E-values: Calibration, combination, and applications

Sam Rosenberg
University of Pennsylvania
rosesamk@sas.upenn.edu

April 7, 2022

Outline

The multiple comparisons problem

- Consider a meta-analysis in which we want to test a hypothesis common to a number of past studies
- What happens if we cannot access each of the original datasets?
  - Can we combine information from published results to perform the analysis?
- In practice, often combine p-values for multiple testing or testing multiple hypotheses
  - Class of functions for merging p-values complicated[1]
  - Class of functions for merging e-values, an alternative, is much nicer[2]

[1] Vovk, B. Wang, and R. Wang 2022.
[2] Vovk and R. Wang 2021.

Setting

- Consider a probability space $(\Omega, \mathcal{A}, Q)$ with sample space Ω (a set), event space $\mathcal{A}$ (a σ-algebra on Ω), and probability measure Q
- For a random variable $X : \Omega \to S \subseteq \overline{\mathbb{R}}$ on our probability space, denote its expectation with respect to Q by $\mathbb{E}^Q[X] := \int_\Omega X \, dQ$
  - We write $\mathbb{E}[X]$ when Q is obvious from context

E-values

- An e-variable $E : \Omega \to [0, \infty]$ is an extended random variable with $\mathbb{E}^Q[E] \leq 1$
  - Any value taken by an e-variable (a realization of the random variable) is called an e-value
  - The set of all e-variables for our probability space is denoted $\mathcal{E}_Q$
  - Larger values of E are stronger evidence against the null hypothesis Q (e.g. since $Q(E \geq e) \leq e^{-1}$ by Markov's inequality)
- Why do we let $E = \infty$?
  - When testing a null hypothesis measure Q with a pre-specified e-variable E, an observed e-value of $E(\omega) = \infty$ allows us to reject Q

P-values

- A p-variable is a $P : \Omega \to [0,1]$ s.t.

$$\forall \epsilon \in (0,1),\ Q(P \leq \epsilon) \leq \epsilon$$

  - P-values are defined similarly to e-values
  - $\mathcal{P}_Q$ denotes the set of all p-variables for our probability space

Calibrators: P-values to e-values

- Calibrators convert between p-values and e-values (and vice versa):
  - A p-to-e calibrator is a decreasing function $f : [0,1] \to [0,\infty]$ if for any probability space $(\Omega, \mathcal{A}, Q)$ and any p-variable $P \in \mathcal{P}_Q$, $f(P) \in \mathcal{E}_Q$
    - i.e. for any probability space, the p-variables on that space are transformed to e-variables on that space by f
  - A calibrator f dominates another calibrator g if $f \geq g$ and strictly dominates if $f \neq g$
  - A calibrator is admissible if it isn't strictly dominated by another calibrator

P-to-e calibrators characterization

We have the following characterization of (admissible) p-to-e calibrators:

Theorem 1

A decreasing function $f : [0, 1] \to [0, \infty]$ is a p-to-e calibrator iff $\int_0^1 f(p)\, dp \leq 1$.
It is admissible iff f is upper semicontinuous, $f(0) = \infty$, and $\int_0^1 f(p)\, dp = 1$.

Since $\int_0^1 \kappa p^{\kappa - 1} = 1$ for $\kappa \in (0, 1)$, an example family of calibrators immediately following from this result is:

$$\left\{ f_\kappa(p) := \kappa p^{\kappa-1} \mid \kappa \in (0, 1) \right\}.$$

P-to-e calibrators characterization: Proof

Proof.

$\boldsymbol{f}$ **is a p-to-e calibrator** $\boldsymbol{\Rightarrow \int_0^1 f \leq 1}$: This follows immediately, taking $(\Omega, \mathcal{A}, Q) = ([0,1], \sigma([0,1]), \mu)$ and $P = \mathrm{id}_{[0,1]}$, where μ is the uniform measure on $[0,1]$.

