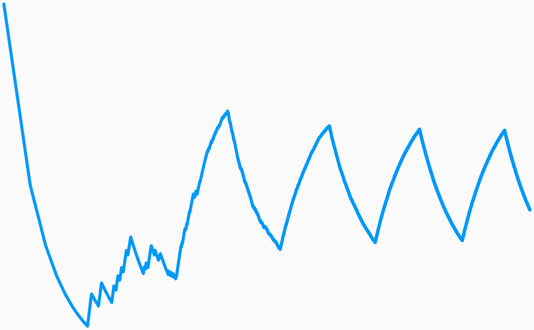
A Composite Ville's Inequality and SLLNs

Wouter M. Koolen

INRIA Seminar, Sep 5, 2022

CWI and University of Twente



Team Effort

Johannes Ruf

Martin Larsson

Aaditya Ramdas

Menu

Introduction

Starting point: Ville's Result

Fix filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_n)_n, P)$ and (tail) event $A \in \mathcal{F}$. Then

Theorem

$P(A) = 0$ **iff** *there is a test martingale* $(M_n)_n$ *for* P *tending to* ∞ *on* A*.*

Why is that useful

If P is true and $P(A) = 0$ then A will not happen.

A martingale $(M_n)_n$ diverging on A allows detecting A in finite time with Type-1 error control.

To do so, we declare A is happening at $\tau = \inf\{n \mid M_n \geq 1/\alpha\}$.

- If A happens, we detect it for sure
- Under P, probability of (false) detection is $\leq \alpha$.

Definition

A stopping time τ detects the event $A \in \mathcal{F}$ in finite time with confidence α if $A \subseteq \{\tau < \infty\}$ and $P\{\tau < \infty\} \leq \alpha$.

Examples of Ville's result

For i.i.d. Bernoulli(1/2), can detect in finite time:

- SLLN: running average converges to 1/2
- LIL: deviations from 1/2 of order $\sqrt{\frac{\ln \ln n}{n}}$
- ...

Composite Upgrade

Question

Suppose we face i.i.d. Bernoulli(p) for unknown $p \in [0, 1]$.

The running average of outcomes will converge.

But can we detect divergence in finite time?

For each fixed p there is a divergence detector.

But it may never happen that all detectors fire.

First attempt

We are interested in a composite null $\mathcal{P}$.

We say event A is $\mathcal{P}$-polar if $P(A) = 0$ for all $P \in \mathcal{P}$.

Maybe

Conjecture

A is $\mathcal{P}$-polar **iff** *there is a $\mathcal{P}$-test (super)martingale tending to ∞ on A.*

But the universe says no.

Fundamental Problems

Some $\mathcal{P}$-polar sets

- *are* finite-time detectable, but not by $\mathcal{P}$-(super)martingales.
- *are not* finite-time detectable at all.

Example

Let $\mathcal{P} = \{P_n \mid n \in \mathbb{N}\}$, where P_n deterministically outputs sequence $0^n \cdot 1 \cdot 0^{\mathbb{N}}$

Let $A = \{0^{\mathbb{N}}\}$, so A is $\mathcal{P}$-polar.

Any purported finite-time detection of A must occur seeing some prefix 0^m.

But then $P_n(\text{detection}) = 1$ for all $n \geq m$. No error control

Successful attempt

Definition

Non-negative $(E_n)_n$ is an E-process for $\mathcal{P}$ if

$$\mathbb{E}_P\left[E_\tau\right] \leq 1 \quad \text{for any } P \in \mathcal{P} \text{ and stopping time } \tau$$

Let's define the minimax finite-time detection probability:

Definition

$$\mu^*(A) \;=\; \inf_{\tau \in \mathcal{T} : A \subseteq \{\tau < \infty\}} \sup_{P \in \mathcal{P}} P(\tau < \infty)$$

We show

Theorem

$\mu^*(A) = 0$ **iff** *there is an E-process for $\mathcal{P}$ tending to ∞ on A.*

