# The only admissible way of merging arbitrary e-values

Ruodu Wang*

March 10, 2025

**Abstract**

We prove that the only admissible way of merging arbitrary e-values is to use a weighted arithmetic average. This result completes the picture of merging methods for arbitrary e-values, and generalizes the result of Vovk and Wang (2021, Annals of Statistics, 49(3), 1736–1754) that the only admissible way of symmetrically merging e-values is to use the arithmetic average combined with a constant. Although the proved statement is naturally anticipated, its proof relies on a sophisticated application of optimal transport duality and a minimax theorem.

**Keywords**: Hypothesis testing, e-variables, arithmetic average, admissible decisions, optimal transport duality

## 1 Introduction

E-values and e-processes in hypothesis testing have been an active research area in the recent years, led by a series of papers including Shafer (2021), Vovk and Wang (2021), Grünwald et al. (2024) and Wasserman et al. (2020). The popularity of e-values came from its various features such as flexibility, anytime validity, robustness, post-hoc decision validity, and a strong connection to betting and martingales. Wang and Ramdas (2022), Grünwald (2024), and Ramdas et al. (2023) offered many discussions on these features.

Arguably, one of the central advantages of e-values, especially in contrast to p-values, is that they are easy to combine. It is elementary that a weighted arithmetic average of e-values, with weights summing to 1, is an e-value. This convenient feature has been used in many applications. For instance, it is used to obtain admissible ways of merging p-values by Vovk et al. (2022), to construct discovery matrices by Vovk and Wang (2023), to deranomize the Model-X knockoffs of Barber and Candès (2015) by Ren and Barber (2024), and to boost the power of the e-Benjamini-Hochberg procedure of Wang and Ramdas (2022) by Lee and Ren (2024).

*Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada. E-mail: wang@uwaterloo.ca.

A natural question is, in addition to the weighted averages, whether there are other useful e-merging functions, i.e., functions that produce an e-value from several input e-values. If we additionally assume that the input e-values are independent or sequential, then many other admissible methods exist, as studied by Vovk and Wang (2021, 2024). Without assuming particular dependence structures, the simple question of identifying useful e-merging functions turns out to be surprisingly nontrivial. As a central result on the combination of e-values, Vovk and Wang (2021, Theorem 3.2) showed, through several additional technical results, that all admissible symmetric e-merging functions take the form of a convex combination of the arithmetic average and the constant 1. From there, one may naturally conjecture that, without assuming symmetry, admissible e-merging functions should take the form of a convex combination of a weighted average and the constant 1. This was not established by Vovk and Wang (2021) or its follow-up papers due to technical challenges.

The main aim of this short paper is to prove the above conjecture, and thus to offer a complete characterization of all admissible e-merging functions. This justifies the use of weighted averages of e-values in many applications (where the weighted averages may have already been used), and frees us of any doubts about whether there are better choices in particular contexts. Although the anticipated mathematical result is natural, the proof requires a delicate analysis via optimal transport duality, very different from the proof for the symmetric case by Vovk and Wang (2021).

At an abstract level, the reason why optimal transport duality is essential is due to the requirement of an e-merging function to work for e-values with any dependence structure. Maximization over all dependence structures is a classic problem in optimal transport theory, and, more precisely, in our context it is a multi-marginal optimal transport problem; see Rachev and Rüschendorf (1998) and Villani (2009) for more on optimal transport theory. On a related matter, optimal transport duality is crucial for results of Vovk et al. (2022), where e-values serve as intermediate tools for characterizing admissible ways of merging p-values.

Merging functions of some subclasses of e-values are discussed in Section 5.

## 2 The main result

Let $K$ be a positive integer and denote by $[K] = \{1, \dots, K\}$. Write $\mathbb{R}_+ = [0, \infty)$ and denote by $\Delta_n$ the standard simplex in $\mathbb{R}^n$ for $n \in \mathbb{N}$, that is,

$$\Delta_n = \left\{ (x_1, \dots, x_n) \in [0,1]^n : \sum_{k=1}^{n} x_k = 1 \right\}.$$

A *hypothesis* is a collection of probability measures on a measurable space $(\Omega, \mathcal{F})$ called the sample space.

