VRIJE UNIVERSITEIT AMSTERDAM

Multiple Testing with Anytime-Valid Evidence

TAIL BOUNDS FOR SUMS OF RUNNING *maximums* OF DEPENDENT E-PROCESSES

Yury Tavyrikov • Jelle Goeman[1] • Rianne de Heide[2]

[1] Leiden University Medical Center [2] University of Twente • SAVI 2026

Sequential evidence, and the question

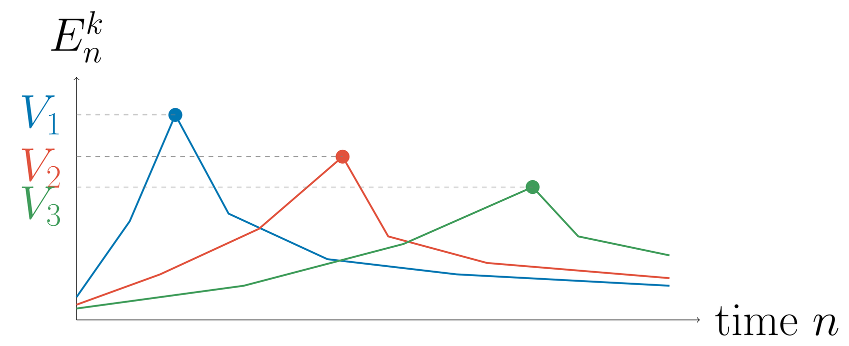
- K hypotheses $H_1, \dots, H_K$. Each is tested over time by an E-process $E^k = (E^k_n)_n$ on a common filtration: $\mathbb{E}[E^k_\tau] \le a_k$ for every stopping time τ (here $a_k = 1$), and the processes may be **arbitrarily dependent**.
- Anytime-valid summary: the **maximum** $V_k := \sup_n E^k_n$. Ville's inequality $\mathbb{P}_{H_k}(V_k \ge t) \le 1/t$ makes $1/V_k$ a p-value. Sort: $V_{(1)} \ge \cdots \ge V_{(K)}$.
- To test a group S, the natural pooled evidence is the **sum of maximums** $\sum_{k\in S} V_k$.

But a maximum is not an e-value: $\mathbb{E}[V_k]$ can be $+\infty$, so e-value methods (e-BH, closed testing) cannot take it directly.

Two ways out. (i) an *adjuster* A with $\int_1^\infty A(u)u^{-2}\,du \le 1$ turns V_k into an e-value $A(V_k)$ — but shrinks it and loses power; (ii) bound $\sum_{k\in S} V_k$ **directly**.

Question: how large can $\sum_{k\in S} V_k$ be, under arbitrary dependence?

The maximums happen at different times



Different experiments get strong at different moments. We want each to contribute its strongest evidence — its own **maximum** — whenever it occurred.

Summary & references

Takeaways. Maximums $V_k = \sup_n E^k_n$ are not e-values, yet $\sum_k V_k$ is controlled at $\frac{2H_K}{\lambda}\sum_k a_k$ — the log price is **sharp** (harmonic ladder), and any aggregator $\sum_k f(V_k)$ obeys the same recipe. One stopped sum + the rank rule; dependence enters only through linearity. Each group's e-value $\Rightarrow$ anytime-valid closed testing / FDR; one functional $\Phi \Rightarrow$ detection over all 2^K groups.

References. Ville (1939) • Vovk–Wang (2020, 2021) • Wang–Ramdas (2022) • Marcus–Peto–Gabriel (1976) • Xu et al. (2025).

Main result

$$\mathbb{P}\Big(\sum_{k=1}^{K} V_k > \lambda\Big) \;\le\; \frac{2H_K}{\lambda}\sum_{k=1}^{K} a_k, \qquad H_K = \sum_{r=1}^{K}\frac{1}{r} \approx \ln K.$$

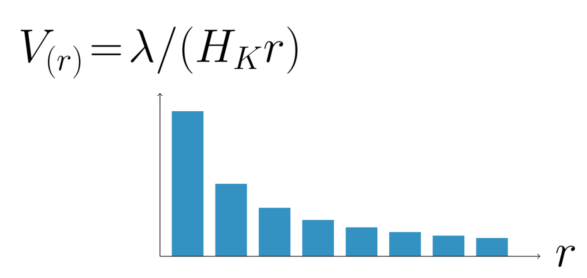
- Equal $a_k = 1$: $\dfrac{2KH_K}{\lambda}$. For a group of size s: $\dfrac{2sH_s}{\lambda}$.
- Holds for **any** dependence. Combining K asynchronous maximums costs a **logarithm** H_K — not a factor K.