Examples

Divergence

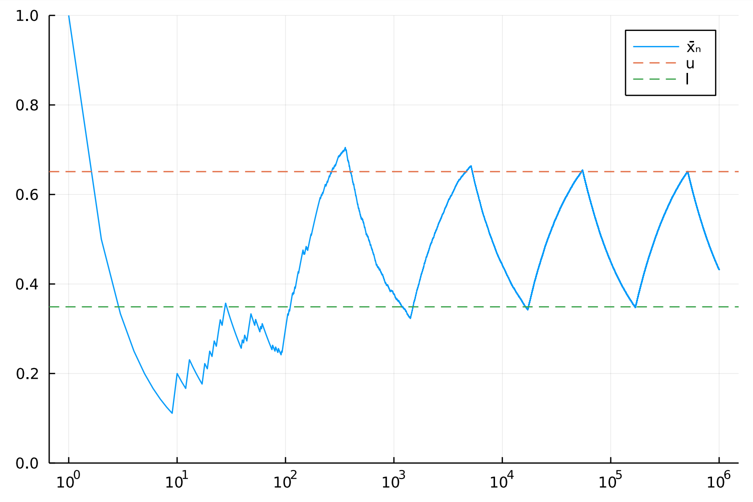
We focus on $X_1, X_2, \ldots$ i.i.d. from $P \in \mathcal{P}$ for different composite $\mathcal{P}$.

We write $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ for the running average.

We focus on the event

$$A_{\text{div}} \;=\; \left\{ \text{the limit } \lim_{n\to\infty} \bar{X}_n \text{ does not exist} \right\}$$

Divergence



$\bar{X}_n$ diverges iff there are rational $l < u$ s.t. both $\bar{X}_n \leq l$ and $\bar{X}_n \geq u$ infinitely often.

Examples

Bernoulli (Positive) Example

Question

Let

$$\mathcal{P} \;=\; \{P|P \text{ is i.i.d. Bernoulli}\}$$

Recall $\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$. Consider the event

$$A_{\text{div}} \;=\; \{\bar{x}_n \text{ does not converge}\}$$

Question

Let

$$\mathcal{P} \;=\; \{P | P \text{ is i.i.d. Bernoulli}\}$$

Recall $\bar{x}_n = \frac{1}{n}\sum_{i=1}^n x_i$. Consider the event

$$A_{\text{div}} \;=\; \{\bar{x}_n \text{ does not converge}\}$$

E-process diverging on A_{div}?

Bernoulli E-process explicit construction: E-values

For i.i.d. Bernoulli p

$$1 + \beta(x - p)$$

is a E-value for p as long as $\frac{-1}{1-p} \leq \beta \leq \frac{1}{p}$.

Bernoulli E-process explicit construction: E-values

For i.i.d. Bernoulli p

$$1 + \beta(x - p)$$

is a E-value for p as long as $\frac{-1}{1-p} \leq \beta \leq \frac{1}{p}$.

A slightly weaker but simpler E-value is

$$1 + \beta(x - p) \;\geq\; e^{\beta(x-p)-\beta^2}$$

as long as $\frac{-1}{2(1-p)} \leq \beta \leq \frac{1}{2p}$

Bernoulli E-process explicit construction: Upcrossings

For now fix $0 \leq l < u \leq 1$.

Suppose that $\bar{x}_{t_1} \leq l$ and $\bar{x}_{t_2} \geq u$.

If we product the weak e-value (with legal β) from time $t_1 + 1$ to time t_2, we make

$$\prod_{i=t_1+1}^{t_2} e^{\beta(x_i - p) - \beta^2} = \exp\Big(\beta\left(t_2\bar{x}_{t_2} - t_1\bar{x}_{t_1} - (t_2 - t_1)p\right) - (t_2 - t_1)\beta^2\Big)$$

Bernoulli E-process explicit construction: Upcrossings

For now fix $0 \leq l < u \leq 1$.

