Discussion of Aaditya Ramdas’ talk *Ville’s inequality, confidence sequences and test supermartingales*

Wouter M. Koolen

CWI
Centrum Wiskunde & Informatica
An elegant framework emerges for the construction of anytime confidence sequences.
Praise

An elegant framework emerges for the construction of anytime confidence sequences.

Empowering theoreticians and practitioners.
An elegant framework emerges for the construction of anytime confidence sequences.

Empowering theoreticians and practitioners.

We need principled (safe) statistics in practice.
Consider a property $\phi$ defined on a class of distributions $\mathcal{P}$. How to build confidence sequence for $\phi$?

**Definition**

$M^P_n$ is a test supermartingale for $P \in \mathcal{P}$ if $M^P_0 = 1$, $M^P_n \geq 0$ and $E^P [M^P_{n+1} | F_n] \leq M^P_n$.

**Idea:** $M^P_n$ is evidence against $\{P\}$.

But what about $\{\phi(P) = \nu\}$?

**Definition**

$M^\nu_n$ is a simultaneous test supermartingale for $P^\nu := \{P \in \mathcal{P} | \phi(P) = \nu\}$ if it is a test supermartingale for each $P \in P^\nu$.

**Idea:** $M^\nu_n$ is evidence against $\{\phi(P) = \nu\}$.
Consider a property $\phi$ defined on a class of distributions $\mathcal{P}$. How to build confidence sequence for $\phi$?

**Definition**

$M_n^P$ is a **test supermartingale** for $P \in \mathcal{P}$ if $M_0^P = 1$, $M_n^P \geq 0$ and

$$E_P \left[ M_{n+1}^P \mid \mathcal{F}_n \right] \leq M_n^P$$

Idea: $M_n^P$ is evidence against $\{P\}$.

**Definition**

$M_n^\nu$ is a **simultaneous test supermartingale** [Vovk et al. 2013] for $P_\nu := \{P \in \mathcal{P} | \phi(P) = \nu\}$ if it is a test supermartingale for each $P \in P_\nu$.

Idea: $M_n^\nu$ is evidence against $\{\phi(P) = \nu\}$. 
Q1: A Fine Distinction (a Confession)

Consider a property \( \phi \) defined on a class of distributions \( \mathcal{P} \). How to build confidence sequence for \( \phi \)?

**Definition**

\( M_n^P \) is a **test supermartingale** for \( P \in \mathcal{P} \) if \( M_0^P = 1, M_n^P \geq 0 \) and

\[
\mathbb{E}_P \left[ M_{n+1}^P \big| \mathcal{F}_n \right] \leq M_n^P
\]

Idea: \( M_n^P \) is **evidence against** \( \{P\} \). But what about \( \{\phi(P) = \nu\} \)?

Wouter Koolen  
Discussion: Ville, confidence, martingale  
Selective Inference
Consider a property \( \phi \) defined on a class of distributions \( \mathcal{P} \). How to build confidence sequence for \( \phi \)?

**Definition**

\( M^P_n \) is a **test supermartingale** for \( P \in \mathcal{P} \) if \( M^P_0 = 1, \ M^P_n \geq 0 \) and

\[
\mathbb{E}_P \left[ M^P_{n+1} \mid \mathcal{F}_n \right] \leq M^P_n
\]

Idea: \( M^P_n \) is **evidence against** \( \{P\} \). But what about \( \{\phi(P) = \nu\} \)?

**Definition**

\( M^\nu_n \) is a **simultaneous test supermartingale** [Vovk et al. 2013] for \( \mathcal{P}_\nu := \{P \in \mathcal{P} | \phi(P) = \nu\} \) if it is a test supermartingale for each \( P \in \mathcal{P}_\nu \).

Idea: \( M^\nu_n \) is **evidence against** \( \{\phi(P) = \nu\} \).
Many martingales considered in the talk/papers are *simultaneous* test supermartingales.
Many martingales considered in the talk/papers are **simultaneous** test supermartingales.

For example the sub-Gaussian $M_n = e^{\lambda S_n - n\lambda^2/2}$, moment-constrained, symmetric, products of “Safe Tests”, etc.
Many martingales considered in the talk/papers are **simultaneous** test supermartingales.

For example the sub-Gaussian \( M_n = e^{\lambda S_n - n\lambda^2/2} \), moment-constrained, symmetric, products of “Safe Tests”, etc.

So simultaneous test supermartingales are **powerful/useful**.
Many martingales considered in the talk/papers are **simultaneous** test supermartingales.

For example the sub-Gaussian $M_n = e^{\lambda S_n - n\lambda^2/2}$, moment-constrained, symmetric, products of “Safe Tests”, etc.

So simultaneous test supermartingales are **powerful/useful**.

Yet the only simultaneous supermartingale for the (parametric!) model \{i.i.d. Bernoulli($\theta$)|$\theta \in [0, 1]$\} is the **trivial** $M_n = 1$. 
Q1: A Fine Distinction (a Confession) ctd.

Many martingales considered in the talk/papers are **simultaneous** test supermartingales.

For example the sub-Gaussian $M_n = e^{\lambda S_n - n\lambda^2/2}$, moment-constrained, symmetric, products of “Safe Tests”, etc.