An *e-variable* $E$ for a hypothesis $\mathcal{H}$ is a $[0, \infty]$-valued random variable satisfying $\mathbb{E}^{\mathbb{P}}[E] \le 1$ for all $\mathbb{P} \in \mathcal{H}$. We have been using the loose term "e-values" for e-variables in the Introduction. An *e-merging function* is an increasing Borel function $F : \mathbb{R}_+^K \to \mathbb{R}_+$ such that for any hypothesis, $F(E_1, \dots, E_K)$ is an e-variable for any e-variables $E_1, \dots, E_K$. We have safely excluded infinite points in the domain of $F$; see Vovk and Wang (2021, Appendix C) for a justification.

An e-merging function $F$ is *admissible* if for any e-merging function $G$, $G \geq F$ implies $G = F$. That is, $F$ cannot be strictly improved. All equalities and inequalities are point-wise.

As explained in Vovk and Wang (2021, Appendix D), to study merging methods, it is necessary and sufficient to consider the simple hypothesis $\{\mathbb{P}\}$ with an atomless probability measure $\mathbb{P}$, and all results carry through to any settings of possibly composite hypotheses. Therefore, in the sequel, all e-variables are defined for the fixed $\mathcal{H} = \{\mathbb{P}\}$, and $\mathbb{E}$ represents the expectation with respect to $\mathbb{P}$. We omit mentioning $\mathbb{P}$ explicitly in most places.

For any vector $\boldsymbol{\lambda} \in \Delta_{K+1}$, define the mapping $M_{\boldsymbol{\lambda}} : \mathbb{R}_+^K \to \mathbb{R}_+$ by $\mathbf{e} \mapsto \boldsymbol{\lambda} \cdot (\mathbf{e}, 1)$. The mapping $M_{\boldsymbol{\lambda}}$ outputs the $\boldsymbol{\lambda}$-weighted average of its input arguments and the constant 1. It is an e-merging function of dimension $K$. The main result in the paper is stated next.

**Theorem 1.** *For a function $F : \mathbb{R}_+^K \to \mathbb{R}_+$,*

*(i) if $F$ is an e-merging function, then $F \leq M_{\boldsymbol{\lambda}}$ for some $\boldsymbol{\lambda} \in \Delta_{K+1}$;*

*(ii) $F$ is an admissible e-merging function if and only if $F = M_{\boldsymbol{\lambda}}$ for some $\boldsymbol{\lambda} \in \Delta_{K+1}$.*

It suffices to show (i), as (ii) follows directly from (i) and the simple fact that $M_{\boldsymbol{\lambda}}$ is an e-merging function not dominated by any other weighted average. In Section 3 we will prove Theorem 1. Before that, we address the simplest case $K = 1$, which will be used in the proof of Theorem 1. In what follows, a random variable is binary if it takes at most two possible values.

**Lemma 2.** *Let $\theta \in [1, \infty]$ and $r \in \mathbb{R}_+$. If a function $g : [0, \theta] \to \mathbb{R}_+$ satisfies $\mathbb{E}[g(E)] \leq r$ for all binary e-variables $E$ taking values in $[0, \theta]$, then there exists $h \in [0, 1]$ such that $g(x) \leq r(1 - h + hx)$ for $x \in [0, \theta] \cap \mathbb{R}$.*

As a particular case of Lemma 2 with $r = 1$, if $g$ is an e-merging function of dimension 1, then there exists $h \in [0, 1]$ such that $g(x) \leq (1 - h) + hx = M_{(h,1-h)}(x)$ for $x \in \mathbb{R}_+$. Hence, Theorem 1 holds for $K = 1$. This result, although very simple, is not found in the literature.

Lemma 2 is a refinement of Lemma EC.2 of Wang et al. (2025); the latter result considered all e-variables instead of those binary and bounded in $[0, \theta]$. Because of its importance for our main result, a self-contained proof, similar to the one in Wang et al. (2025), is presented in Section 4.