Suppose that $\bar{x}_{t_1} \leq l$ and $\bar{x}_{t_2} \geq u$.

If we product the weak e-value (with legal β) from time $t_1 + 1$ to time t_2, we make

$$\prod_{i=t_1+1}^{t_2} e^{\beta(x_i - p) - \beta^2} = \exp\Big(\beta\left(t_2\bar{x}_{t_2} - t_1\bar{x}_{t_1} - (t_2 - t_1)p\right) - (t_2 - t_1)\beta^2\Big)$$

Is that good?

Bernoulli E-process explicit construction: Upcrossings

If $\bar{x}_{t_1} \leq l < u \leq \bar{x}_{t_2}$, then

$$t_2 \;\geq\; t_1 \frac{1-l}{1-u}$$

If $u \geq p$, the hardest t_2 is that value. We then get

$$e^{t_1 \frac{u-l}{1-u}\left(\beta(1-p)-\beta^2\right)}$$

If we run with $\beta = (u-p)_+$ (which is legal), we guarantee

$$e^{t_1(u-l)(u-p)_+}$$

And as $t_1 \geq 1$ w.l.o.g., this is at least

$$e^{(u-l)(u-p)_+}$$

Down and up

A downcrossing from u at t_1 to l at t_2 can be symmetrically exploited with negative $\beta = -(p-l)_+$ for a gain of

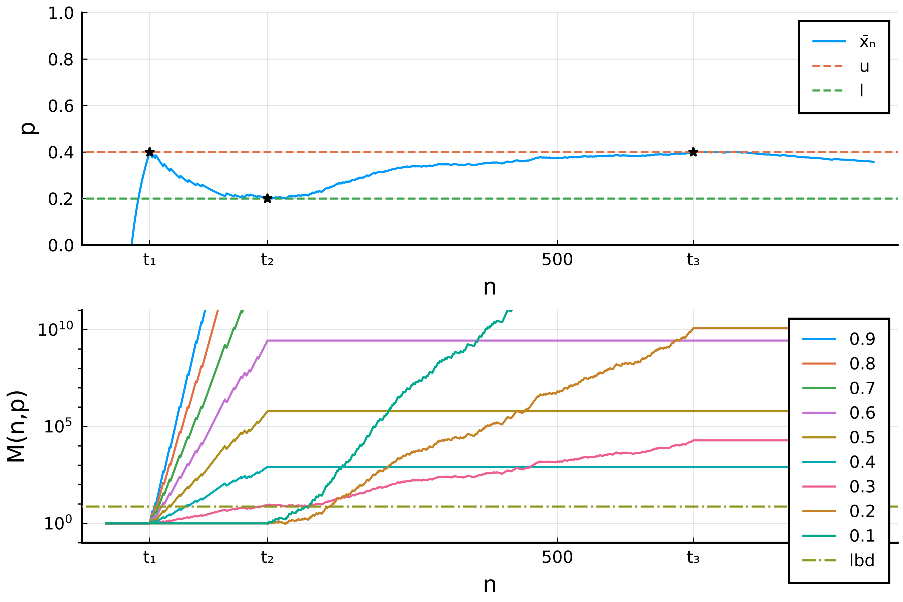
$$e^{(u-l)(p-l)_+}$$

Now for any p, a full down-up cycle gives at least

$$\min_{p \in [0,1]} e^{(u-l)((u-p)_+ + (p-l)_+)} \;=\; e^{(u-l)^2}$$

So the minimum of running products is an E-process winning $e^{(u-l)^2}$ every cycle.

In pictures



1.0
0.8
0.6
0.4
0.2
0.0
p
t₁
t₂
500
t₃
n
x̄ₙ
u
l
10^10
10^5
10^0
M(n,p)
t₁
t₂
500
t₃
n
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
lbd

Bernoulli Conclusion

Given $0 < l < u < 1$, we can make an E-process that multiplies by $e^{(u-l)^2}$ every down-up cycle. This hence has infinite supremum if $\liminf_n \bar{X}_n \leq l$ and $\limsup_n \bar{X}_n \geq u$.

