E_Q\left[f\left(\Sigma_{j=1}^{i-1}X_j\right) + \operatorname*{ess\,sup}_{Q'\in\mathcal{P}(Q,i-1)} E_{Q'}\left[h\left(\Sigma_{j=1}^{i-1}X_j\right)X_i^2|\mathcal{G}_{i-1}\right]\right]. \end{aligned} \tag{3.6}$$

When $\mathcal{P}$ is a singleton, (3.6) is implied by the law of iterated expectations, which applies to much more general functions of $(X_1, ..., X_{i-1}, X_i)$. Quadratic-consistency

[18] In fact, we do not need the previous assumptions that $\Omega_i$ is identical for all $i$ and finite. Here the $\Omega_i$s are arbitrary.

is a counterpart for the upper expectation given nonsingleton sets, though only with the functional form restrictions, (including quadratic in $X_i$), imposed in (3.6).

Importantly, quadratic-consistency is satisfied in the bandit model.[19]

**Lemma 3.1.** *The outcomes $(X_i)$ and the set $\mathcal{P}$ defined in (3.1) satisfy quadratic-consistency.*

**Proof:** For any continuous bounded functions $f$ and $h$, and $i \geq 1$,

$$\begin{aligned}
&\sup_{Q\in\mathcal{P}} E_Q\left[f\left(\Sigma_{j=1}^{i-1}X_i\right)+h\left(\Sigma_{j=1}^{i-1}X_j\right)X_i^2\right]\\
=&\sup_{s\in\mathcal{S}} E_{P^s}\left[f\left(\Sigma_{j=1}^{i-1}X_j\right)+h\left(\Sigma_{j=1}^{i-1}X_j\right)X_i^2\right]\\
=&\sup_{s\in\mathcal{S}} E_{P^s}\left[f\left(\Sigma_{j=1}^{i-1}X_j\right)+E_{P_i^s}\left[h\left(\Sigma_{j=1}^{i-1}X_j\right)X_i^2|\mathcal{G}_{i-1}\right]\right]\\
=&\sup_{s\in\mathcal{S}} E_{P^s}\left[f\left(\Sigma_{j=1}^{i-1}X_j\right)+E_{P_i^{s_i}}\left[h\left(\Sigma_{j=1}^{i-1}X_j\right)X_i^2|a_s^{(i-1)},\omega^{(i-1)}\right]\right],
\end{aligned}$$

where $a_s^{(i-1)}$ is given by (2.3) via strategy $s$.

For any $s_i' \in \mathcal{S}_i$ and $s \in \mathcal{S}$, there exist a strategy $\hat{s} \in \mathcal{S}(s, i-1)$ such that $\hat{s}_j = s_j$ for $j < i$ and $\hat{s}_i = s_i'$. Therefore,

$$\begin{aligned}
&\sup_{s\in\mathcal{S}} E_{P^s}\left[f\left(\Sigma_{j=1}^{i-1}X_j\right)+E_{P_i^{s_i}}\left[h\left(\Sigma_{j=1}^{i-1}X_j\right)X_i^2|a_s^{(i-1)},\omega^{(i-1)}\right]\right]\\
=&\sup_{s\in\mathcal{S}}\sup_{s_i'\in\mathcal{S}_i} E_{P^s}\left[f\left(\Sigma_{j=1}^{i-1}X_j\right)+E_{P_i^{s_i'}}\left[h\left(\Sigma_{j=1}^{i-1}X_j\right)X_i^2|a_s^{(i-1)},\omega^{(i-1)}\right]\right]\\
\geq&\sup_{s\in\mathcal{S}} E_{P^s}\left[f\left(\Sigma_{j=1}^{i-1}X_j\right)+E_{P_i^{s_i^*}}\left[h\left(\Sigma_{j=1}^{i-1}X_j\right)X_i^2|a_s^{(i-1)},\omega^{(i-1)}\right]\right]\\
=&\sup_{s\in\mathcal{S}} E_{P^s}\left[f\left(\Sigma_{j=1}^{i-1}X_j\right)+\underset{s'\in\mathcal{S}(s,i-1)}{ess\sup} E_{P^{s'}}[X_i^2|\mathcal{G}_{i-1}]\left[h(\Sigma_{j=1}^{i-1}X_j)\right]^+\right.\\
&\qquad\left.-\underset{s'\in\mathcal{S}(s,i-1)}{ess\inf} E_{P^{s'}}[X_i^2|\mathcal{G}_{i-1}]\left[h(\Sigma_{j=1}^{i-1}X_j)\right]^-\right]\\
=&\sup_{Q\in\mathcal{P}} E_Q\left[f\left(\Sigma_{j=1}^{i-1}X_j\right)+\underset{Q'\in\mathcal{P}(Q,i-1)}{ess\sup} E_{Q'}\left[h\left(\Sigma_{j=1}^{i-1}X_j\right)X_i^2|\mathcal{G}_{i-1}\right]\right],
\end{aligned}$$

[19] An earlier version of the paper relied on a CLT that assumed the stronger property - rectangularity of $\mathcal{P}$ (Epstein and Schneider 2003). We are grateful to Luciano Pomatto for alerting us to the fact that rectangularity is not satisfied by the bandit model when there is learning, which realization led us to recognize quadratic-consistency as the property that "works" for both the bandit model and the CLT.

where the strategy $s_i^*$ is defined by

$$s_i^*(a_s^{(i-1)}, \omega^{(i-1)}) = \begin{cases} \arg\max\limits_{a \in \mathcal{A}} E_{P_i^a}\left[X_i^2 | a_s^{(i-1)}, \omega^{(i-1)}\right] & \text{if} \quad h(\Sigma_{j=1}^{i-1} X_j) \geq 0 \\ \arg\min\limits_{a \in \mathcal{A}} E_{P_i^a}\left[X_i^2 | a_s^{(i-1)}, \omega^{(i-1)}\right] & \text{if} \quad h(\Sigma_{j=1}^{i-1} X_j) < 0 \end{cases}$$

On the other hand,

$$\begin{aligned} &\sup_{Q \in \mathcal{P}} E_Q\left[f\left(\Sigma_{j=1}^{i-1} X_j\right) + h\left(\Sigma_{j=1}^{i-1} X_j\right) X_i^2\right] \\ \leq &\sup_{Q \in \mathcal{P}} E_Q\left[f\left(\Sigma_{j=1}^{i-1} X_j\right) + \underset{Q' \in \mathcal{P}(Q, i-1)}{ess\sup} E_{Q'}\left[h\left(\Sigma_{j=1}^{i-1} X_j\right) X_i^2 | \mathcal{G}_{i-1}\right]\right], \end{aligned}$$

proving quadratic-consistency. ■

## 3.2 The theorem

We extend (a version of) the classic martingale CLT to admit a set of variances while maintaining the assumption of a fixed zero mean. Throughout $(B_t)$ denotes a standard Brownian motion under a probability space $(\Omega^*, \mathcal{F}^*, P^*)$ and $(\mathcal{F}_t)_{t \geq 0}$ is the natural filtration generated by $(B_t)$.

In the classic case, the limiting distribution is normal, which is the distribution of $B_1$. In the more general case, the corresponding (upper) limit is not given by the normal distribution, but is described instead by the time 1 value of an *oscillating Brownian motion* (Keilson and Wellner 1978; Lejay and Pigato 2018), defined as follows: Given $\overline{\sigma} \geq \underline{\sigma} > 0$ and threshold $c \in \mathbb{R}$, let $(W_t^c)$ denote the unique strong solution, (which exists by Le Gall (1984)), of the stochastic differential equation (SDE)

$$Y_t = \int_0^t \sigma\left(Y_s\right) dB_s, \quad t \geq 0, \tag{3.7}$$

where the diffusion coefficient $\sigma$ is the positive two-valued function, discontinuous at the threshold $c$,

$$\sigma(y) = \underline{\sigma} I_{[c,\infty)}(y) + \overline{\sigma} I_{(-\infty,c)}(y), \ \forall y \in \mathbb{R}. \tag{3.8}$$

There is a seeming connection to the bandit model - smaller volatility in the region $(c, \infty)$ of gains where there is risk aversion, and greater volatility in the region of losses $(-\infty, c)$ where there is risk loving.[20] In fact, by Keilson and Wellner (1978, Theorem 1), the time 1 value $W_1^c$ of the oscillating Brownian motion has distribution given by the density $q$ referred to in Theorem 2.1(i).

[20] Reversing the roles of $\overline{\sigma}$ and $\underline{\sigma}$ also defines an oscillating Brownian motion, but one that is irrelevant here given the assumption of loss aversion.

**Theorem 3.2.** *Let $(X_i)$ be such that $X_i \in \mathcal{H}$ for each $i$, and let $(X_i)$ and $\mathcal{P}$ satisfy (3.3) and (3.4). Assume also the Lindeberg condition (3.5), that measures in $\mathcal{P}$ are equivalent on each $\mathcal{G}_i$, and that quadratic-consistency is satisfied. Set $\theta = \underline{\sigma}/\overline{\sigma}$. For any $c \in \mathbb{R}$ and $\varphi_1 \in C_b^3(\mathbb{R}_+)$, with $\varphi_1(0) = 0$, define $\varphi$ by*

$$\varphi(x) = \begin{cases} \varphi_1(x-c) & x \geq c \\ -\frac{1}{\theta}\varphi_1(-\theta(x-c)) & x < c \end{cases} \tag{3.9}$$

*If $\varphi_1''(x) \leq 0$ for $x \geq 0$, then*

$$\lim_{n\to\infty} \sup_{Q\in\mathcal{P}} E_Q\left[\varphi\left(\frac{\Sigma_{i=1}^n X_i}{\sqrt{n}}\right)\right] = E_{P^*}[\varphi(W_1^c)]. \tag{3.10}$$

The most important point to make about the theorem is that all its assumptions are satisfied by the bandit model with $\mathcal{P}$ defined by (3.1). (The Lindeberg condition (3.5) is satisfied because of the finiteness of $\Omega_i = \overline{\Omega}$.) Therefore, using also the noted density for $W_1^c$, the CLT implies Theorem 2.1(i). Though the bandit theorem is stated with reference only to a density and not to oscillating Brownian motions, we prefer to include the latter here because it is more revealing of what underlies the limit and, to a degree, how the limit result is proven.

For perspective, if instead of defining $\varphi$ by (3.9), we took $\varphi$ to be any (suitably bounded, smooth and) *globally concave* function, then the limit in (3.10) would equal the expected value of $\varphi$ under $\mathbb{N}(0, \underline{\sigma}^2)$, as in the classic case with fixed variance $\underline{\sigma}^2$. Informally, this result is suggested by taking $c \to -\infty$ above. (For a rigorous argument, see Proposition 2.2.15 and Theorem 2.4.4 in Peng (2019).)

