Sequential Methods with Martingales

Liam Donovan
Department of Mathematics, University of Massachusetts Amherst

Introduction

Sequential analysis studies hypothesis testing that allow decisions without fixing the sample size in advance. These methods often lead to earlier stopping and greater efficiency.

We focus on procedures based on **likelihood ratios**, which:

- Are updated in time as new data arrives
- Satisfy the Markov property
- Form **nonnegative martingales** under the null hypothesis

This martingale structure allows us to:

- Derive valid stopping rules
- Control type I error via optional stopping and Ville's Inequality

We concern ourselves with simple null hypotheses, which are often used in practice. We demonstrate and compare these methods across simulated Bernoulli trials.

Simple Method: Sequential Probability Ratio Test (SPRT)

Suppose data arrive one at a time and we wish to test, for a parameter, θ

$$H_0 : \theta = \theta_0 \quad \text{vs.} \quad H_1 : \theta = \theta_1 \quad \big(X_t \sim P_\theta\big).$$

as data arrives one sample at a time.

Let $\{X_t\}_{t\geq 1}$ be independent observations. Define the **likelihood ratio process**:

$$L_t = \prod_{i=1}^{t} \frac{P_{\theta_1}(X_i)}{P_{\theta_0}(X_i)}$$

This process is the sequential analogue of the classical Likelihood Ratio Test.

- Under simple H_0, $\{L_t\}_{t\geq 1}$ is a **nonnegative martingale** with $\mathbb{E}_0[L_t] = 1$.
  - Thus, by the **Optional Stopping Theorem**, if H_0 is true:
  $$E_0(L_\tau) = E_0(L_0) = 1.$$
  for any stopping time τ.
- Intuitively, L_t grows when H_1 holds—so we stop once it's "too big."
- Further, $1/L_t$ is a martingale under H_1.

From here, one may apply the OST to each and derive the following decision rule. Define the stopping time

$$\tau = \inf\{\, t : L_t \geq A \text{ or } L_t \leq B \}.$$

and let

$$A = \frac{1-\beta}{\alpha}, \quad B = \frac{\beta}{1-\alpha}.$$

where α, β are the type-I and type-II error rates, respectively.

Decision rule:

- If $L_\tau \geq A$, reject H_0.
- If $L_\tau \leq B$, fail to reject H_0.

This is called **Wald's Sequential Probability Ratio Test (SPRT)**.

Robust Tests for Composite Hypotheses

When the alternative hypothesis is composite, i.e. $\theta \in \Theta_1$, the likelihood ratio is typically no longer a martingale under H_0, but instead a supermartingale. This complicates type I error control. We present two approaches that restore the martingale property under the simple null $\theta = \theta_0$

1.) Bayesian Methods: Mixture Martingales

Take a prior over the chosen parameter: $\pi(\theta)$ and compute the average over the alternate possibilities. Computing the posterior for θ:

$$L_t^{(\text{mix})} = \int_{\Theta_1} \prod_{i=1}^{t} \frac{P_\theta(X_i)}{P_{\theta_0}(X_i)} \pi(\theta)\, d\theta.$$

- For conjugate priors, this is tractable
- For others, this integral is not tractable. One could use:
  - Discretization: take a grid of possible θ
  - Numerical methods: quadrature, sampling methods, Monte Carlo integration
- **Challenge:** Computationally expensive.

This is a nonnegative martingales, under H_0 for *any choice of prior*.

2.) CGF-Tilted Martingales

One can derive a method from the desired multiplicity of the likelihood ratios and the moment generating function:

$$T_t(\lambda) = \exp\left(\lambda \sum_{i=1}^{t} X_i - t\,\Lambda(\lambda)\right)$$

where $\Lambda(\lambda)$ denotes the **cumulant generating function (CGF)**:

$$\Lambda(\lambda) = \log E\left(e^{X\lambda}\right)$$

, i.e., the log of the moment generating function. with chosen parameter λ. This is also a nonnegative martingale under H_0.

- **Challenge:** Requires "nice" moment generating function and requires tuning λ.

There does not exist a simple associated martingale under H_1 for either of these martingales, so one may create a stopping rule using **Ville's inequality**:

$$P_0\left(\sup_t L_t^{(\text{mix})} \geq \frac{1}{\alpha}\right) \leq \alpha$$

$$P_0\left(\sup_t T_t \geq \frac{1}{\alpha}\right) \leq \alpha$$

Limitation: Slow Detection for Small Effects

Let $\theta = \theta_0 + \delta$, where δ is the **effect size** (deviation from null)

Problem: When δ is small"

- A fixed-λ CGF martingale tuned to δ needs many samples to reject with high probability.
- Mixture martingales (integrating over λ) guard uniformly over δ, but the likelihood ratio will be about 1 each step leading to slow convergence.