$\boldsymbol{\int_0^1 f \leq 1 \Rightarrow f}$ **is a p-to-e calibrator**: Suppose P is a p-variable and P' is uniform on $[0,1]$. Then $Q(P < x) \leq x =: Q(P' < x)$ for any $x \in (0,1)$ since P is a p-variable and by definition of P' as uniform.

So $Q(f(P) > y) \leq Q(f(P') > y)$ for any $y \in (0,1)$ since f is decreasing.

Then $\mathbb{E}[f(P)] \leq \mathbb{E}[f(P')] = \int_0^1 f(p)dp \leq 1$, where the equality holds since P' is uniform on $[0,1]$.

We omit the proof of the latter statement for brevity, but it follows readily from the definitions.

Calibrators: E-values to p-values

- The case for converting from e-values to p-values is similar:
  - An e-to-p calibrator is a decreasing function $f : [0, \infty] \to [0, 1]$ iff for any probability space $(\Omega, \mathcal{A}, Q)$ and any e-variable $E \in \mathcal{E}_Q$, $f(E) \in \mathcal{P}_Q$
  - (Strict) domination and admissibility are defined analogously to before

E-to-p calibrators characterization

Similar to Theorem 1, we can characterize (admissible) e-to-p calibrators:

Theorem 2

The function $f : [0, \infty] \to [0, 1]$ *given by* $f(t) := \min(1, t^{-1})$ *is an e-to-p calibrator. In fact, it dominates every other e-to-p calibrator and is the only admissible e-to-p calibrator.*

So, there are effectively many admissible p-to-e calibrators, but only a single admissible e-to-p calibrator.

E-to-p calibrators characterization: Proof

Proof.
***f(t)* is an e-to-p calibrator**: For $E \in \mathcal{E}_Q$ and $\epsilon \in (0,1)$,

$$Q(f(E) \leq \epsilon) = Q(\epsilon^{-1} \leq f(E)^{-1}) = Q(\epsilon^{-1} \leq E) \leq \frac{\mathbb{E}^Q[E]}{\epsilon^{-1}} \leq \epsilon,$$

where the second equality is because $\epsilon^{-1} > 1$ so $f(E)^{-1} = E$ when $\epsilon^{-1} \leq f(E)^{-1}$, the first inequality uses Markov's inequality, and the last inequality is true because E is an e-variable.

E-to-p calibrators characterization: Proof (con.)

Proof.

***f*(*t*) dominates any e-to-p calibrator**: Let g be another e-to-p calibrator. Note that it is sufficient to show that f dominates g. Suppose that for some $t_0 \in [0, \infty]$, $g(t_0) < f(t_0) = \min(1, t_0^{-1})$. We have two cases:

- When $g(t_0) < f(t_0) = t_0^{-1}$ for some $t_0 > 1$, consider an e-variable E that equals t_0 with probability $1/t_0$ and is 0 otherwise. Note that $g(E) = g(t_0) < t_0^{-1}$ with probability t_0^{-1}. But if $g(E)$ was a p-variable, then $P(g(E) \leq g(t_0)) \leq g(t_0) < t_0^{-1}$, indicating a contradiction (since $t_0^{-1} = P(g(E) = g(t_0)) \leq P(G(E) \leq g(t_0))$.
- When $g(t_0) < f(t_0) = 1$ for some $t_0 \in [0, 1]$, consider an e-variable E that equals 1 a.s. We have $g(E) = g(t_0) < 1$ a.s., so $P(g(E) < t_0) = 1 > t_0$ and $g(E)$ cannot be a p-variable.