Some extensions of the CLT are possible. For example, one can obtain similar limits with any combination of the modifications $\theta = \overline{\sigma}/\underline{\sigma}$, $\varphi_1''(x) \geq 0$ on $(0, \infty)$, and/or considering the limit of the lower expectation $\inf_{Q\in\mathcal{P}} E_Q\left[\varphi\left(\Sigma_1^n X_i/\sqrt{n}\right)\right]$. These extensions do not seem relevant to the bandit problem, but the reader can find them in our working paper version listed in the bibliography. It is also possible to derive closed-form limiting results for other integrands (functions $\varphi$), for example, for some indicator functions (Appendix A.3). For many other functions $\varphi$, the corresponding expressions for the limit are more complex, less transparent and arguably intractable, and consequently are excluded.

We conclude with mention of related CLTs in the literature. Chen and Epstein (2022) establish CLTs assuming, contrary to (3.3)-(3.4), that conditional means lie in an interval $\left[\underline{\mu}, \overline{\mu}\right]$ while all conditional variances equal a constant $\sigma^2$. However, their theorems are substantially different, for example, limits have a different form and proofs are much different. There exist other generalizations of the classic CLT that are motivated by robustness to ambiguity. In both Marinacci (1999, Theorem 16) and Epstein et al. (2016), experiments are not ordered and their analyses are

better suited for a cross-sectional, rather than sequential, context. Another difference is that in both cases, limiting distributions are normal. Peng (2007, 2019) and Fang et al. (2019) assume that experiments are ordered. Comparison with Theorem 3.2 of the latter is representative. It is more general in permitting ambiguity about both mean and variance. For purposes of comparison, limit attention to the special case of their theorem where there is ambiguity about variance only. Even then, an important difference, particularly given the application developed here, is that greater generality comes arguably at the cost of reduced tractability. In particular, limits are much more complicated (they involve Peng's (2007) notion of a "G-normal" distribution), and a counterpart of Theorem 3.2 is not apparent from their results. Moreover, their CLT is not applicable to the bandit problem with learning, because their key "consistency" assumption is closely related to rectangularity, (see the footnote prior to Lemma 3.1), and suffers from the same limitations. Finally, none of the above papers recognize the potential application to (Bayesian) sequential decision problems.

# A Appendix: Main Proofs

The notation and assumptions in Theorem 3.2 are adopted throughout this appendix. In addition, for any $X$ in $\mathcal{H}$, its *upper expectation* is defined by

$$\mathbb{E}[X] \equiv \sup_{Q \in \mathcal{P}} E_Q[X].$$

Let $(B_t)$ be the standard Brownian motion under a probability space $(\Omega^*, \mathcal{F}^*, P^*)$, and let $(\mathcal{F}_t)_{t \geq 0}$ be the natural filtration generated by $(B_t)_{t \geq 0}$.

## A.1 Lemmas

For a small fixed $h > 0$, and any fixed $(t, x, c) \in [0, 1+h] \times \mathbb{R} \times \mathbb{R}$, $(Y_s^{t,x,c})_{s \in [t,1+h]}$ denotes the solution of the SDE

$$\begin{cases} dY_s^{t,x,c} = \sigma\left(Y_s^{t,x,c}\right) dB_s, \quad s \in [t, 1+h] \\ Y_t^{t,x,c} = x, \end{cases} \tag{A.1}$$

where $\sigma(y) = \underline{\sigma} I_{[c,\infty)}(y) + \overline{\sigma} I_{(-\infty,c)}(y), \ \forall y \in \mathbb{R}$.

By Keilson and Wellner (1978, Theorem 1), (see also Chen and Zili (2015)), the transition probability density of $(Y_s^{t,x,c})_{s \in [t,1+h]}$ is given by, for any $t < s \leq 1+h$ and $y \in \mathbb{R}$,

$$q^c(t, x; s, y) = \frac{1}{\sqrt{2\pi(s-t)}} \frac{1}{\sigma(y)} \exp\left( - \frac{\left( \frac{x-c}{\sigma(x)} - \frac{y-c}{\sigma(y)} \right)^2}{2(s-t)} \right)$$

$$+\frac{\overline{\sigma}-\underline{\sigma}}{\overline{\sigma}+\underline{\sigma}}\frac{1}{\sqrt{2\pi(s-t)}}\frac{sgn(y-c)}{\sigma(y)}\exp\left(-\frac{\left(\left|\frac{x-c}{\sigma(x)}\right|+\left|\frac{y-c}{\sigma(y)}\right|\right)^2}{2(s-t)}\right). \tag{A.2}$$

Given $\varphi_1 \in C_b^3(\mathbb{R}_+)$, $\varphi$ is defined by (3.9). Then

$$\varphi \in C_b^1(\mathbb{R}) \text{ and } \varphi''(z+c) = -\tfrac{\underline{\sigma}}{\overline{\sigma}}\varphi''(-\tfrac{\underline{\sigma}}{\overline{\sigma}}z+c),\ \forall z<0.$$

Define the set of functions $\{H_t\}_{t\in[0,1+h]}$ by

$$H_t(x) = E_{P^*}\left[\varphi\left(Y_{1+h}^{t,x,c}\right)\right], \quad \forall x \in \mathbb{R}. \tag{A.3}$$

Then

$$H_{1+h}(x) = \varphi(x), \quad H_0(0) = E_{P^*}[\varphi(Y_{1+h}^{0,0,c})] = E_{P^*}[\varphi(W_{1+h}^c)].$$

The following lemma describes some properties of the functions $\{H_t\}_{t\in[0,1+h]}$.

**Lemma A.1.** *The functions $\{H_t\}$ defined by (A.3) satisfy:*

**(1)** *For any $t \in [0,1]$, $H_t \in C_b^2(\mathbb{R})$, and the first and second derivatives of $H_t$ are bounded uniformly in $t \in [0,1]$.*

**(2)** *There exists a constant $L$ such that, for any $x_1, x_2 \in \mathbb{R}$ and $t \in [0,1]$,*

$$|H_t''(x_1) - H_t''(x_2)| \le L|x_1 - x_2|.$$

**(3)** *If $\varphi''(x) \le 0$ for $x > c$, then*

$$\begin{cases} H_t''(x) \le 0 & \text{for } x \ge c \\ H_t''(x) \ge 0 & \text{for } x \le c. \end{cases}$$

**(4)** *For any $r \in [0, 1+h-t]$,*

$$H_t(x) = E_{P^*}\left[H_{t+r}\left(Y_{t+r}^{t,x,c}\right)\right],\ \forall x \in \mathbb{R}.$$

**(5)** *If $\varphi''(x) \leq 0$ for $x > c$, then*

$$\lim_{n\to\infty} \sum_{m=1}^{n} \sup_{x\in\mathbb{R}} \left| H_{\frac{m-1}{n}}(x) - H_{\frac{m}{n}}(x) - \frac{\overline{\sigma}^2}{2n}\left[H''_{\frac{m}{n}}(x)\right]^+ + \frac{\underline{\sigma}^2}{2n}\left[H''_{\frac{m}{n}}(x)\right]^- \right| = 0.$$

**(6)** *There exists a constant $C_0$ such that*

$$\sup_{x\in\mathbb{R}} |H_1(x) - \varphi(x)| \leq C_0\sqrt{\underline{\sigma}^2 + \overline{\sigma}^2}\sqrt{h}.$$

**Proof: (1)** Given the transition probability density in (A.2), we have, for $t \in [0,1]$,

$$H_t(x) = \int_{-\infty}^{\infty} \varphi(y) q^c(t,x;1+h,y)dy, \quad \forall x \in \mathbb{R}.$$

For $T = 1 + h$, we have

$$H'_t(x) = \begin{cases} \frac{1}{\underline{\sigma}\sqrt{2\pi(T-t)}} \int_0^\infty \varphi'_1(y)\left[e^{-\frac{(x-c+y)^2}{2\underline{\sigma}^2(T-t)}} + e^{-\frac{(x-c-y)^2}{2\underline{\sigma}^2(T-t)}}\right] dy & \text{if } x \geq c \\ \frac{1}{\overline{\sigma}\sqrt{2\pi(T-t)}} \int_0^\infty \varphi'_1(\frac{\underline{\sigma}}{\overline{\sigma}}y)\left[e^{-\frac{(x-c+y)^2}{2\overline{\sigma}^2(T-t)}} + e^{-\frac{(x-c-y)^2}{2\overline{\sigma}^2(T-t)}}\right] dy & \text{if } x \leq c \end{cases}$$

$$H''_t(x) = \begin{cases} \frac{1}{\underline{\sigma}\sqrt{2\pi(T-t)}} \int_0^\infty \varphi''_1(y) e^{-\frac{(x-c-y)^2}{2\underline{\sigma}^2(T-t)}}\left[1 - e^{-\frac{2y(x-c)}{\underline{\sigma}^2(T-t)}}\right] dy & \text{if } x \geq c \\ \frac{-1}{\overline{\sigma}\sqrt{2\pi(T-t)}} \int_0^\infty \frac{\underline{\sigma}}{\overline{\sigma}}\varphi''_1(\frac{\underline{\sigma}}{\overline{\sigma}}y) e^{-\frac{(x-c+y)^2}{2\overline{\sigma}^2(T-t)}}\left[1 - e^{\frac{2y(x-c)}{\overline{\sigma}^2(T-t)}}\right] dy & \text{if } x \leq c. \end{cases}$$

The assertion follows from $\varphi_1 \in C_b^3(\mathbb{R}_+)$ and the definition of $\varphi$ in (3.9).
**(2)** For any $x > c$, $H'''_t(x) =$

$$\frac{1}{\underline{\sigma}\sqrt{2\pi(T-t)}}\left[2\varphi''_1(0)e^{-\frac{(x-c)^2}{2\underline{\sigma}^2(T-t)}} + \int_0^\infty \varphi'''_1(y)\left(e^{-\frac{(x-c-y)^2}{2\underline{\sigma}^2(T-t)}} + e^{-\frac{(x-c+y)^2}{2\underline{\sigma}^2(T-t)}}\right) dy\right],$$

and, for $x < c$, $H'''_t(x) =$

$$\frac{1}{\overline{\sigma}\sqrt{2\pi(T-t)}}\left[-2\varphi''_1(0)e^{-\frac{(x-c)^2}{2\overline{\sigma}^2(T-t)}} + \int_0^\infty \frac{\underline{\sigma}^2}{\overline{\sigma}^2}\varphi'''_1(\frac{\underline{\sigma}}{\overline{\sigma}}y)\left(e^{-\frac{(x-c-y)^2}{2\overline{\sigma}^2(T-t)}} + e^{-\frac{(x-c+y)^2}{2\overline{\sigma}^2(T-t)}}\right) dy\right].$$

Since $\varphi_1 \in C_b^3(\mathbb{R}_+)$, there exists a constant $L$ such that

$$\sup_{x\in\mathbb{R}, x\neq c} |H'''_t(x)| \leq L \text{ for all } t \in [0,1].$$

The assertion follows by the Mean Value Theorem.