Two more standard techniques finish the job

- A countable Bayesian mixture over rational l, u has ∞ supremum if $\lim_n \bar{X}_n$ does not exist.
- Bayesian mixture of height-stopped copies gets ∞ limit given ∞ supremum.

Examples

Truncated Cauchy (Negative) Example

Cauchy Example

Now let $\mathcal{P} = \{\text{i.i.d. } P_a | a > 0\}$ where P_a is Cauchy truncated to $[\pm a]$.

Since P_a has bounded support, $P_a(A_{\text{div}}) = 0$.

So A_{div} is $\mathcal{P}$-polar.

But $\mu^*(A_{\text{div}}) = 1$.

Why? A finite-time detector τ for A_{div} will fire on P_{Cauchy} since $P_{\text{Cauchy}}(A_{\text{div}}) = 1$.

But for large a, P_a looks enough like P_{Cauchy} at time τ.

So by our result, no diverging E-process exists.

More details

Consider any stopping time τ such that $A_{\text{div}} \subseteq \{\tau < \infty\}$. We have

$$\begin{aligned} P_a(\tau(X) < \infty) &\geq P_a(\tau(X) < \infty \text{ and } |X_t| < a \text{ for all } t \leq \tau(X)) \\ &= \mathbb{Q}(\tau(X) < \infty \text{ and } |X_t| < a \text{ for all } t \leq \tau(X)), \end{aligned}$$

Next, since $\mathbb{Q}(\tau(X) < \infty) \geq \mathbb{Q}(A_{\text{div}}) = 1$

$$\begin{aligned} &= \mathbb{Q}(|X_t| < a \text{ for all } t \leq \tau(X)) \\ &\geq \mathbb{Q}(|X_t| < a \text{ for all } t \leq T) - \mathbb{Q}(\tau(X) > T) \end{aligned}$$

for any $T \in \mathbb{N}$. Fix any $\varepsilon > 0$ and choose T large enough that $\mathbb{Q}(\tau(X) > T) \leq \varepsilon$

$$\geq \mathbb{Q}(|X_t| < a \text{ for all } t \leq T) - \varepsilon.$$

Sending a to ∞ we find that $\sup_{a \in (0,\infty)} P_a(\tau(X) < \infty) \geq 1 - \varepsilon$. Since τ was an arbitrary stopping time with $A_{\text{div}} \subseteq \{\tau < \infty\}$ it follows that $\mu^*(A_{\text{div}}) \geq 1 - \varepsilon$. Since this holds for every $\varepsilon > 0$, we obtain $\mu^*(A_{\text{div}}) = 1$.

Examples

Uniform Strong Law (Positive) Example

Recovering from Cauchy

Bernoulli is ok, but truncated Cauchy is too wild.

Need some uniformity.

Theorem

Let $\mathcal{P}$ be a family of probability measures P under which $(X_t)_{t \in \mathbb{N}}$ is i.i.d. with $\mathbb{E}_P[|X_1|] < \infty$. Assume also that $\mathcal{P}$ satisfies the centered uniform integrability condition

$$\lim_{K \to \infty} \sup_{P \in \mathcal{P}} \mathop{\mathbb{E}}_{P} \left[\left| X_1 - \mathop{\mathbb{E}}_{P}[X_1] \right| \mathbf{1}_{|X_1 - \mathbb{E}_P[X_1]| > K} \right] = 0.$$

Then $\mu^*(A_{\mathrm{div}}) = 0$.

Conclusion

Conclusion

We talked about confident finite-time detection of events in composite settings.

Paper also contains

- E-process characterisation for $\mu^*(A) > 0$
- Pricing interpretation of μ^*
- Line crossing inequality

References