Merging arbitrary e-values

- An e-merging function of $K \geq 2$ e-values is an increasing Borel function $F : [0, \infty)^K \to [0, \infty)$ s.t. for any probability space $(\Omega, \mathcal{A}, Q)$ and random variables $E_1, \dots, E_K$ on the space,

$$E_1, \dots, E_K \in \mathcal{E}_Q \Rightarrow F(E_1, \dots, E_K) \in \mathcal{E}_Q$$

- An e-merging function F dominates another G if $F \geq G$ and is strict if $F(e) > G(e)$ for some $e \in [0, \infty)^K$
  - F essentially dominates G if for any $e \in [0, \infty)^K$,

$$G(e) > 1 \Rightarrow F(e) \geq G(e)$$

(i.e. F is at least as good as G when G is useful)

- An e-merging function is admissible if it is not dominated by another e-merging function
  - i.e. it is maximal in the partial order defined by the relation of domination

Domination of e-merging functions

Theorem 3

The arithmetic mean $M_K : [0, \infty)^K \to [0, \infty)$*, given by*

$$M_K(e_1, \ldots, e_K) := \frac{e_1 + \cdots + e_K}{K},$$

essentially dominates any symmetric e-merging function.

Theorem 4

Suppose that F *is a symmetric e-merging function. Then* F *is dominated by a function in the class*

$$M_{K,\lambda} := \left\{ \lambda + (1 - \lambda) M_k : \lambda \in [0, 1] \right\}.$$

Specifically, F *is admissible iff* $F \in M_{K,\lambda}$ *with* $F(0) = \lambda$*.*

Domination of e-merging functions: Proof

We will provide a proof of Theorem 3, but omit that of Theorem 4 for expositional brevity (see Vovk and R. Wang 2021).

Proof.

Suppose F is a symmetric e-merging function. Suppose for contradiction that there is some $(e_1, \ldots, e_K) \in [0, \infty)^K$ s.t.

$$b := F(e_1, \ldots, e_K) > \max\Big(M_K(e_1, \ldots, e_K), 1\Big) =: a.$$

Let π by uniformly selected from S_K and $(D_1, \ldots, D_K) := (e_{\pi(1)}, \ldots, e_{\pi(K)})$ (i.e. a random permutation of our e-values).
Take $(D'_1, \ldots, D'_K) := (D_1, \ldots, D_K)1_A$, where A is independent of π and $P(A) = a^{-1}$.

Proof.

For each k, $\mathbb{E}[D_k] = M_K(e_1, \ldots, e_K)$ since π was uniform on S_K implies that D_k equals $e_1, \ldots, e_K$ with equal probability.
Then $\mathbb{E}[D'_k] = M_K(e_1, \ldots, e_K)/a \leq 1$ by construction.
Since D'_k is nonnegative, we must have $D'_k \in \mathcal{E}_Q$.
Then $F(D'_1, \ldots, D'_K) \in \mathcal{E}_Q$ since F is an e-merging function, so $\mathbb{E}[F(D'_1, \ldots, D'_K)] \leq 1$.
But by symmetry,

$$\begin{aligned}\mathbb{E}[F(D'_1, \ldots, D'_K)] &= Q(A)F(e_1, \ldots, e_K) + (1 - Q(A))F(0, \ldots, 0)\\ &\geq b/a\\ &> 1,\end{aligned}$$

a contradiction.
So, no such $(e_1, \ldots, e_K)$ exists and M_K essentially dominates F.

Merging independent e-values

- Can also consider merging functions specifically for independent e-values
- An ie-merging function of $K \geq 2$ e-values is an increasing Borel function $F : [0, \infty)^K \to [0, \infty)$ s.t. for any probability space $(\Omega, \mathcal{A}, Q)$,

$$E_1, \ldots, E_k \in \mathcal{E}_Q \text{ are independent } \Rightarrow F(E_1, \ldots, E_K) \in \mathcal{E}_Q$$