**(3)** It follows from the explicit form of $H_t''(x)$ given above.

**(4)** Since $(Y_s^{t,x,c})$ is a time-homogeneous Markov process, for any $r \in [0, 1+h-t]$,

$$H_t(x) = E_{P^*}[\varphi(Y_{1+h}^{t,x,c})] = E_{P^*}\left[E_{P^*}[\varphi(Y_{1+h}^{t,x,c})|\mathcal{F}_{t+r}]\right] = E_{P^*}[H_{t+r}(Y_{t+r}^{t,x,c})].$$

**(5)** It follows from part (4) that, for any $1 \le m \le n$,

$$H_{\frac{m-1}{n}}(x) = E_{P^*}\left[H_{\frac{m}{n}}\left(Y_{\frac{m}{n}}^{\frac{m-1}{n},x,c}\right)\right].$$

Apply Itô's formula to $H_{\frac{m}{n}}\left(Y_{\frac{m}{n}}^{\frac{m-1}{n},x,c}\right)$ to derive

$$\begin{aligned} H_{\frac{m}{n}}\left(Y_{\frac{m}{n}}^{\frac{m-1}{n},x,c}\right) =& H_{\frac{m}{n}}(x) + \int_{\frac{m-1}{n}}^{\frac{m}{n}} H'_{\frac{m}{n}}\left(Y_s^{\frac{m-1}{n},x,c}\right)\sigma\left(Y_s^{\frac{m-1}{n},x,c}\right)dB_s \\ &+ \frac{1}{2}\int_{\frac{m-1}{n}}^{\frac{m}{n}} H''_{\frac{m}{n}}\left(Y_s^{\frac{m-1}{n},x,c}\right)\left(\sigma\left(Y_s^{\frac{m-1}{n},x,c}\right)\right)^2 ds \end{aligned}$$

Using parts (3) and (4), we have

$$\begin{aligned} H_{\frac{m-1}{n}}(x) &= E_{P^*}\left[H_{\frac{m}{n}}\left(Y_{\frac{m}{n}}^{\frac{m-1}{n},x,c}\right)\right] = \\ &E_{P^*}\left[H_{\frac{m}{n}}(x) + \frac{\overline{\sigma}^2}{2}\int_{\frac{m-1}{n}}^{\frac{m}{n}}\left[H''_{\frac{m}{n}}\left(Y_s^{\frac{m-1}{n},x,c}\right)\right]^+ ds - \frac{\underline{\sigma}^2}{2}\int_{\frac{m-1}{n}}^{\frac{m}{n}}\left[H''_{\frac{m}{n}}\left(Y_s^{\frac{m-1}{n},x,c}\right)\right]^- ds\right] \end{aligned}$$

Thus

$$\begin{aligned} &\sum_{m=1}^n \sup_{x\in\mathbb{R}}\left|H_{\frac{m-1}{n}}(x) - H_{\frac{m}{n}}(x) - \frac{\overline{\sigma}^2}{2n}\left[H''_{\frac{m}{n}}(x)\right]^+ + \frac{\underline{\sigma}^2}{2n}\left[H''_{\frac{m}{n}}(x)\right]^-\right| \\ \le& \sum_{m=1}^n \sup_{x\in\mathbb{R}} E_{P^*}\left[\frac{\underline{\sigma}^2+\overline{\sigma}^2}{2}\int_{\frac{m-1}{n}}^{\frac{m}{n}}\left|H''_{\frac{m}{n}}\left(Y_s^{\frac{m-1}{n},x,c}\right) - H''_{\frac{m}{n}}(x)\right| ds\right] \\ \le& \sum_{m=1}^n \sup_{x\in\mathbb{R}} \frac{(\underline{\sigma}^2+\overline{\sigma}^2)L}{2n} E_{P^*}\left[\sup_{s\in[\frac{m-1}{n},\frac{m}{n}]}\left|Y_s^{\frac{m-1}{n},x,c} - x\right|\right] \\ \le& \sum_{m=1}^n \sup_{x\in\mathbb{R}} \frac{C}{n}\left(E_{P^*}\left[\int_{\frac{m-1}{n}}^{\frac{m}{n}}\left(\sigma\left(Y_s^{\frac{m-1}{n},x,c}\right)\right)^2 dr\right]\right)^{\frac{1}{2}} \le \frac{C\sqrt{\underline{\sigma}^2+\overline{\sigma}^2}}{\sqrt{n}}, \end{aligned}$$

where $C$ is a constant that depends only on $\underline{\sigma}, \overline{\sigma}, L$.

**(6)** Since $\varphi \in C_b^1(\mathbb{R})$, $C_0 \equiv \|\varphi'\| = \sup_{x\in\mathbb{R}}|\varphi'(x)| < \infty$, and

$$\sup_{x\in\mathbb{R}}|H_1(x) - \varphi(x)| = \sup_{x\in\mathbb{R}}\left|E_{P^*}[\varphi(Y_{1+h}^{1,x,c})] - \varphi(x)\right|$$

$$
\begin{aligned}
&\leq \sup_{x\in\mathbb{R}} E_{P^*}\left[\left|\varphi(Y_{1+h}^{1,x,c})-\varphi(x)\right|\right]\\
&\leq \sup_{x\in\mathbb{R}} C_0 E_{P^*}\left[\left|\int_1^{1+h}\sigma\left(Y_s^{1,x,c}\right)dB_s\right|\right]\\
&\leq \sup_{x\in\mathbb{R}} C_0\left(E_{P^*}\left[\int_1^{1+h}\left(\sigma\left(Y_s^{1,x,c}\right)\right)^2 ds\right]\right)^{\frac{1}{2}}\\
&\leq C_0\sqrt{\underline{\sigma}^2+\overline{\sigma}^2}\sqrt{h}. \qquad \blacksquare
\end{aligned}
$$

**Lemma A.2.** *Let $\{H_t\}_{t\in[0,1]}$ be the functions defined in (A.3), and define the family of functions $\{L_{m,n}\}_{m=1}^n$ by*

$$
L_{m,n}(x)=H_{\frac{m}{n}}(x)+\frac{\overline{\sigma}^2}{2n}\left[H''_{\frac{m}{n}}(x)\right]^+-\frac{\underline{\sigma}^2}{2n}\left[H''_{\frac{m}{n}}(x)\right]^-. \tag{A.4}
$$

*Then*

$$
\lim_{n\to\infty}\sum_{m=1}^n\left|\mathbb{E}\left[H_{\frac{m}{n}}\left(\frac{\Sigma_1^m X_i}{\sqrt{n}}\right)\right]-\mathbb{E}\left[L_{m,n}\left(\frac{\Sigma_1^{m-1}X_i}{\sqrt{n}}\right)\right]\right|=0. \tag{A.5}
$$

**Proof:** It suffices to prove

$$
\lim_{n\to\infty}\sum_{m=1}^n\left|\mathbb{E}\left[H_{\frac{m}{n}}\left(\frac{\Sigma_1^m X_i}{\sqrt{n}}\right)\right]-f(m,n)\right|=0 \text{ and} \tag{A.6}
$$

$$
\lim_{n\to\infty}\sum_{m=1}^n\left|f(m,n)-\mathbb{E}\left[L_{m,n}\left(\frac{\Sigma_1^{m-1}X_i}{\sqrt{n}}\right)\right]\right|=0, \tag{A.7}
$$

where

$$
f(m,n)=\mathbb{E}\left[H_{\frac{m}{n}}\left(\frac{\Sigma_1^{m-1}X_i}{\sqrt{n}}\right)+H'_{\frac{m}{n}}\left(\frac{\Sigma_1^{m-1}X_i}{\sqrt{n}}\right)\frac{X_m}{\sqrt{n}}+H''_{\frac{m}{n}}\left(\frac{\Sigma_1^{m-1}X_i}{\sqrt{n}}\right)\frac{X_m^2}{2n}\right].
$$

By Lemma A.1, there exists $L>0$ such that

$$
\sup_{t\in[0,1]}\sup_{x\in\mathbb{R}}|H''_t(x)|\leq L,\ \sup_{t\in[0,1]}\sup_{x,y\in\mathbb{R},x\neq y}\frac{|H''_t(x)-H''_t(y)|}{|x-y|}\leq L.
$$

By the Taylor expansion of $H_t\in C_b^2(\mathbb{R})$, $\forall\epsilon>0$ $\exists\delta>0$ ($\delta$ depends only on $L$ and $\epsilon$), such that, for any $x,y\in\mathbb{R}$ and $t\in[0,1]$,

$$
\left|H_t(x+y)-H_t(x)-H'_t(x)y-\frac{1}{2}H''_t(x)y^2\right|\leq\epsilon|y|^2I_{\{|y|<\delta\}}+L|y|^2I_{\{|y|\geq\delta\}}. \tag{A.8}
$$

Let $x = \Sigma_1^{m-1} X_i/\sqrt{n}$ and $y = X_m/\sqrt{n}$ in (A.8) to derive

$$\sum_{m=1}^{n} \left| \mathbb{E}\left[ H_{\frac{m}{n}}\left( \frac{\Sigma_1^m X_i}{\sqrt{n}} \right) \right] - f(m,n) \right| \le \overline{var}\,\epsilon + \frac{L}{n} \sum_{m=1}^{n} \mathbb{E}\left[ |X_m|^2 I_{\{|X_m| \ge \sqrt{n}\delta\}} \right].$$

By the arbitrariness of $\epsilon$ and the Lindeberg condition (3.5), we obtain (A.6).