Stitching Martingales

Solution for small effects: Construct a family of CGF-tilted martingales $M_t(\lambda_j) = \exp(\lambda_j S_t - t\Lambda(\lambda_j))$ for a grid $\{\lambda_j\}$. Form the mixture

$$M_t^{\text{stitch}} = \sum_j w_j\, M_t(\lambda_j), \quad \sum_j w_j = 1.$$

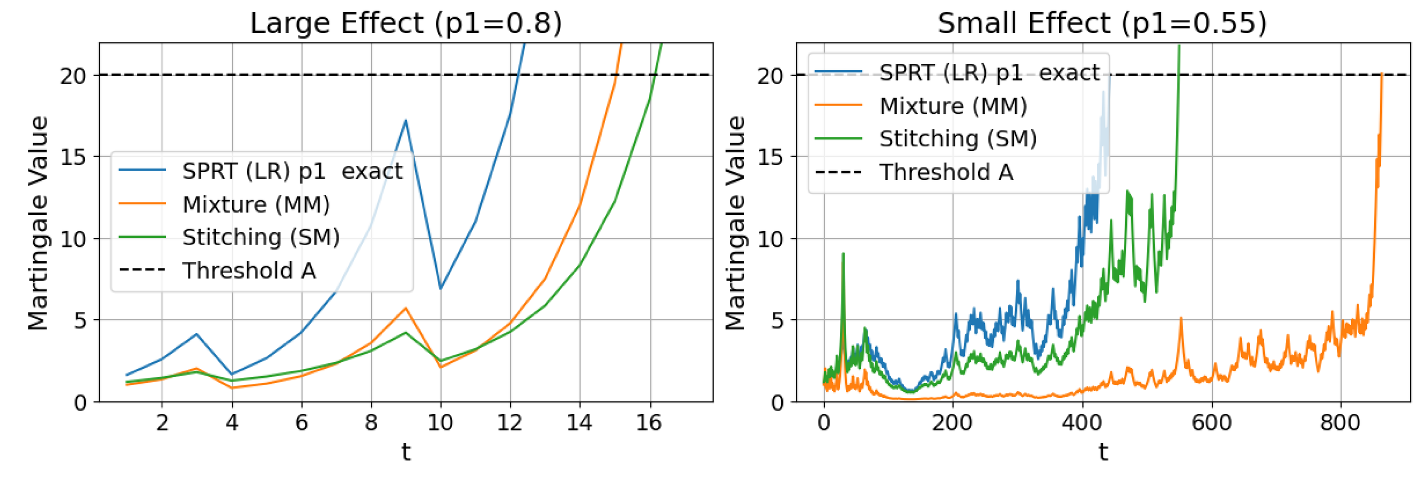
By linearity, $E_0[M_t^{\text{stitch}}] = 1$, so Ville's inequality gives $P_0(\sup_t M_t^{\text{stitch}} \geq 1/\alpha) \leq \alpha$.

Empirical Comparison of Methods

Example: Bernoulli trials

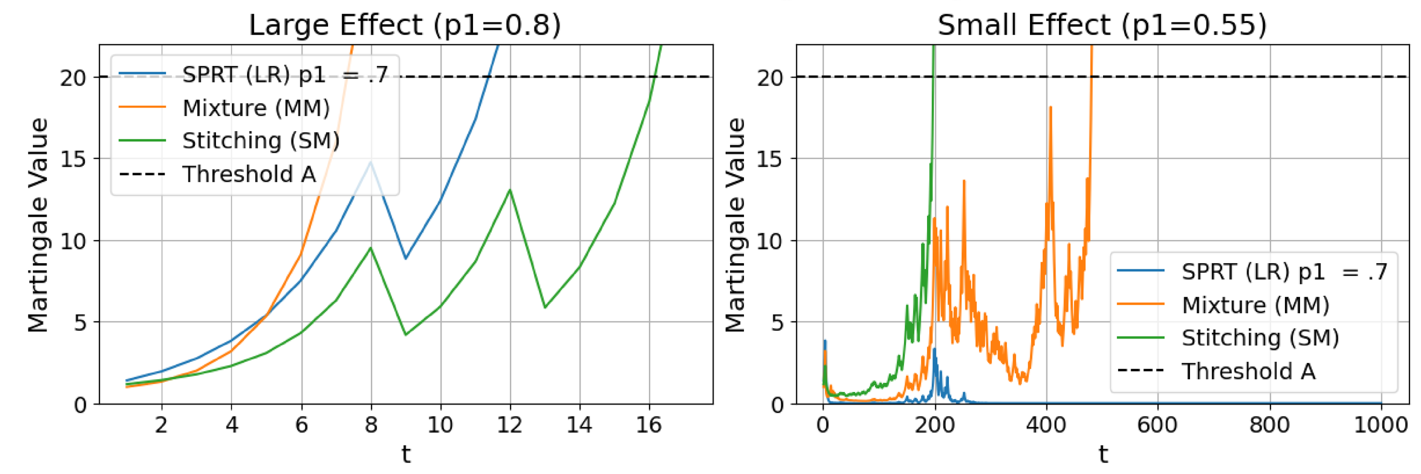
- Uninformative ($\beta(1,1)$) **uniform prior** for mixture.
- 25 λ values from 0.1 to 2.5 spaced evenly for stitching. Weighting scheme favors **smaller** values.
- All computed at 5% confidence level; $A = 1/\alpha = 20$
- We compare when SPRT uses exact p_1 value and when SPRT is not exact:

Exact SPRT (well-specified)



Exact SPRT converges fastest. Stitching martingale lags slightly under small effects.

SPRT with misspecified p_1



Misspecification of SPRT slows convergence dramatically. Stitching remains robust. Note: the scale of the x-axis differs—large effects require far fewer samples, so differences are magnified.

References

- Wald (1947), *Sequential Analysis*
- Ville (1939), *Étude Critique de la Notion de Collectif*
- Howard et al. (2021), *Time-uniform Chernoff bounds via nonnegative supermartingales.*, 18