- Define domination, strict domination, and admissibility for ie-merging functions analogously to case of e-merging functions
- Define $i\mathcal{E}_Q^K \subseteq \mathcal{E}_Q^K$ to be the set of component-wise independent random vectors in $\mathcal{E}_Q^K$ and $\mathbf{1} := (1, \ldots, 1) \in \mathbb{R}^K$
- Say an ie-merging function F weakly dominates another ie-merging function G if for all $e_1, \ldots, e_K$,

$$(e_1, \ldots, e_K) \in [1, \infty)^K \Rightarrow F(e_1, \ldots, e_K) \geq G(e_1, \ldots, e_K)$$

  - i.e. F is at least as good as G when all e-value inputs are useful

(i)e-merging functions characterization

We have the following sufficient condition for being an (i)e-merging function:

Theorem 5

For an increasing Borel function $F : [0,\infty)^K \to [0,\infty)$, if $\mathbb{E}[F(E)] = 1$ for all $E \in \mathcal{E}_Q^K$ s.t. $\mathbb{E}[(E_1,\dots,E_K)] = \mathbf{1}$ (resp. for all $E \in i\mathcal{E}_Q^K$ s.t. $\mathbb{E}[(E_1,\dots,E_K)] = \mathbf{1}$), then F is an admissible e-merging (resp. ie-merging) function.

So, the M_K is an admissible e-merging function and the U-statistics

$$U_n(e_1,\dots,e_K) := \frac{1}{\binom{K}{n}} \sum_{\{k_1,\dots,k_n\}\subseteq\{1,\dots,K\}} e_{k_1}\dots e_{k_n},\ n \in \{0,\dots,K\}$$

and their convex mixtures are admissible ie-merging functions

- Contains the product ($n = K$), arithmetic average M_K ($n = 1$), and constant function 1 ($n = 0$)
- Family is not complete class of admissible ie-merging functions

(i)e-merging functions characterization: Proof

Proof.

Clearly, $F(E) \in \mathcal{E}_Q^K$ (resp. $F(E) \in i\mathcal{E}_Q^K$), so F is an e-merging (resp. ie-merging function).

Suppose for contradiction that there is an (i)e-merging function F such that $G \geq F$ and $G(e_1, \ldots, e_K) > F(e_1, \ldots, e_K)$ for some $(e_1, \ldots, e_K) \in [0, \infty)^K$.

Take $(E_1, \ldots, E_K) \in \mathcal{E}_Q^K$ (resp. $i\mathcal{E}_Q^K$) with $\mathbb{E}[(E_1, \ldots, E_K)] = \mathbf{1}$ s.t. $Q((E_1, \ldots, E_K) = (e_1, \ldots, e_K)) > 0$ (e.g. by considering a distribution with a positive mass on $e_1, \ldots, e_K$).

Then

$$Q(G(E_1, \ldots, E_K) > F(E_1, \ldots, E_K)) > 0$$

since $(E_1, \ldots, E_K) = (e_1, \ldots, e_K) \Rightarrow G(E_1, \ldots, E_K) > F(E_1, \ldots, E_K)$.

So,

$$\mathbb{E}[G(E_1, \ldots, E_K)] > \mathbb{E}[F(E_1, \ldots, E_K)] = 1,$$

contradicting that G is an (i)e-merging function.

Thus, F is admissible.

Domination of ie-merging functions

We have an additional nice property of the product merging function:

Theorem 6

The product $(e_1, \ldots, e_K) \mapsto e_1 \ldots e_K$ weakly dominates any ie-merging function.

Proof.

Suppose for contradiction that there exists $(e_1, \dots, e_K) \in [1, \infty)^K$ s.t. $F(e_1, \dots, e_K) > e_1 \dots e_K$ for some ie-merging function F.

Let $E_1, \dots, E_K$ be independent random variables such that for $k \in \{1, \dots, K\}$, $E_k = e_k$ with probability e_k^{-1} and $E_k = 0$ otherwise.

Clearly, each E_k is an e-variable, but

$$\begin{aligned} \mathbb{E}[F(E_1, \dots, E_K)] &\geq F(e_1, \dots, e_K) Q(E_1 = e_1, \dots, E_K = e_K) \\ &> (e_1 \dots e_K)(e_1^{-1} \dots e_K^{-1}) \\ &= 1. \end{aligned}$$

So, F is not an ie-merging function and the product weakly dominates any ie-merging function.