By Lemma 3.1 and assumptions (3.3) and (3.4), we have

$$\begin{aligned}
&\sum_{m=1}^{n} \left| f(m,n) - \mathbb{E}\left[ L_{m,n}\left( \frac{\Sigma_1^{m-1} X_i}{\sqrt{n}} \right) \right] \right| \\
=&\sum_{m=1}^{n} \left| \mathbb{E}\left[ H_{\frac{m}{n}}\left( \frac{\Sigma_1^{m-1} X_i}{\sqrt{n}} \right) + H'_{\frac{m}{n}}\left( \frac{\Sigma_1^{m-1} X_i}{\sqrt{n}} \right) \frac{X_m}{\sqrt{n}} + H''_{\frac{m}{n}}\left( \frac{\Sigma_1^{m-1} X_i}{\sqrt{n}} \right) \frac{X_m^2}{2n} \right] \right. \\
&\left. -\mathbb{E}\left[ H_{\frac{m}{n}}\left( \frac{\Sigma_1^{m-1} X_i}{\sqrt{n}} \right) + \frac{\overline{\sigma}^2}{2n}\left[ H''_{\frac{m}{n}}\left( \frac{\Sigma_1^{m-1} X_i}{\sqrt{n}} \right) \right]^+ - \frac{\underline{\sigma}^2}{2n}\left[ H''_{\frac{m}{n}}\left( \frac{\Sigma_1^{m-1} X_i}{\sqrt{n}} \right) \right]^- \right] \right| \\
=&\sum_{m=1}^{n} \left| \sup_{Q\in\mathcal{P}} E_Q\left[ H_{\frac{m}{n}}\left( \frac{\Sigma_1^{m-1} X_i}{\sqrt{n}} \right) + \underset{Q'\in\mathcal{P}(Q,m-1)}{ess\,\sup} E_{Q'}\left[ H''_{\frac{m}{n}}\left( \frac{\Sigma_1^{m-1} X_i}{\sqrt{n}} \right) \frac{X_m^2}{2n} \middle| \mathcal{G}_{m-1} \right] \right] \right. \\
&\left. - \sup_{Q\in\mathcal{P}} E_Q\left[ H_{\frac{m}{n}}\left( \frac{\Sigma_1^{m-1} X_i}{\sqrt{n}} \right) + \frac{\overline{\sigma}^2}{2n}\left[ H''_{\frac{m}{n}}\left( \frac{\Sigma_1^{m-1} X_i}{\sqrt{n}} \right) \right]^+ - \frac{\underline{\sigma}^2}{2n}\left[ H''_{\frac{m}{n}}\left( \frac{\Sigma_1^{m-1} X_i}{\sqrt{n}} \right) \right]^- \right] \right| \\
\le&\frac{1}{2n}\sum_{m=1}^{n} \sup_{Q\in\mathcal{P}} E_Q\left[ \left| \overline{\sigma}^2 - \underset{Q'\in\mathcal{P}(Q,m-1)}{ess\,\sup} E_{Q'}[X_m^2|\mathcal{G}_{m-1}] \right| \left[ H''_{\frac{m}{n}}\left( \frac{\Sigma_1^{m-1} X_i}{\sqrt{n}} \right) \right]^+ \right. \\
&\qquad \left. + \left| \underline{\sigma}^2 - \underset{Q'\in\mathcal{P}(Q,m-1)}{ess\,\inf} E_{Q'}[X_m^2|\mathcal{G}_{m-1}] \right| \left[ H''_{\frac{m}{n}}\left( \frac{\Sigma_1^{m-1} X_i}{\sqrt{n}} \right) \right]^- \right] \\
\le&\frac{C}{2n}\sum_{m=1}^{n} \sup_{Q\in\mathcal{P}} E_Q\left[ \left| \overline{\sigma}^2 - \underset{Q'\in\mathcal{P}(Q,m-1)}{ess\,\sup} E_{Q'}[X_m^2|\mathcal{G}_{m-1}] \right| + \left| \underline{\sigma}^2 - \underset{Q'\in\mathcal{P}(Q,m-1)}{ess\,\inf} E_{Q'}[X_m^2|\mathcal{G}_{m-1}] \right| \right] \\
\to&0 \quad \text{as } n \to \infty,
\end{aligned}$$

where $C$ is a constant uniform bound for $|H_t''(x)|$.

This implies (A.7) and completes the proof. ■

## A.2 Proof of the CLT (Theorem 3.2)

For $h > 0$ sufficiently small, let $\{H_t\}_{t\in[0,1+h]}$ be the functions defined by (A.3). First prove

$$\lim_{n\to\infty} \left| \mathbb{E}\left[ H_1\left( \frac{\Sigma_1^n X_i}{\sqrt{n}} \right) \right] - E_{P^*}\left[ \varphi\left( W_{1+h}^c \right) \right] \right| = 0.$$

We have

$$\begin{aligned}
&\mathbb{E}\left[H_1\left(\frac{\Sigma_1^n X_i}{\sqrt{n}}\right)\right] - E_{P^*}\left[\varphi\left(W_{1+h}^c\right)\right] \\
=&\mathbb{E}\left[H_1\left(\frac{\Sigma_1^n X_i}{\sqrt{n}}\right)\right] - H_0(0) \\
=&\sum_{m=1}^{n}\left\{\mathbb{E}\left[H_{\frac{m}{n}}\left(\frac{\Sigma_1^m X_i}{\sqrt{n}}\right)\right] - \mathbb{E}\left[H_{\frac{m-1}{n}}\left(\frac{\Sigma_1^{m-1} X_i}{\sqrt{n}}\right)\right]\right\} \\
=&\sum_{m=1}^{n}\left\{\mathbb{E}\left[H_{\frac{m}{n}}\left(\frac{\Sigma_1^m X_i}{\sqrt{n}}\right)\right] - \mathbb{E}\left[L_{m,n}\left(\frac{\Sigma_1^{m-1} X_i}{\sqrt{n}}\right)\right]\right\} \\
&+\sum_{m=1}^{n}\left\{\mathbb{E}\left[L_{m,n}\left(\frac{\Sigma_1^{m-1} X_i}{\sqrt{n}}\right)\right] - \mathbb{E}\left[H_{\frac{m-1}{n}}\left(\frac{\Sigma_1^{m-1} X_i}{\sqrt{n}}\right)\right]\right\} \\
=&: I_{1n} + I_{2n},
\end{aligned}$$

where $L_{m,n}(x) = H_{\frac{m}{n}}(x) + \frac{\overline{\sigma}^2}{2n}\left[H''_{\frac{m}{n}}(x)\right]^+ - \frac{\underline{\sigma}^2}{2n}\left[H''_{\frac{m}{n}}(x)\right]^-,\ 1 \le m \le n.$

By Lemma A.2,

$$|I_{1n}| \le \sum_{m=1}^{n}\left|\mathbb{E}\left[H_{\frac{m}{n}}\left(\frac{\Sigma_1^m X_i}{\sqrt{n}}\right)\right] - \mathbb{E}\left[L_{m,n}\left(\frac{\Sigma_1^{m-1} X_i}{\sqrt{n}}\right)\right]\right| \to 0 \quad \text{as } n \to \infty.$$

Furthermore, by Lemma A.1(5), as $n \to \infty$,

$$\begin{aligned}
|I_{2n}| &\le \sum_{m=1}^{n}\mathbb{E}\left[\left|L_{m,n}\left(\frac{\Sigma_1^{m-1} X_i}{\sqrt{n}}\right) - H_{\frac{m-1}{n}}\left(\frac{\Sigma_1^{m-1} X_i}{\sqrt{n}}\right)\right|\right] \\
&\le \sum_{m=1}^{n}\sup_{x\in\mathbb{R}}\left|L_{m,n}(x) - H_{\frac{m-1}{n}}(x)\right| \\
&= \sum_{m=1}^{n}\sup_{x\in\mathbb{R}}\left|H_{\frac{m-1}{n}}(x) - H_{\frac{m}{n}}(x) - \frac{\overline{\sigma}^2}{2n}\left[H''_{\frac{m}{n}}(x)\right]^+ + \frac{\underline{\sigma}^2}{2n}\left[H''_{\frac{m}{n}}(x)\right]^-\right| \to 0.
\end{aligned}$$

By Lemma A.1(6), $\lim_{n\to\infty}\left|\mathbb{E}\left[\varphi\left(\frac{\Sigma_1^n X_i}{\sqrt{n}}\right)\right] - E_{P^*}\left[\varphi\left(W_1^c\right)\right]\right| \le$
$\lim_{n\to\infty}\left|\mathbb{E}\left[\varphi\left(\frac{\Sigma_1^n X_i}{\sqrt{n}}\right)\right] - \mathbb{E}\left[H_1\left(\frac{\Sigma_1^n X_i}{\sqrt{n}}\right)\right]\right| +$
$\lim_{n\to\infty}\left|\mathbb{E}\left[H_1\left(\frac{\Sigma_1^n X_i}{\sqrt{n}}\right)\right] - E_{P^*}\left[\varphi\left(W_{1+h}^c\right)\right]\right| + \left|E_{P^*}\left[\varphi\left(W_{1+h}^c\right)\right] - E_{P^*}\left[\varphi\left(W_1^c\right)\right]\right|$
$\le \sup_{x\in\mathbb{R}}\left|H_1(x) - \varphi(x)\right| + C_0\sqrt{\overline{\sigma}^2 + \underline{\sigma}^2}\sqrt{h} \ \le\ 2C_0\sqrt{\overline{\sigma}^2 + \underline{\sigma}^2}\sqrt{h}.$
Since $h$ is arbitrary, the proof is complete. ■

## A.3 A corollary

Indicator functions for one-sided intervals $[c, \infty)$ can be suitably approximated by functions $\varphi$ satisfying the conditions in Theorem 3.2, which suggests that the limiting result (3.10) is valid also for such indicators. The following corollary confirms this, and is of interest also because it is used below in the proof of Theorem 2.2. See our working paper version (Corollary 3.4) for a more general result that considers also indicators for intervals of the form $(-\infty, c]$.

**Corollary A.3.** *Adopt the assumptions in Theorem 3.2. Then, for any $c \in \mathbb{R}$,*

$$\lim_{n\to\infty} \sup_{Q\in\mathcal{P}} Q\left(\frac{\Sigma_1^n X_i}{\sqrt{n}} \geq c\right) = P^*\left(W_1^c \geq c\right) \tag{A.9}$$

*and*

$$P^*\left(W_1^c \geq c\right) = \begin{cases} \frac{2\overline{\sigma}}{\overline{\sigma}+\underline{\sigma}}\Phi\left(-\frac{c}{\overline{\sigma}}\right) & c > 0 \\ 1 - \frac{2\underline{\sigma}}{\overline{\sigma}+\underline{\sigma}}\Phi\left(\frac{c}{\underline{\sigma}}\right) & c \leq 0 \end{cases}, \tag{A.10}$$

*where $\Phi$ is the standard normal cdf.*

**Proof:** For any $c \in \mathbb{R}$ and $\varepsilon > 0$, suppose that $f_1, g_1 \in C_b^3(\mathbb{R}_+)$ satisfy

$$\begin{cases} f_1(x) = 1 \text{ for } x \geq \frac{\sigma}{\overline{\sigma}}\varepsilon \\ f_1''(x) \leq 0 \text{ for } x \geq 0 \\ f_1(0) = \frac{\overline{\sigma}}{\overline{\sigma}+\underline{\sigma}} \end{cases} \quad \begin{cases} g_1(x) = 1 \text{ for } x \geq \varepsilon \\ g_1''(x) \leq 0 \text{ for } x \geq 0 \\ g_1(0) = \frac{\overline{\sigma}}{\overline{\sigma}+\underline{\sigma}} \end{cases}$$