*Remark* 3. For $\alpha \in [0, 1]$, a level-$\alpha$ test for a hypothesis $\mathcal{H}$ is a $[0, 1]$-valued random variable $\tau$ satisfying $\mathbb{E}^{\mathbb{P}}[\tau] \leq \alpha$ for all $\mathbb{P} \in \mathcal{H}$, with $\tau = 1$ indicating a rejection, $\tau = 0$ indicating no rejection, and $\tau \in (0, 1)$ indicating a randomized decision. For $\alpha \neq 0$, $\tau/\alpha$ is an e-variable. For the connection between tests and e-values, see Koning (2024) and Ramdas and Wang (2024). Applying the proof of Theorem 1 to e-variables that are bounded in $[0, 1/\alpha]$, we get that using a weighted average is the only admissible way to merge arbitrary level-$\alpha$ tests. Merging tests at different levels is similar, up to an adjustment of the weights.

# 3 Proof of Theorem 1

Let us first prove a small lemma. In what follows, $a \vee b$ is the maximum of $a, b$, $a \wedge b$ is the minimum of $a, b$, and $\mathbf{0}$ represents a zero vector of the appropriate dimension.

**Lemma 4.** *Any e-merging function $F$ satisfies $F(\mathbf{e}) \le 1 \vee \max(\mathbf{e})$ for all $\mathbf{e} \in \mathbb{R}_+^K$.*

*Proof.* Suppose that there exists $\mathbf{e} \in \mathbb{R}_+^K$ such that $F(\mathbf{e}) > 1 \vee \max(\mathbf{e})$. Let $\bar{e} = \max\{\mathbf{e}\}$. If $\bar{e} \le 1$, then $\mathbf{e}$ is a vector of constant e-variables, but $F(\mathbf{e}) > 1$ is not an e-variable, a contradiction. Next, consider $\bar{e} > 1$. Take e-variables $E_1, \dots, E_K$ with $\mathbb{P}((E_1, \dots, E_K) = \mathbf{e}) = 1/\bar{e}$ and $\mathbb{P}((E_1, \dots, E_K) = \mathbf{0}) = 1 - 1/\bar{e}$. It is straightforward to see that $E_1, \dots, E_K$ are e-variables. Moreover, $\mathbb{E}[F(E_1, \dots, E_K)] > \bar{e}/\bar{e} = 1$, a contradiction to the assumption that $F$ is an e-merging function. □

Fix an e-merging function $F$. Lemma 4 implies in particular

$$F(e_1, \dots, e_K) \le 1 + \sum_{k=1}^{K} e_k \qquad \text{for all } (e_1, \dots, e_K) \in \mathbb{R}_+^K. \tag{1}$$

This allows us to apply optimal transport duality. Let us state the duality first. Let $\Pi(\mu_1, \dots, \mu_K)$ be the set of all Borel measures on $\mathbb{R}^K$ with marginal distributions $\mu_1, \dots, \mu_K$ on $\mathbb{R}$. Let $\mathcal{B}$ be the set of Borel-measurable functions on $\mathbb{R}_+$, and write $\bigoplus_{k=1}^K \phi_k : (x_1, \dots, x_K) \mapsto \sum_{k=1}^K \phi_k(x_k)$ for functions $\phi_1, \dots, \phi_K \in \mathcal{B}$. Define

$$D_F = \left\{ (\phi_1, \dots, \phi_K) \in \mathcal{B}^K : \bigoplus_{k=1}^{K} \phi_k \ge F \right\},$$

that is, the set of tuples of univariate functions whose sum dominates $F$. The set $D_F$ is important in optimal transport theory because it compares $F$ with a *separable* function $G$ of the from $G = \bigoplus_{k=1}^K \phi_k$, with the key property that $\mathbb{E}[G(E_1, \dots, E_K)]$ depends only on the marginals of $(E_1, \dots, E_K)$ but not the dependence structure.

Let $\mathcal{M}_{\mathcal{E}}$ be the set of all distributions on $\mathbb{R}_+$ with mean no larger than 1, i.e., the set of all distributions of e-variables, and $\mathcal{M}_{\mathcal{E}}^{\theta}$ be the subset of $\mathcal{M}_{\mathcal{E}}$ containing all distributions on $[0, \theta]$ for $\theta \ge 1$. In what follows, we always write $\boldsymbol{\phi} = (\phi_1, \dots, \phi_K)$ and $\boldsymbol{\mu} = (\mu_1, \dots, \mu_K)$.

Condition (1) implies that the set $D_F$ is not empty, and moreover some elements $\boldsymbol{\phi}$ in $D_F$ satisfies $\sum_{k=1}^K \int \phi_k \, \mathrm{d}\mu_k < \infty$ when $\boldsymbol{\mu} \in (\mathcal{M}_{\mathcal{E}})^K$. With this, optimal transport duality holds.