So simultaneous test supermartingales are **powerful/useful**.

Yet the only simultaneous supermartingale for the (parametric!) model \{i.i.d. Bernoulli($\theta$)|$\theta \in [0, 1]$\} is the **trivial** $M_n = 1$.

So one cannot disqualify i.i.d. Bernoulli? (E.g. using a time series)?
Many martingales considered in the talk/papers are **simultaneous** test supermartingales.

For example the sub-Gaussian $M_n = e^{\lambda S_n - n\lambda^2/2}$, moment-constrained, symmetric, products of “Safe Tests”, etc.

So simultaneous test supermartingales are **powerful/useful**.

Yet the only simultaneous supermartingale for the (parametric!) model \{i.i.d. Bernoulli($\theta$)$|\theta \in [0, 1]$\} is the **trivial** $M_n = 1$.

So one cannot disqualify i.i.d. Bernoulli? (E.g. using a time series)?

Yes **of course** one can. But not with a **simultaneous** test supermartingale.
Q1: A Fine Distinction (a Confession) ctd.

Many martingales considered in the talk/papers are simultaneous test supermartingales.

For example the sub-Gaussian $M_n = e^{\lambda S_n - n\lambda^2/2}$, moment-constrained, symmetric, products of “Safe Tests”, etc.

So simultaneous test supermartingales are powerful/useful.

Yet the only simultaneous supermartingale for the (parametric!) model \{i.i.d. Bernoulli($\theta$)|$\theta \in [0, 1]$\} is the trivial $M_n = 1$.

So one cannot disqualify i.i.d. Bernoulli? (E.g. using a time series)?

Yes of course one can. But not with a simultaneous test supermartingale.

Q: power and limits of simultaneous supermartingales. Orthogonality?
Q2: A converse

Consider a property $\phi$ defined on a class of distributions $\mathcal{P}$.
Q2: A converse

Consider a property $\phi$ defined on a class of distributions $\mathcal{P}$.

Definition
A random process $C_n$ is an $\alpha$-confidence sequence if

$$\forall P \in \mathcal{P}: \quad P(\forall n: \phi(P) \in C_n) \geq 1 - \alpha$$
Q2: A converse

Consider a property $\phi$ defined on a class of distributions $\mathcal{P}$.

**Definition**

A random process $C_n$ is an $\alpha$-confidence sequence if

$$\forall P \in \mathcal{P} : \quad P(\forall n : \phi(P) \in C_n) \geq 1 - \alpha$$

Q: can we **always** obtain $C_n$ using Ville?

**Conjecture (Universality)**

For each $\alpha$-confidence sequence $C_n$ there is a family of test martingales $\{M_P^n | P \in \mathcal{P}\}$ such that $C_n \supseteq \{\phi(P) | P \in \mathcal{P} \text{ and } M_P^n \leq 1/\alpha\}$.
Q2: A converse

Consider a property $\phi$ defined on a class of distributions $\mathcal{P}$.

**Definition**

A random process $C_n$ is an $\alpha$-confidence sequence if

$$\forall P \in \mathcal{P} : \quad P \left( \forall n : \phi(P) \in C_n \right) \geq 1 - \alpha$$

Q: can we *always* obtain $C_n$ using Ville?

**Conjecture (Universality)**

For each $\alpha$-confidence sequence $C_n$ there is a family of test martingales $\left\{ M_n^P \mid P \in \mathcal{P} \right\}$ such that

$$C_n \supseteq \left\{ \phi(P) \mid P \in \mathcal{P} \text{ and } M_n^P \leq 1/\alpha \right\}.$$
Q2: A converse

This forward “constructive” direction

\[ C_n := \left\{ \phi(P) \mid P \in \mathcal{P} \text{ and } M_n^P \leq 1/\alpha \right\} \]

raises the design question:
This forward “constructive” direction

\[ C_n := \left\{ \phi(P) \mid P \in \mathcal{P} \text{ and } M_n^P \leq 1/\alpha \right\} \]

raises the design question:

Q: how to craft \( \{ M_n^P \mid P \in \mathcal{P} \} \) if you are interested in \( \phi \) on \( \mathcal{P} \)?
Q2: A converse

This forward “constructive” direction

\[ C_n := \left\{ \phi(P) \middle| P \in \mathcal{P} \text{ and } M_n^P \leq \frac{1}{\alpha} \right\} \]

raises the design question:

Q: how to craft \( \{ M_n^P \middle| P \in \mathcal{P} \} \) if you are interested in \( \phi \) on \( \mathcal{P} \)?

Sequential GROW criterion [Grünwald, De Heide, Koolen, 2019]?
What is the “lego” of test supermartingales? We saw constructions of the “sub-parametric” form

\[ e^{\lambda S_n - \psi(\lambda) V_n} \]

and moment constrained form [Agrawal, Juneja, Glynn, ALT 2020]

\[ \prod_{t=1}^{n} \left( 1 + \lambda_1 (X_t - \mu) + \lambda_2 (X_t^2 - b) \right) \]

What else is out there? Are all forms extremal likelihood ratios? Can we do tight stitching efficiently \textbf{in software}?