Define $f_\varepsilon$ and $g_\varepsilon$ by

$$f_\varepsilon(x) = \begin{cases} f_1(x - c - \varepsilon) & \text{for } x \geq c + \varepsilon \\ -\frac{\overline{\sigma}}{\underline{\sigma}} f_1\left(-\frac{\underline{\sigma}}{\overline{\sigma}}(x - c - \varepsilon)\right) + \frac{\overline{\sigma}}{\underline{\sigma}} & \text{for } x \leq c + \varepsilon \end{cases} \tag{A.11}$$

$$g_\varepsilon(x) = \begin{cases} g_1(x - c + \varepsilon) & \text{for } x \geq c - \varepsilon \\ -\frac{\overline{\sigma}}{\underline{\sigma}} g_1\left(-\frac{\underline{\sigma}}{\overline{\sigma}}(x - c + \varepsilon)\right) + \frac{\overline{\sigma}}{\underline{\sigma}} & \text{for } x \leq c - \varepsilon \end{cases} \tag{A.12}$$

It can be checked that

$$g_\varepsilon(x) \geq I_{[c,\infty)}(x) \geq f_\varepsilon(x) \text{ and}$$
$$|g_\varepsilon(x) - f_\varepsilon(x)| \leq I_{\left[c-(1+\frac{\overline{\sigma}}{\underline{\sigma}})\varepsilon, c+(1+\frac{\sigma}{\overline{\sigma}})\varepsilon\right]}(x), \ \forall x \in \mathbb{R}.$$

Consider the solution $(\widetilde{W}_t^x)_{t\geq 0}$ of the SDE

$$\begin{cases} d\widetilde{W}_t^x = \left(\underline{\sigma} I_{[0,\infty)}\left(\widetilde{W}_t^x\right) + \overline{\sigma} I_{(-\infty,0)}\left(\widetilde{W}_t^x\right)\right) dB_t, \quad t \geq 0 \\ \widetilde{W}_0^x = x. \end{cases} \tag{A.13}$$

Then $W_1^c$ and $c + \widetilde{W}_1^{-c}$ are described by the same law, and

$$
\begin{aligned}
&\left| \sup_{Q \in \mathcal{P}} Q\left( \frac{\Sigma_1^n X_i}{\sqrt{n}} \geq c \right) - P^*\left( W_1^c \geq c \right) \right| \\
&\leq \left| \mathbb{E}\left[ f_\varepsilon\left( \frac{\Sigma_1^n X_i}{\sqrt{n}} \right) \right] - E_{P^*}\left[ g_\varepsilon\left( W_1^c \right) \right] \right| + \left| \mathbb{E}\left[ g_\varepsilon\left( \frac{\Sigma_1^n X_i}{\sqrt{n}} \right) \right] - E_{P^*}\left[ f_\varepsilon\left( W_1^c \right) \right] \right| \\
&\leq \left| \mathbb{E}\left[ f_\varepsilon\left( \frac{\Sigma_1^n X_i}{\sqrt{n}} \right) \right] - E_{P^*}\left[ f_\varepsilon\left( W_1^{c+\varepsilon} \right) \right] \right| + \left| \mathbb{E}\left[ g_\varepsilon\left( \frac{\Sigma_1^n X_i}{\sqrt{n}} \right) \right] - E_{P^*}\left[ g_\varepsilon\left( W_1^{c-\varepsilon} \right) \right] \right| \\
&\quad + \left| E_{P^*}\left[ f_\varepsilon\left( W_1^{c+\varepsilon} \right) - f_\varepsilon\left( W_1^c \right) \right] \right| + \left| E_{P^*}\left[ g_\varepsilon\left( W_1^{c-\varepsilon} \right) - g_\varepsilon\left( W_1^c \right) \right] \right| \\
&\quad + 2 \left| E_{P^*}\left[ f_\varepsilon\left( W_1^c \right) - g_\varepsilon\left( W_1^c \right) \right] \right| \\
&\leq \left| \mathbb{E}\left[ f_\varepsilon\left( \frac{\Sigma_1^n X_i}{\sqrt{n}} \right) \right] - E_{P^*}\left[ f_\varepsilon\left( W_1^{c+\varepsilon} \right) \right] \right| + \left| \mathbb{E}\left[ g_\varepsilon\left( \frac{\Sigma_1^n X_i}{\sqrt{n}} \right) \right] - E_{P^*}\left[ g_\varepsilon\left( W_1^{c-\varepsilon} \right) \right] \right| \\
&\quad + E_{P^*}\left[ \left| f_\varepsilon\left( c + \varepsilon + \widetilde{W}_1^{-c-\varepsilon} \right) - f_\varepsilon\left( c + \widetilde{W}_1^{-c} \right) \right| + \left| g_\varepsilon\left( c - \varepsilon + \widetilde{W}_1^{-c+\varepsilon} \right) - g_\varepsilon\left( c + \widetilde{W}_1^{-c} \right) \right| \right] \\
&\quad + 2 \left| E_{P^*}\left[ f_\varepsilon\left( W_1^c \right) - g_\varepsilon\left( W_1^c \right) \right] \right| \\
&\leq \left| \mathbb{E}\left[ f_\varepsilon\left( \frac{\Sigma_1^n X_i}{\sqrt{n}} \right) \right] - E_{P^*}\left[ f_\varepsilon\left( W_1^{c+\varepsilon} \right) \right] \right| + \left| \mathbb{E}\left[ g_\varepsilon\left( \frac{\Sigma_1^n X_i}{\sqrt{n}} \right) \right] - E_{P^*}\left[ g_\varepsilon\left( W_1^{c-\varepsilon} \right) \right] \right| \\
&\quad + C_0 E_{P^*}\left[ 2\varepsilon + \left| \widetilde{W}_1^{-c-\varepsilon} - \widetilde{W}_1^{-c} \right| + \left| \widetilde{W}_1^{-c+\varepsilon} - \widetilde{W}_1^{-c} \right| \right] + 2P^*\left( c - \left( 1 + \frac{\overline{\sigma}}{\underline{\sigma}} \right) \varepsilon \leq W_1^c \leq c + \left( 1 + \frac{\underline{\sigma}}{\overline{\sigma}} \right) \varepsilon \right),
\end{aligned}
$$

where $C_0$ is a constant that depends on $\|f_\varepsilon'\|, \|g_\varepsilon'\|$. With Le Gall (1984, Theorem 1.5) and Theorem 3.2, the upper probability equation in (A.9) is proven.

The expression (A.10) may be derived by integrating the pdf (B.1)-(B.2). ■

# B Appendix: Proofs for bandits

## B.1 An explicit density

Let $W_1^c$ be the $t = 1$ value of the oscillating Brownian motion defined by (3.7)-(3.8). Keilson and Wellner (1978, Theorem 1) give the following expression for its pdf: For $c \geq 0$,

$$
q(y) = \begin{cases} \frac{1}{\underline{\sigma}\sqrt{2\pi}} \left( e^{-\frac{(\frac{-c}{\overline{\sigma}} - \frac{y-c}{\underline{\sigma}})^2}{2}} + \frac{\overline{\sigma}-\underline{\sigma}}{\underline{\sigma}+\overline{\sigma}} e^{-\frac{(\frac{c}{\overline{\sigma}} + \frac{y-c}{\underline{\sigma}})^2}{2}} \right) & y \geq c \\ \frac{1}{\overline{\sigma}\sqrt{2\pi}} \left( e^{-\frac{(\frac{-c}{\overline{\sigma}} - \frac{y-c}{\overline{\sigma}})^2}{2}} - \frac{\overline{\sigma}-\underline{\sigma}}{\underline{\sigma}+\overline{\sigma}} e^{-\frac{(\frac{c}{\overline{\sigma}} + \frac{c-y}{\overline{\sigma}})^2}{2}} \right) & y < c \end{cases} \tag{B.1}
$$

and for $c < 0$,

$$q(y) = \begin{cases} \frac{1}{\underline{\sigma}\sqrt{2\pi}} \left( e^{-\frac{(\frac{-c}{\overline{\sigma}} - \frac{y-c}{\underline{\sigma}})^2}{2}} + \frac{\overline{\sigma}-\underline{\sigma}}{\underline{\sigma}+\overline{\sigma}} e^{-\frac{(\frac{-c}{\overline{\sigma}} + \frac{y-c}{\underline{\sigma}})^2}{2}} \right) & y \geq c \\ \frac{1}{\overline{\sigma}\sqrt{2\pi}} \left( e^{-\frac{(\frac{-c}{\overline{\sigma}} - \frac{y-c}{\overline{\sigma}})^2}{2}} - \frac{\overline{\sigma}-\underline{\sigma}}{\underline{\sigma}+\overline{\sigma}} e^{-\frac{(\frac{c}{\overline{\sigma}} + \frac{c-y}{\overline{\sigma}})^2}{2}} \right) & y < c \end{cases} \tag{B.2}$$

These expressions are used to derive (2.21) and to prove Corollary A.3 and Theorem 2.1.

## B.2 Proof of Theorem 2.1

As indicated in the text, (i) follows from Theorem 3.2 and the above density; and (ii) follows from (i) by inspection of the above density. It remains to prove (iii).

Take $c \geq 0$. The proof for $c < 0$ is similar. In light of (3.2) and (3.10), it suffices to compute $E_{P^*}[\varphi(W_1^c)]$. Use the pdf of $W_1^c$ in (B.1), to deduce that, for $c \geq 0$,