**Lemma 5** (Optimal transport duality)**.** *For $\mu_1, \dots, \mu_K \in \mathcal{M}_{\mathcal{E}}$ and an e-merging function $F$, we have*

$$1 \geq \sup_{\pi \in \Pi(\boldsymbol{\mu})} \int F \, \mathrm{d}\pi = \inf_{\boldsymbol{\phi} \in D_F} \sum_{k=1}^{K} \int \phi_k \, \mathrm{d}\mu_k, \tag{2}$$

*where in the infimum we only consider those with $\sum_{k=1}^{K} \int \phi_k \, \mathrm{d}\mu_k$ well-defined.*

The inequality in (2) follows because $F$ is an e-merging function, and the equality is a classic form of optimal transport duality in the form of Remark 2.1.2 of Rachev and Rüschendorf (1998).

To interpret (2), let $E_1, \dots, E_K$ be e-variables with distributions $\mu_1, \dots, \mu_K$. Since $F \leq \bigoplus_{k=1}^{K} \phi_k$ for $\boldsymbol{\phi} \in D_F$, the equality in (2) means that the supremum of $\mathbb{E}[F(E_1, \dots, E_K)]$ over dependence structures of $(E_1, \dots, E_K)$ is equal to the infimum of $\mathbb{E}[\sum_{k=1}^{K} \phi_k(E_k)]$ over $\boldsymbol{\phi}$.

We explain our main idea in the proof first. The inequality (2) implies

$$\sup_{\boldsymbol{\mu} \in (\mathcal{M}_{\mathcal{E}})^K} \inf_{\boldsymbol{\phi} \in D_F} \sum_{k=1}^{K} \int \phi_k \, \mathrm{d}\mu_k \leq 1.$$

If we could interchange the order of sup and inf in the above equation, then for any $\epsilon > 0$, $\sum_{k=1}^{K} \int \phi_k \, \mathrm{d}\mu_k \leq 1 + \epsilon$ for some $\boldsymbol{\phi}$ and all $\boldsymbol{\mu} \in (\mathcal{M}_{\mathcal{E}})^K$. Using this and Lemma 2 yields linear upper bounds on $\boldsymbol{\phi}$, with which we can further use $F \leq \bigoplus_{k=1}^{K} \phi_k$ to show $F \leq M_{\boldsymbol{\lambda}}$ for some $\boldsymbol{\lambda}$. The following lemma gives the desired maximin interchangeability under the assumption of bounded compact. Let $\mathcal{M}_0$ be the set of all distributions on $\mathbb{R}_+$.

**Lemma 6.** *Suppose $\mathcal{M} \subseteq (\mathcal{M}_0)^K$ is compact with respect to weak convergence and $F : \mathbb{R}_+^K \to \mathbb{R}_+$ is bounded and upper semicontinuous. Then*

$$\sup_{\boldsymbol{\mu} \in \mathcal{M}} \inf_{\boldsymbol{\phi} \in D_F} \sum_{k=1}^{K} \int \phi_k \, \mathrm{d}\mu_k = \inf_{\boldsymbol{\phi} \in D_F^+} \sup_{\boldsymbol{\mu} \in \mathcal{M}} \sum_{k=1}^{K} \int \phi_k \, \mathrm{d}\mu_k, \tag{3}$$

*where $D_F^+$ is the subset of $D_F$ containing all $\boldsymbol{\phi}$ with nonnegative components.*

We will prove Lemma 6 later. We now use this lemma to complete the proof of Theorem 1. Fix $\theta \in [1, \infty)$. We need to verify a few things for an e-merging function $F$, which justifies that we can apply Lemma 6.

(i) Since $[0, \theta]$ is compact, the set $(\mathcal{M}_{\mathcal{E}}^{\theta})^K$ equipped with the weak topology is tight. By Prokhorov's theorem, it is sequentially compact, and hence compact.

(ii) Since $F$ is bounded on $[0, \theta]^K$ and its value outside $[0, \theta]^K$ does not matter in (3), we can treat $F$ as bounded when applying Lemma 6.