Setting for multiple testing

- Let $(\Omega, \mathcal{A})$ be a measurable space (the sample space and event space, resp.) and $\mathcal{B}(\Omega)$ be the family of all probability measures on the space
- Say E is an e-variable w.r.t. a composite null hypothesis $H \subseteq \mathcal{B}(\Omega)$ if $\mathbb{E}^Q[E] \leq 1$ for any measure $Q \in H$
- In multiple testing, have a set of composite null hypotheses $\{H_k : 1 \leq k \leq K\}$
- For each k, have an e-variable E_k w.r.t. H_k

Family-wise validity (FWV)

- A conditional e-variable is a family of extended nonnegative random variables $E_Q, Q \in \mathcal{B}(\Omega)$ satisfying

$$\forall Q \in \mathcal{B}(Q), \mathbb{E}^Q[E_Q] \leq 1$$

(i.e. each $E_Q \in \mathcal{E}_Q$)

- Say that extended random variables $E_1^*, \ldots, E_K^*$ are family-wise valid (FWV) for testing $H_1, \ldots, H_K$ if there exists a conditional e-variable $(E_Q)_{Q \in \mathcal{B}(\Omega)}$ s.t.

$$\forall k \in \{1, \ldots, K\}, \forall Q \in H_k : E_Q \geq E_k^*$$

  - From the definition of a conditional e-variable, we see that this equivalent to

$$\forall Q \in \mathcal{B}(\Omega) : \mathbb{E}^Q\left[\max_{k : Q \in H_k} E_k^*\right] \leq 1$$

  (i.e. joint validity of the e-variables E_k^*)
  - Say that $(E_Q)_{Q \in \mathcal{B}(\Omega)}$ witnesses that $E_1^*, \ldots E_K^*$ are FWV

Adjusting arbitrary e-values for multiple testing

Algorithm 1 Adjusting arbitrary e-values for multiple testing

Require: Arbitrary e-values $e_1, \dots, e_K$.

Find the ordering permutation $\pi \in S_k$ s.t. $e_{\pi(1)} \leq \cdots \leq e_{\pi(K)}$

Define the order statistics $e_{(k)} := e_{\pi(k)}$ for $k \in \{1, \dots, K\}$

Set $S_0 := 0$

for $i \in \{1, \dots, K\}$ **do**

  $S_i := S_{i-1} + e_{(i)}$ (cumulative sum of ordered values)

for $k \in \{1, \dots, K\}$ **do**

  $e^*_{\pi(k)} := e_{\pi(k)}$

  for $i \in \{1, \dots, k-1\}$ **do**

    $e := \frac{e_{\pi(k)} + S_i}{i+1}$ (average of first i and k-th ordered values)

    if $e < e^*_{\pi(k)}$ **then**

      $e^*_{\pi(k)} := e$

return $e^*_{\pi(1)}, \dots, e^*_{\pi(K)}$

FWV arbitrary e-value adjustment

Theorem 7

Algorithm 1 is family-wise valid. It has a computational complexity of $O(K^2)$.

FWV arbitrary e-value adjustment: Proof

Proof.

Note that the computational complexity immediately follows from the nested for-loops, so we need only show FWV.

It suffices to check that the e-variables $E_1^*, \ldots, E_k^*$ from Algorithm 1 are FWV. For any subset $I \subseteq \{1, \ldots, K\}$, the composite hypothesis H_I is

$$H_I := \left(\bigcap_{k \in I} H_k \right) \bigcap \left(\bigcap_{k \in \{1, \ldots, K\} \setminus I} H_k^c \right),$$

with H_k^c defined as the complement of H_k. The conditional e-variable witnessing the FWV of $E_1^*, \ldots, E_K^*$ is the arithmetic mean

$$E_Q := \frac{1}{|I_Q|} \sum_{k \in I_Q} E_k,$$

with $I_Q := \{k | Q \in H_k\}$ and E_Q defined as 1 when $I_Q = \emptyset$.