$$\begin{aligned} E_{P^*}\left[\varphi\left(W_1^c\right)\right] &= \int_c^\infty q\left(y\right)\varphi_1\left(y-c\right)dy + \int_{-\infty}^{c} q\left(y\right)\left[-\frac{\overline{\sigma}}{\underline{\sigma}}\varphi_1\left(-\frac{\underline{\sigma}}{\overline{\sigma}}(y-c)\right)\right]dy \\ &= \frac{1}{\sqrt{2\pi}}\frac{1}{\underline{\sigma}}\int_c^\infty \varphi_1\left(y-c\right)\left[\frac{2\overline{\sigma}}{\overline{\sigma}+\underline{\sigma}}\right]e^{-\frac{\left(\frac{c}{\overline{\sigma}}+\frac{y-c}{\underline{\sigma}}\right)^2}{2}}dy \\ &- \frac{1}{\sqrt{2\pi}}\frac{1}{\underline{\sigma}}\int_{-\infty}^{c} \varphi_1\left(-\frac{\underline{\sigma}}{\overline{\sigma}}(y-c)\right)\left(e^{-\frac{(\frac{-c}{\overline{\sigma}}-\frac{y-c}{\overline{\sigma}})^2}{2}} - \frac{\overline{\sigma}-\underline{\sigma}}{\underline{\sigma}+\overline{\sigma}}e^{-\frac{(\frac{c}{\overline{\sigma}}+\frac{c-y}{\overline{\sigma}})^2}{2}}\right)dy \\ &= \frac{1}{\sqrt{2\pi}}\frac{1}{\underline{\sigma}}\int_0^\infty \varphi_1\left(y\right)\left[\frac{2\overline{\sigma}}{\overline{\sigma}+\underline{\sigma}}\right]e^{-\frac{\left(\frac{c}{\overline{\sigma}}+\frac{y}{\underline{\sigma}}\right)^2}{2}}dy \\ &- \frac{1}{\sqrt{2\pi}}\frac{1}{\underline{\sigma}}\int_{-\infty}^{0} \varphi_1\left(-\frac{\underline{\sigma}}{\overline{\sigma}}y\right)\left(e^{-\frac{(\frac{-c}{\overline{\sigma}}-\frac{y}{\overline{\sigma}})^2}{2}} - \frac{\overline{\sigma}-\underline{\sigma}}{\underline{\sigma}+\overline{\sigma}}e^{-\frac{(\frac{c}{\overline{\sigma}}-\frac{y}{\overline{\sigma}})^2}{2}}\right)dy \\ &= \frac{1}{\sqrt{2\pi}}\frac{1}{\underline{\sigma}}\int_0^\infty \varphi_1\left(y\right)\left[\frac{2\overline{\sigma}}{\overline{\sigma}+\underline{\sigma}}\right]e^{-\frac{\left(\frac{c}{\overline{\sigma}}+\frac{y}{\underline{\sigma}}\right)^2}{2}}dy \\ &- \frac{1}{\sqrt{2\pi}}\frac{1}{\underline{\sigma}}\frac{\overline{\sigma}}{\underline{\sigma}}\int_0^\infty \varphi_1\left(z\right)\left(e^{-\frac{(\frac{-c}{\overline{\sigma}}+\frac{z}{\underline{\sigma}})^2}{2}} - \frac{\overline{\sigma}-\underline{\sigma}}{\underline{\sigma}+\overline{\sigma}}e^{-\frac{(\frac{c}{\overline{\sigma}}+\frac{z}{\underline{\sigma}})^2}{2}}\right)dz \\ &= \frac{1}{\sqrt{2\pi}\underline{\sigma}}\int_0^\infty \varphi_1\left(y\right)\frac{\overline{\sigma}}{\underline{\sigma}}\left[e^{-\frac{(y+m)^2}{2\underline{\sigma}^2}} - e^{-\frac{(y-m)^2}{2\underline{\sigma}^2}}\right]dy \end{aligned}$$

where $m = \frac{\underline{\sigma}}{\overline{\sigma}} c$. Thus we want to prove that

$$\frac{1}{\sqrt{2\pi}\underline{\sigma}} \int_0^\infty \varphi_1(y) \frac{\overline{\sigma}}{\underline{\sigma}} \left[ e^{-\frac{(y+m)^2}{2\underline{\sigma}^2}} - e^{-\frac{(y-m)^2}{2\underline{\sigma}^2}} \right] dy \geq -\frac{\overline{\sigma}}{\underline{\sigma}} \varphi_1\left(\frac{\underline{\sigma}}{\overline{\sigma}} c\right),$$

with equality if and only if $c = 0$. .

It is evident that $E_{P^*}[\varphi(W_1^c)] = 0 = \varphi(0)$ if $c = 0$. Henceforth, take $c > 0$ and prove that

$$\int_0^\infty \varphi_1(y) \left[ \left( e^{-\frac{(y-m)^2}{2\underline{\sigma}^2}} - e^{-\frac{(y+m)^2}{2\underline{\sigma}^2}} \right) / \sqrt{2\pi}\underline{\sigma} \right] dy < \varphi_1(m).$$

Denote by $f(y)$ the expression in the square bracket, (thus $f(y) > 0$ for all $y > 0$), and let $F \equiv \int_0^\infty f(y)\, dy$, $0 < F < 1$. Then $f/F$ is a density. If its mean is $\mu$, then, by strict concavity of $\varphi_1$,

$$\int_0^\infty \varphi_1(y) f(y)\, dy < F\varphi_1(\mu). \tag{B.3}$$

Next we prove that $F\mu = m$:

$$\begin{aligned}
F\mu &= \int_0^\infty y \left[ \left( e^{-\frac{(y-m)^2}{2\underline{\sigma}^2}} - e^{-\frac{(y+m)^2}{2\underline{\sigma}^2}} \right) / \sqrt{2\pi}\underline{\sigma} \right] dy \\
&= \int_{-m}^\infty (z+m) \left[ e^{-\frac{z^2}{2\underline{\sigma}^2}} / \sqrt{2\pi}\underline{\sigma} \right] dz - \int_m^\infty (z-m) \left[ \left( e^{-\frac{z^2}{2\underline{\sigma}^2}} \right) / \sqrt{2\pi}\underline{\sigma} \right] dz \\
&= \int_{-m}^m z \left[ e^{-\frac{z^2}{2\underline{\sigma}^2}} / \sqrt{2\pi}\underline{\sigma} \right] dz + m \int_{-m}^\infty \left[ e^{-\frac{z^2}{2\underline{\sigma}^2}} / \sqrt{2\pi}\underline{\sigma} \right] dz + m \int_m^\infty \left[ e^{-\frac{z^2}{2\underline{\sigma}^2}} / \sqrt{2\pi}\underline{\sigma} \right] dz \\
&= 0 + m\left[ \Pr(Z > -m) + \Pr(Z > m) \right] \quad = m,
\end{aligned}$$

where probabilities are computed according to $\mathbb{N}(0, \underline{\sigma}^2)$.

Finally, $F\mu = m \implies F\varphi_1(\mu) = F\varphi_1(m/F) \leq \varphi_1(m)$, by $F < 1$, $\varphi_1(0) = 0$, and the concavity of $\varphi_1$. Combine with (B.3) to complete the proof. ■

## B.3 Proof of Theorem 2.2

**(i)** We are given that $c = 0$. For small enough $h > 0$, let $\{H_t\}_{t \in [0, 1+h]}$ be the corresponding functions defined by (A.3).

First prove

$$\lim_{n \to \infty} \left| E_{P^{s*}} \left[ H_1 \left( \frac{\Sigma_1^n X_i}{\sqrt{n}} \right) \right] - E_{P^*} \left[ \varphi\left( W_{1+h}^0 \right) \right] \right| = 0 \tag{B.4}$$

We have

$$\begin{aligned}
&E_{P^{s^*}}\left[H_1\left(\frac{\Sigma_1^n X_i}{\sqrt{n}}\right)\right] - E_{P^*}\left[\varphi\left(W_{1+h}^0\right)\right] \\
=&E_{P^{s^*}}\left[H_1\left(\frac{\Sigma_1^n X_i}{\sqrt{n}}\right)\right] - H_0(0) \\
=&\sum_{m=1}^{n}\left\{E_{P^{s^*}}\left[H_{\frac{m}{n}}\left(\frac{\Sigma_1^m X_i}{\sqrt{n}}\right)\right] - E_{P^{s^*}}\left[H_{\frac{m-1}{n}}\left(\frac{\Sigma_1^{m-1} X_i}{\sqrt{n}}\right)\right]\right\} \\
=&\sum_{m=1}^{n}\left\{E_{P^{s^*}}\left[H_{\frac{m}{n}}\left(\frac{\Sigma_1^m X_i}{\sqrt{n}}\right)\right] - E_{P^{s^*}}\left[L_{m,n}\left(\frac{\Sigma_1^{m-1} X_i}{\sqrt{n}}\right)\right]\right\} \\
&+\sum_{m=1}^{n}\left\{E_{P^{s^*}}\left[L_{m,n}\left(\frac{\Sigma_1^{m-1} X_i}{\sqrt{n}}\right)\right] - E_{P^{s^*}}\left[H_{\frac{m-1}{n}}\left(\frac{\Sigma_1^{m-1} X_i}{\sqrt{n}}\right)\right]\right\} \\
=&: J_{1n} + J_{2n},
\end{aligned}$$

where $L_{m,n}(x) = H_{\frac{m}{n}}(x) + \frac{\overline{\sigma}^2}{2n}\left[H''_{\frac{m}{n}}(x)\right]^+ - \frac{\underline{\sigma}^2}{2n}\left[H''_{\frac{m}{n}}(x)\right]^-$, $1 \le m \le n$.

By a similar argument to that in the proof of Lemma A.2, (using Lemma A.1(3) and the fact that $E_{P^{s^*}}[X_m^2|\mathcal{G}_{m-1}] = I_{\{\Sigma_1^{m-1}X_i \le 0\}}\overline{\sigma}^2 + I_{\{\Sigma_1^{m-1}X_i > 0\}}\underline{\sigma}^2$), deduce that

$$\lim_{n\to\infty} |J_{1n}| = 0.$$

On the other hand, by Lemma A.1(5), (argue as in the proof that $|I_{2n}| \to 0$ in Appendix A.2), we have $\lim_{n\to\infty} |J_{2n}| = 0$. Thus we obtain (B.4).

By the definition of functions $\{H_t\}$ and Lemma A.1(6), and arguing as at the end of Appendix A.2, the proof of (i) is complete.

**(ii)** By Corollary A.3, we have that, for any $N > 0$,

$$P^{s^*}\left(\cap_{n=N}^{\infty}\{\Sigma_1^n X_i > 0\}\right) \le \lim_{n\to\infty}\sup_{s\in\mathcal{S}} P^s\left(\Sigma_1^n X_i/\sqrt{n} > 0\right) = \frac{\overline{\sigma}}{\overline{\sigma}+\underline{\sigma}} < 1.$$

By the corresponding result for the indicator of $(-\infty, c]$, (see Corollary 3.4 in our working paper version),

$$P^{s^*}\left(\cap_{n=N}^{\infty}\{\Sigma_1^n X_i \le 0\}\right) \le \lim_{n\to\infty}\sup_{s\in\mathcal{S}} P^s\left(\Sigma_1^n X_i/\sqrt{n} \le 0\right) = \frac{\overline{\sigma}}{\overline{\sigma}+\underline{\sigma}} < 1.$$

**(iii)** To derive (2.25), argue first, as in Corollary A.3, that the indicator for $[0, \infty)$ can be approximated by a function $\varphi$ satisfying conditions of the CLT and the bandit application. Then it can be shown that (2.24) is asymptotically optimal also when the indicator replaces $\varphi$, that is, when DM solves $\sup_{s\in\mathcal{S}} P^s\left(\Sigma_1^n X_i/\sqrt{n} > d\right)$. Finally, apply the closed-form expression in the noted corollary. ■

## B.4 Proof of Theorem 2.4

**Proof of Lemma 2.3**: (i) Let $s \in \mathcal{S}$. Bayesian updating implies that $\{\mu_n\}$ is a $P^s$-martingale adapted to $\{\mathcal{G}_n\}$. Since $\{\mu_n\}$ is uniformly bounded, there exists a random variable $\mu$ such that

$$\lim_{n\to\infty} \mu_n = \mu \quad P^s\text{-}a.s. \tag{B.5}$$

Argue that $\mu = 0$ or $1$ $P^s$-$a.s.$, which implies (2.32). Purely for simplicity, we give the argument when $\underline{p} + \overline{p} = 1$.