(iii) There exists an upper semicontinuous e-merging function $F^*$ with $F^* \geq F$ by Proposition E.1 of Vovk and Wang (2021). Therefore, to prove Theorem 1, it suffices to consider upper semicontinuous e-merging functions.

By using Lemmas 5 and 6 with $\mathcal{M} = (\mathcal{M}_{\mathcal{E}}^{\theta})^K$ and an upper semicontinuous e-merging function $F$, we have

$$1 \geq \inf_{\boldsymbol{\phi} \in D_F^+} \sup_{\boldsymbol{\mu} \in (\mathcal{M}_{\mathcal{E}}^{\theta})^K} \sum_{k=1}^{K} \int \phi_k \, \mathrm{d}\mu_k = \inf_{\boldsymbol{\phi} \in D_F^+} \sum_{k=1}^{K} \sup_{\mu_k \in \mathcal{M}_{\mathcal{E}}^{\theta}} \int \phi_k \, \mathrm{d}\mu_k. \tag{4}$$

Denote by $T_\phi = \sup_{\mu \in \mathcal{M}_{\mathcal{E}}^{\theta}} \int \phi \, \mathrm{d}\mu$ for $\phi \in \mathcal{B}$. For any $\epsilon > 0$, by (4), we can find $(\phi_1, \dots, \phi_K) \in D_F^+$ such that $\sum_{k=1}^{K} T_{\phi_k} \leq 1 + \epsilon$. For each $k \in [K]$, by applying Lemma 2 to $\phi_k$, there exists a constant $h_{\phi_k} \in [0,1]$ such that

$$\phi_k(x) \leq T_{\phi_k} \left(1 - h_{\phi_k} + h_{\phi_k} x\right) \text{ for all } x \in [0, \theta].$$

Since $(\phi_1, \dots, \phi_K) \in D_F^+$, we have

$$F(x_1, \dots, x_K) \leq \sum_{k=1}^{K} T_{\phi_k} \left(1 - h_{\phi_k} + h_{\phi_k} x_k\right) \text{ for all } x_1, \dots, x_K \in [0, \theta].$$

This means that there exists $\boldsymbol{\lambda}_{\epsilon,\theta} \in \Delta_{K+1}$ such that $F \leq (1+\epsilon) M_{\boldsymbol{\lambda}_{\epsilon,\theta}}$ on $[0,\theta]^K$. Since $\Delta_{K+1}$ is compact, we can find a convergent subsequence of $\boldsymbol{\lambda}_{\epsilon,\theta}$ by sending $\epsilon \downarrow 0$ and $\theta \to \infty$, with its limit denoted by $\boldsymbol{\lambda}_0 \in \Delta_{K+1}$. Continuity of $\boldsymbol{\lambda} \mapsto M_{\boldsymbol{\lambda}}$ yields $F \leq M_{\boldsymbol{\lambda}_0}$ on $\mathbb{R}_+^K$, thus the desired statement. □

*of Lemma 6.* We assume $0 \leq F \leq 1$ without loss of generality. We first argue that we can require $(\phi_1, \dots, \phi_K) \in D_F$ to be upper semicontinuous and take values in $[0,1]$. That is, we will show, for $(\mu_1, \dots, \mu_K) \in \mathcal{M}$,

$$\inf_{\boldsymbol{\phi} \in D_F} \sum_{k=1}^{K} \int \phi_k \, \mathrm{d}\mu_k = \inf_{\boldsymbol{\phi} \in \widetilde{D}_F} \sum_{k=1}^{K} \int \phi_k \, \mathrm{d}\mu_k, \tag{5}$$

where

$$\widetilde{D}_F = \left\{ \boldsymbol{\phi} \in D_F : \ \phi_k \text{ is upper semicontinuous and } 0 \leq \phi_k \leq 1 \text{ for } k \in [K] \right\}.$$

(a) For $(\phi_1, \dots, \phi_K) \in D_F$, denote by $c_k = \inf_{x \in \mathbb{R}_+} \phi_k(x)$ for $k \in [K]$. Since $\sum_{k=1}^{K} \phi_k(x_k) \geq F(x_1, \dots, x_K) \geq 0$ for all $(x_1, \dots, x_K) \in \mathbb{R}_+^K$, the function $\tilde{\phi}_k : x \mapsto \phi_k(x) - c_k + \sum_{m=1}^{K} c_m / K$ is nonnegative. Moreover, for all $(\mu_1, \dots, \mu_K) \in \mathcal{M}$,

$$\sum_{k=1}^{K} \int \tilde{\phi}_k \, \mathrm{d}\mu_k = \sum_{k=1}^{K} \int \phi_k \, \mathrm{d}\mu_k.$$

Hence, we can safely restrict the functions $\phi_1, \dots, \phi_K$ to be nonnegative.