FWV arbitrary e-value adjustment: Proof (con.)

Proof.

We use the the following (conservative) definition for our adjusted e-variables E_k^*:

$$E_k^* := \min_{I \subseteq \{1,\ldots,K\}: k \in I} \frac{1}{|I|} \sum_{i \in I} E_i.$$

Note that for each $k \in \{1, \ldots, K\}$,

$$E_{\pi(k)}^* = \min_{i \in \{0,\ldots,k-1\}} \frac{E_{\pi(k)} + E_{(1)} + \cdots + E_{(i)}}{i+1}.$$

Proof.

In lines 3-5 of Algorithm 1, we compute

$$S_i := e_{(1)} + \cdots + e_{(i)},\ i \in \{1, \dots, K\}$$

and in lines 8-9 we calculate

$$e_{k,i} := \frac{e_{\pi(k)} + e_{(1)} + \cdots + e_{(i)}}{i+1},\ i \in \{1, \dots, k-1\}$$

Then

$$e^*_{\pi(k)} = \min_{i \in \{1,\dots,k-1\}} e_{k,i} = \min_{i \in \{0,\dots,k-1\}} \frac{e_{\pi(k)} + e_{(1)} + \cdots + e_{(i)}}{i+1},$$

so the algorithm is in fact FWV since these are e-values associated with the adjusted e-variables E^*_k.

Algorithm 2 Adjusting independent e-values for multiple testing

Require: Independent e-values $e_1, \ldots, e_K$.

1: Let $a := \prod_{k \in \{1,\ldots,K\}: e_k < 1} e_k$ (with empty product defined as 1)
2: **for** $k \in \{1, \ldots, K\}$ **do**
3: $\quad e_k^* := a e_k$
4: **return** $e_1^*, \ldots, e_K^*$

FWV independent e-value adjustment

Theorem 8

Algorithm 2 is family-wise valid. It has a computational complexity of $O(K)$.

Note that Algorithm 2 is actually FWV for sequential e-variables, but we focus on independent e-variables for brevity.

FWV independent e-value adjustment: Proof

Proof.

Again, the computational complexity immediately follows from the algorithm pseudocode so we focus on proving FWV.

As before, take $I_Q := \{k | Q \in H_k\}$. Our conditional e-variable that witnesses that $e_1^*, \ldots, e_K^*$ are FWV are the ones given by the product ie-merging function, i.e.

$$E_Q := \prod_{k \in I_Q} E_k,$$

where our adjusted e-variables are

$$E_k^* := \min_{I \subseteq \{1, \ldots, K\} : k \in I} \prod_{i \in I} E_i.$$

By inspection, we can see that $e_1^*, \ldots, e_K^*$ in Algorithm 2 are realizations of $E_1^*, \ldots, E_K^*$, so the Algorithm is FWV.

Conclusion

- E-values as an alternative to p-values
- Conversion between p- and e-values with calibrators that have known domination structure
- Combine e-values together with e-merging functions that have complicated domination structure
- FWV valid algorithms for adjusting (independent) e-values
- Applications to multiple testing

References

Vovk, Vladimir, Bin Wang, and Ruodu Wang (2022). "Admissible ways of merging p-values under arbitrary dependence". In: *The Annals of Statistics* 50.1, pp. 351–375. DOI: 10.1214/21-AOS2109. URL: https://doi.org/10.1214/21-AOS2109.

Vovk, Vladimir and Ruodu Wang (2021). "E-values: Calibration, combination and applications". In: *The Annals of Statistics* 49.3, pp. 1736–1754. DOI: 10.1214/20-AOS2020. URL: https://doi.org/10.1214/20-AOS2020.