We have $P^s(\widehat{\Omega}) = 1$, where $\widehat{\Omega} = \{\omega \in \Omega \mid \lim_{n\to\infty} \mu_n(\omega) = \mu(\omega)\}$. For any $\omega \in \widehat{\Omega}$,

$$\mu_n(\omega) = \frac{\underline{p}\mu_{n-1}(\omega)}{\underline{p}\mu_{n-1}(\omega) + \overline{p}(1-\mu_{n-1}(\omega))} \quad \text{or} \quad \frac{\overline{p}\mu_{n-1}(\omega)}{\overline{p}\mu_{n-1}(\omega) + \underline{p}(1-\mu_{n-1}(\omega))}.$$

Thus, without loss of generality, there exists a subsequence $\{\mu_{k_n}\}$ satisfying

$$\mu_{k_n}(\omega) = \frac{\underline{p}\mu_{k_n-1}(\omega)}{\underline{p}\mu_{k_n-1}(\omega) + \overline{p}(1-\mu_{k_n-1}(\omega))},$$

which implies that

$$\mu(\omega) = \frac{\underline{p}\mu(\omega)}{\underline{p}\mu(\omega) + \overline{p}(1-\mu(\omega))}.$$

Conclude that $\mu(\omega) = 0$ or $1$.

(ii) Let $\nu_n = \mu_n/(1-\mu_n)$ and apply (2.28) to derive, for any $s$,

$$\begin{aligned}
&\log \nu_{n+1} - \log \nu_1 \\
&= \left[(f_a^s(\omega^{(n)}) - f_{a,0}^s(\omega^{(n)})) - (f_b^s(\omega^{(n)}) - f_{b,0}^s(\omega^{(n)}))\right] \log\left(\frac{\underline{p}}{\overline{p}}\right) \\
&+ \left[f_{a,0}^s(\omega^{(n)}) - f_{b,0}^s(\omega^{(n)})\right] \log\left(\frac{1-\underline{p}}{1-\overline{p}}\right).
\end{aligned}$$

Define the sets

$$\begin{aligned}
N_a &= \left\{\omega : \lim_{n\to\infty} f_a^s(\omega^{(n)}) = \infty,\ \lim_{n\to\infty} f_b^s(\omega^{(n)}) < \infty\right\}, \\
N_b &= \left\{\omega : \lim_{n\to\infty} f_a^s(\omega^{(n)}) < \infty,\ \lim_{n\to\infty} f_b^s(\omega^{(n)}) = \infty\right\}, \\
N_{a,b} &= \left\{\omega : \lim_{n\to\infty} f_a^s(\omega^{(n)}) = \infty,\ \lim_{n\to\infty} f_b^s(\omega^{(n)}) = \infty\right\},
\end{aligned}$$

$$M_a = \left\{ \omega : \lim_{n\to\infty} \frac{f^s_{a,0}(\omega^{(n)})}{f^s_a(\omega^{(n)})} = 1 - \underline{p} \right\},$$
$$M_b = \left\{ \omega : \lim_{n\to\infty} \frac{f^s_{b,0}(\omega^{(n)})}{f^s_b(\omega^{(n)})} = 1 - \overline{p} \right\}.$$

Consider $\omega \in N_{a,b} \cap M_a \cap M_b$. Then $\log \nu_{n+1} - \log \nu_1 =$

$$\begin{aligned} &- f^s_a \left[ \underline{p} \log \left( \frac{\overline{p}}{\underline{p}} \right) + (1 - \underline{p}) \log \left( \frac{1 - \overline{p}}{1 - \underline{p}} \right) \right] \\ &- f^s_b \left[ \overline{p} \log \left( \frac{\underline{p}}{\overline{p}} \right) + (1 - \overline{p}) \log \left( \frac{1 - \underline{p}}{1 - \overline{p}} \right) \right] \\ &\equiv -f^s_a H_1 - f^s_b H_2. \end{aligned}$$

By the concavity of log, both $H_1$ and $H_2$ are negative. Therefore, $\nu_n \to \infty$, equivalently $\mu_n \to 1$, on $N_{a,b} \cap M_a \cap M_b$. By the LLN, $Q^s(N_{a,b} \cap M_a \cap M_b) = Q^s(N_{a,b})$. Conclude that

$$Q^s \left( N_{a,b} \cap \{\omega : \mu_n \to 1\} \right) = Q^s \left( N_{a,b} \right).$$

Similar equations apply if $N_{a,b}$ is replaced by either $N_a$ or $N_b$. Finally, since $\{N_a, N_b, N_{a,b}\}$ is a partition of $\Omega$, conclude that $Q^s(\{\omega : \mu_n \to 1\}) = 1$.
(iii) Prove (2.7). By (i), $P^s(C_s) = 1$ for every $s \in \mathcal{S}$, where

$$C_s = \left\{ \omega \in \Omega : \{ \lim_{n\to\infty} E_{P^a_n}[X^2_n | a^{(n-1)}_s, \omega^{(n-1)}](\omega), \lim_{n\to\infty} E_{P^b_n}[X^2_n | a^{(n-1)}_s, \omega^{(n-1)}](\omega) \} = \{\overline{p}, \underline{p}\} \right\}.$$

where $a^{(n-1)}_s$ is given by (2.3) via strategy $s$. Define

$$A_s = \{\omega \in \Omega : \lim_{n\to\infty} ess \sup_{s' \in \mathcal{S}(s,n-1)} E_{P^{s'}}[X^2_n | \mathcal{G}_{n-1}](\omega) = \overline{p}\}$$
$$B_s = \{\omega \in \Omega : \lim_{n\to\infty} ess \inf_{s' \in \mathcal{S}(s,n-1)} E_{P^{s'}}[X^2_n | \mathcal{G}_{n-1}](\omega) = \underline{p}\}$$

and observe that $C_s \subset A_s, C_s \subset B_s$, and hence that $P^s(A_s) = P^s(B_s) = 1$. This proves (2.7) with $(\underline{\sigma}^2, \overline{\sigma}^2) = (\underline{p}, \overline{p})$.

The bound $\overline{var}$ required by Assumption-Variance can be taken to be $\overline{p}$. ■

**Proof of Theorem** 2.4:
Step 1: For $n \geq 1$, define

$$\underline{M}_n = \min\{\mu_n, 1 - \mu_n\}, \ \overline{M}_n = \max\{\mu_n, 1 - \mu_n\}$$

Then, by the dominated convergence theorem and Lemma 2.3(i),

$$\lim_{n\to\infty} E_{P^{s^*}}[\underline{M}_n] = E_{P^{s^*}}\left[ \lim_{n\to\infty} \underline{M}_n \right] = E_{P^{s^*}}[\min\{\mu, 1 - \mu\}] = 0,$$

where $\mu$ is defined by (B.5).

For small enough $h > 0$, let $\{H_t\}_{t\in[0,1+h]}$ be the functions defined in (A.3), and let $\{L_{m,n}\}_{m=1}^n$ be the functions defined in (A.4). We prove below that

$$\lim_{n\to\infty}\sum_{m=1}^{n}\left|E_{P^{s^*}}\left[H_{\frac{m}{n}}\left(\frac{\Sigma_1^m X_i}{\sqrt{n}}\right)\right]-E_{P^{s^*}}\left[L_{m,n}\left(\frac{\Sigma_1^{m-1}X_i}{\sqrt{n}}\right)\right]\right|=0. \tag{B.6}$$

This is the counterpart for the present setting of the limit result (A.5) in the proof of our CLT (Lemma A.2), where instead of the expectation with respect to the single measure $P^{s^*}$, one has the upper expectation $\mathbb{E}$ corresponding to the set of measures $\mathcal{P}$. The proof of (B.6) roughly parallels the earlier arguments but the difference between $E_{P^{s^*}}$ and $\mathbb{E}$ necessitates some adjustments (notably in Step 3).

Define

$$d(m,n)=E_{P^{s^*}}\left[H_{\frac{m}{n}}\left(\frac{\Sigma_1^{m-1}X_i}{\sqrt{n}}\right)+H'_{\frac{m}{n}}\left(\frac{\Sigma_1^{m-1}X_i}{\sqrt{n}}\right)\frac{X_m}{\sqrt{n}}+H''_{\frac{m}{n}}\left(\frac{\Sigma_1^{m-1}X_i}{\sqrt{n}}\right)\frac{X_m^2}{2n}\right].$$

It suffices for (B.6) to prove that

$$\sum_{m=1}^{n}\left|E_{P^{s^*}}\left[H_{\frac{m}{n}}\left(\frac{\Sigma_1^m X_i}{\sqrt{n}}\right)\right]-d(m,n)\right|\to 0 \quad \text{and} \tag{B.7}$$

$$\sum_{m=1}^{n}\left|d(m,n)-E_{P^{s^*}}\left[L_{m,n}\left(\frac{\Sigma_1^{m-1}X_i}{\sqrt{n}}\right)\right]\right|\longrightarrow 0. \tag{B.8}$$

Step 2: Prove (B.7). The argument is similar to that for (A.6).