The bound is sharp

There are explicit dependent E-processes ($a_k = 1$) with

$$\mathbb{P}\Big(\sum_k V_k \ge \lambda\Big) = \min\Big(\tfrac{KH_K}{\lambda}, 1\Big).$$

The worst case is the **harmonic ladder**: maximums $V_{(r)} = \dfrac{\lambda}{H_K\, r}$ (sum $= \lambda$), all firing together. Built from a random row-permutation of the staircase matrix $A_{ij} = \mathbf{1}\{j \le i\}$.



Why it holds (proof idea)

Gather the maximums into one **stopped sum**. Place levels $0 < x_1 < \cdots < x_M = \lambda$; each time a process crosses a level, stop it and read $E^k \ge x_i$. Let $S_M = \sum_k \sum_{m=1}^{M} E^k_{\tau^k_m}$.

- **Budget** (linear, dependence-free): $\mathbb{E}[S_M] \le M\sum_k a_k$.
- **Rank rule**: the r-th process at a level gets r stops, so (with $p_{(r)}$ the depth of the r-th deepest process)

$$S_M \;\ge\; \frac{\lambda}{2M}\sum_{r=1}^{K} r\, p_{(r)}^2.$$

- The event forces $\sum_k p_k \ge M$; minimizing $\sum_r r\, p_{(r)}^2$ gives the ladder $p_{(r)} \propto 1/r$, value M^2/H_K. Hence $S_M \ge \frac{\lambda M}{2H_K}$, and Markov gives the bound.

The rank-weight r is what produces the harmonic number.

Any aggregator $\sum_k f(V_k)$

For any nondecreasing f with $f(0) = 0$, warp the grid to $x_i = f^{-1}(i\lambda/M)$. With μ solving $\sum_{r=1}^{K} f(\mu/r) = \lambda$,

$$\mathbb{P}\Big(\sum_k f(V_k) > \lambda\Big) \le \frac{\lambda}{\Lambda_K(\lambda)}\sum_k a_k,$$

$$\Lambda_K(\lambda) = \lambda\mu - \sum_{r=1}^{K} r\int_0^{\mu/r} f.$$

Powers $f(x) = x^s$ ($H_K^{(s)} = \sum_r r^{-s}$):

$$\mathbb{P}\big(\textstyle\sum_k V_k^s > \lambda\big) \;\le\; \frac{s+1}{s}\cdot\frac{K\,(H_K^{(s)})^{1/s}}{\lambda^{1/s}}$$

$s = 1$: $2KH_K/\lambda$ (price $\sim \ln K$). $s = 2$: $H_K^{(2)} \le \pi^2/6$, so the price is **bounded** — more weight on the top maximums is cheaper.

Applications

1. Closed testing

For each group S, the normalized sum $\bar V_S = \dfrac{\sum_{k\in S} V_k}{2|S|H_{|S|}}$ has **Ville-type tails**

$$\mathbb{P}_{H_S}\big(\bar V_S \ge t\big) \;\le\; 1/t \quad (t > 0),$$

the very tail an e-value obeys by Markov — which is all the **e-closed** procedure needs. So $\bar V_S$ certifies H_S (reject when $\bar V_S \ge 1/\alpha$) and plugs into e-closed testing as the per-group certificate, *without* itself being an e-value ($\mathbb{E}[\bar V_S]$ may be $+\infty$): reject H_i when every group containing it is rejected — only *marginal* validity per group, no price for the 2^K groups. $\Rightarrow$ anytime-valid familywise / true-discovery-proportion control, each hypothesis at its own maximum.

One group: $\sum_{k\in S} V_k = \sum 1/p_k$ is the harmonic mean of the p-values (Vovk–Wang) — here sequential, with sharp constant.

2. Signal in some unknown group

One functional $\Phi = \sum_{r=1}^{K} r\, V_{(r)}^2$ controls *all* 2^K groups at once:

$$\mathbb{P}\Big(\exists\, S:\ \tfrac{1}{\sqrt{H_{|S|}}}\textstyle\sum_{k\in S} V_k > \lambda\Big) \le \frac{2K\sqrt{H_K}}{\lambda}.$$

Price $\sqrt{H_K}\,K$, not 2^K — a global test for signal in a group not chosen in advance. (Powers: divide by $(H_{|S|}^{(s)})^{1/(s+1)}$, bounded for $s > 1$.)