Step 3: Prove (B.8). By (2.29), for any $m \geq 1$, $E_{P^{s^*}}[X_m|\mathcal{G}_{m-1}] = 0$, and

$$E_{P^{s^*}}[X_m^2|\mathcal{G}_{m-1}]=\begin{cases}\overline{\sigma}^2\overline{M}_m+\underline{\sigma}^2\underline{M}_m & \text{if} \quad \Sigma_1^{m-1}X_i\leq 0\\ \\ \underline{\sigma}^2\overline{M}_m+\overline{\sigma}^2\underline{M}_m & \text{if} \quad \Sigma_1^{m-1}X_i>0\end{cases} \tag{B.9}$$

Therefore, for $C_1$ equal to the uniform bounded of $|H_t^{''}(x)|$,

$$\begin{aligned}&\sum_{m=1}^{n}\left|d(m,n)-E_{P^{s^*}}\left[L_{m,n}\left(\tfrac{\Sigma_1^{m-1}X_i}{\sqrt{n}}\right)\right]\right|\\ \leq&\sum_{m=1}^{n}E_{P^{s^*}}\left[\frac{1}{2n}\left[H''_{\frac{m}{n}}\left(\tfrac{\Sigma_1^{m-1}X_i}{\sqrt{n}}\right)\right]^+\left(\overline{\sigma}^2-\overline{\sigma}^2\overline{M}_m-\underline{\sigma}^2\underline{M}_m\right)\right]\\ &+\sum_{m}^{n}E_{P^{s^*}}\left[\frac{1}{2n}\left[H''_{\frac{m}{n}}\left(\tfrac{\Sigma_1^{m-1}X_i}{\sqrt{n}}\right)\right]^-\left(\underline{\sigma}^2\overline{M}_m+\overline{\sigma}^2\underline{M}_m-\underline{\sigma}^2\right)\right]\\ \leq&\frac{C_1(\overline{\sigma}^2-\underline{\sigma}^2)}{n}\sum_{m=1}^{n}E_{P^{s^*}}\left[\underline{M}_m\right]\ \longrightarrow 0 \text{ (by Step 1).}\end{aligned}$$

**Remark B.1.** *Step 3 involves a departure from the arguments of the CLT. In the latter, we had by assumption (3.4) that upper and lower conditional variances converge to $\overline{\sigma}^2$ and $\underline{\sigma}^2$ respectively, while here the relevant conditional variances are under $P^{s^*}$ and are shown in (B.9). Also noteworthy is that, while preceding steps in the argument are valid for all strategies $s$, Steps 3 and 4 rely explicitly on $s = s^*$.*

Step 4: Complete the proof. It can be checked that

$$\begin{aligned}
&E_{P^{s^*}}\left[H_1\left(\frac{\Sigma_1^n X_i}{\sqrt{n}}\right)\right] - H_0(0) \\
=&\sum_{m=1}^{n}\left\{E_{P^{s^*}}\left[H_{\frac{m}{n}}\left(\frac{\Sigma_1^m X_i}{\sqrt{n}}\right)\right] - E_{P^{s^*}}\left[H_{\frac{m-1}{n}}\left(\frac{\Sigma_1^{m-1} X_i}{\sqrt{n}}\right)\right]\right\} \\
=&\sum_{m=1}^{n}\left\{E_{P^{s^*}}\left[H_{\frac{m}{n}}\left(\frac{\Sigma_1^m X_i}{\sqrt{n}}\right)\right] - E_{P^{s^*}}\left[L_{m,n}\left(\frac{\Sigma_1^{m-1} X_i}{\sqrt{n}}\right)\right]\right\} \\
&+\sum_{m=1}^{n}\left\{E_{P^{s^*}}\left[L_{m,n}\left(\frac{\Sigma_1^{m-1} X_i}{\sqrt{n}}\right)\right] - E_{P^{s^*}}\left[H_{\frac{m-1}{n}}\left(\frac{\Sigma_1^{m-1} X_i}{\sqrt{n}}\right)\right]\right\} \\
=&: \hat{J}_{1n} + \hat{J}_{2n}.
\end{aligned}$$

By (B.6), we have $\lim_{n\to\infty}|\hat{J}_{1n}| = 0$. By Lemma A.1(5), (argue as in the proof that $|I_{2n}| \to 0$ in Appendix A.2), we have $\lim_{n\to\infty}|\hat{J}_{2n}| = 0$. Therefore,

$$\lim_{n\to\infty}\left|E_{P^{s^*}}\left[H_1\left(\frac{\Sigma_1^n X_i}{\sqrt{n}}\right)\right] - H_0(0)\right| = 0.$$

By the definition of functions $\{H_t\}$, with arguments similar to those at the end of Appendix A.2, we have

$$\left|E_{P^{s^*}}\left[\varphi\left(\frac{\Sigma_1^n X_i}{\sqrt{n}}\right)\right] - E_{P^*}[\varphi(W_1^0)]\right| \to 0 \quad \text{as} \quad n \to \infty. \qquad \blacksquare$$

# Acknowledgements

We have benefited from comments by Peter Wakker, an editor (Marzena Rostek) and two referees, one of whom, Luciano Pomatto, led us to recognize a significant error in an earlier version. Chen gratefully acknowledges the support of the National Key R&D Program of China (grant No. ZR2019ZD41), and the Taishan Scholars Project. Zhang gratefully acknowledges the support of Shandong Provincial Natural Science Foundation, China (grants ZR2022QA063 and ZR2021MA098).

# References


[1] Banks, J. and Sundaram, R.K. (1992). A class of bandit problems yielding myopic optimal strategies. *J. Appl. Probab.* (1992), 625-632.

[2] Barberis, N.C. (2013). Thirty years of prospect theory in economics: a review and assessment. *J. Econ. Persp.* 27, 173-196.

[3] Bergemann, D. and Välimäki, J. (2008). Bandit problems. In Palgrave Macmillan (eds.) *The New Palgrave Dictionary of Economics*. Palgrave Macmillan, London.

[4] Berry, D. and Fristedt, B. (1985). *Bandit Problems*. Chapman Hall, London.

[5] Bradt, R. N., Johnson, S. M., and Karlin, S. (1956). On sequential designs for maximizing the sum of n observations. *Ann. Math. Stat.* 27(4), 1060-1074.

[6] Chen, Z. and Epstein, L. G. (2022). A central limit theorem for sets of probability measures. *Stoch. Process. Appl.* 152, 424-451.

[7] Chen, Z., Epstein, L. G., and Zhang, G. (2021). A central limit theorem, loss aversion and multi-armed bandits. arXiv preprint arXiv:2106.05472v1.

[8] Chen, Z. and Zili, M. (2015). One-dimensional heat equation with discontinuous conductance. *Science China Math.* 58(1), 97-108.

[9] De Finetti, B. (1938). English translation is "On the condition of partial exchangeability." In R. Jeffrey (ed.) *Studies in Inductive Logic and Probability*, vol. 2. 1980, U. California Press, Berkeley.

[10] Diaconis, P. and Freedman, D. (1982). Partial exchangeability and sufficiency. Tech Report 190, Statistics Department, Stanford University.

[11] Dreze, J. (1987). Decision theory with moral hazard and state-dependent preference. pp. 23-89 in J. Dreze (ed.) *Essays on Economic Decisions under Uncertainty*. Cambridge U. Press, Cambridge.

[12] Easley, D. and Yang, L. (2015). Loss aversion, survival and asset prices. *J. Econ. Theory* 160, 494-516.

[13] Ebert, S. and Strack, P. (2015). Until the bitter end: On prospect theory in a dynamic context. *Amer. Econ. Rev.* 105, 1618-1633.

[14] Epstein, L.G., Kaido, H. and Seo, K. (2016). Robust confidence regions for incomplete models. *Econometrica* 84, 1799-1838.

[15] Epstein, L.G. and Schneider, M. (2003). Recursive multiple-priors. *J. Econ. Theory* 113, 1-31.

[16] Fang, X., Peng, S., Shao, Q. M., and Song, Y. (2019). Limit theorems with rate of convergence under sublinear expectations. *Bernoulli* 25(4A), 2564-2596.

[17] Feldman, D. (1962). Contributions to the "two-armed bandit" problem. *Ann. Math. Statist.* 33, 847-856.

[18] Gittins, J. and Jones, D. (1974). A dynamic allocation index for the sequential allocation of experiments. In J. Gani (ed.) *Progress in Statistics*. North-Holland, Amsterdam.

[19] Guasoni, P., Huberman, G., and Ren, D. (2020). Shortfall aversion. *Math. Finan.* 30, 869-920.

[20] Hirano, K. and Porter, J.R. (2020). Asymptotic analysis of statistical decision rules in econometrics. pp. 283-354 in S.N. Durlauf, L.P. Hansen, J.J. Heckman, and R.L. Matzkin eds. *Handbook of Econometrics* Vol. 7A. Elsevier.

[21] Huo, X. and Fu, F. (2017). Risk-aware multi-armed bandit problem with application to portfolio selection. *Royal Soc.open sci.* 4:171377.

[22] Kahneman D. and Tversky, A. (2000). *Choices, Values and Frames*. Cambridge U. Press, N.Y.

[23] Karni, E. (2011). A theory of Bayesian decision making with action dependent subjective probabilities. *Econ. Theory* 48, 125-146.

[24] Kallenberg, O. (2005). *Probabilistic Symmetries and Invariance Principles*. Springer, N.Y.

[25] Keilson, J. and Wellner, J. A. (1978). Oscillating Brownian motion. *J. Appl. Probab.* 15(2), 300-310.

[26] Kelsey, D. and Milne, F. (1999). Induced preferences, nonadditive beliefs, and multiple priors. *Intern. Econ. Rev.* 40, 455-477.

[27] Kobberling, V. and Wakker, P.P. (2005). An index of loss aversion. *J. Econ. Theory* 122, 119-131.

[28] Le Gall, J. F. (1984). One-dimensional stochastic differential equations involving the local times of the unknown process. In A. Taubman and D. Williams (eds.) *Stochastic Analysis and Applications* (pp. 51-82), LNM vol 1095. Springer, Berlin.

[29] Lejay, A. and Pigato, P. (2018). Statistical estimation of the oscillating Brownian motion. *Bernoulli* 24(4B), 3568-3602.

[30] Link, G. (1980). Representation theorems of the de Finetti type for (partially) symmetric probability measures. In R. Jeffrey (ed.) *Studies in Inductive Logic and Probability*, vol. 2. U. California Press, Berkeley.

[31] Marinacci, M. (1999). Limit laws for non-additive probabilities and their frequentist interpretation. *J. Econ. Theory* 84, 145-195.

[32] Peng, S. (2007). G-expectation, G-Brownian motion and related stochastic calculus of Itô type. In: Benth, F.E., Di Nunno, G., Lindstrøm, T., Øksendal, B., Zhang, T. (eds) *Stoch. Anal. Appl.* Abel Symposia, vol 2. Springer, Berlin, https://doi.org/10.1007/978-3-540-70847-6_25.

[33] Peng, S. (2019). *Nonlinear Expectations and Stochastic Calculus under Uncertainty: with Robust CLT and G-Brownian Motion.* Springer Nature.

[34] Rothschild, M. (1974). A two-armed bandit theory of market pricing. *J. Econ. Theory* 9, 185-202.

[35] Sani, A., Lazaric, A. and Munos, R. (2013). Risk-aversion in multi-armed bandits. arXiv:1301.1936v1 [cs.LG]

[36] Shi, Y., Cui, X., Yao, J., and Li, D. (2015). Dynamic trading with reference point adaptation and loss aversion. *Oper. Res.* 63, 789-806.

[37] Slivkins, A. (2019). Introduction to multi-armed bandits. *Foundations and Trends®in Machine Learning*, 12 (1-2), 1-286 http://dx.doi.org/10.1561/220000006.

[38] Tversky, A. and Kahneman, D. (1992). Advances in prospect theory: cumulative representation of uncertainty. *J. Risk Uncert.* 5, 297-323.

[39] Wakker, P. P. and Tversky, A. (1993). An axiomatization of cumulative prospect theory. *J. Risk Uncert.* 7, 147-176.

[40] Xu, Z.Q. and Zhou, X.Y. (2013). Optimal stopping under probability distortion. *Ann. Appl. Probab.* 23, 251-282.