(b) For $(\mu_1, \dots, \mu_K) \in \mathcal{M}$, since $F \leq 1$, together with (a) we can also safely truncate $\phi_1, \dots, \phi_K$ at 1.

(c) Since $F$ is upper semicontinuous and bounded, by Proposition 1.31 of Kellerer (1984) the functions $(\phi_1, \dots, \phi_K)$ can be chosen as upper semicontinuous.

The above three arguments together show the assertion (5).

Next, we verify

$$\sup_{\boldsymbol{\mu} \in \mathcal{M}} \inf_{\boldsymbol{\phi} \in \widetilde{D}_F} \sum_{k=1}^{K} \int \phi_k \, \mathrm{d}\mu_k = \inf_{\boldsymbol{\phi} \in \widetilde{D}_F} \sup_{\boldsymbol{\mu} \in \mathcal{M}} \sum_{k=1}^{K} \int \phi_k \, \mathrm{d}\mu_k. \tag{6}$$

To apply the minimax theorem of Sion (1958), we need to check a few things. Define the mapping

$$J : (\boldsymbol{\mu}, \boldsymbol{\phi}) \mapsto \sum_{k=1}^{K} \int \phi_k \, \mathrm{d}\mu_k.$$

(i) The mapping $J$ is bilinear, and therefore both convex and concave.

(ii) The set $\mathcal{M}$ equipped with the weak topology is compact.

(iii) Since each component of $\boldsymbol{\phi}$ is bounded in $[0, 1]$, $\boldsymbol{\phi} \mapsto J(\boldsymbol{\mu}, \boldsymbol{\phi})$ is continuous with respect to point-wise convergence in $\boldsymbol{\phi}$ for each $\boldsymbol{\mu}$. This continuity is not needed if we use the minimax theorem of Fan (1953, Theorem 2).

(iv) Since each component of $\boldsymbol{\phi}$ is upper semicontinuous, $\boldsymbol{\mu} \mapsto J(\boldsymbol{\mu}, \boldsymbol{\phi})$ is upper semicontinuous with respect to weak convergence in $\boldsymbol{\mu}$ for each $\boldsymbol{\phi}$.

Therefore, Sion's minimax theorem implies that (6) holds. Putting (5) and (6) together we get

$$\sup_{\boldsymbol{\mu} \in \mathcal{M}} \inf_{\boldsymbol{\phi} \in D_F} J(\boldsymbol{\mu}, \boldsymbol{\phi}) = \inf_{\boldsymbol{\phi} \in \widetilde{D}_F} \sup_{\boldsymbol{\mu} \in \mathcal{M}} J(\boldsymbol{\mu}, \boldsymbol{\phi}).$$

Finally, by using $\widetilde{D}_F \subseteq D_F^+ \subseteq D_F$ we obtain (3). □

# 4 Proof of Lemma 2

*Proof.* The case $r = 0$ is trivial as $g$ is the constant function 0. Otherwise, $r > 0$, and without loss of generality we can assume $r = 1$.

If $\theta = 1$, then $g(x) \leq 1$ for all $x$, and taking $h = 0$ gives the desired inequality. In what follows, we assume $\theta \in (1, \infty]$, and all points $x, y$ that appear below are in $[0, \theta] \cap \mathbb{R}$.

First, it is easy to note that $g(y) \leq y$ for $y > 1$; indeed, if $g(y) > y$ then taking a random variable $X$ with $\mathbb{P}(X = y) = 1/y$ and 0 otherwise gives $\mathbb{E}[g(X)] > 1$ and breaks the assumption. Moreover, $g(y) \leq 1$ for $y \leq 1$ is also clear, which in particular implies $g(1) \leq 1$.

Suppose for the purpose of contradiction that the statement in the lemma does not hold. This means that for each $\lambda \in [0,1]$, either (a) $g(x) > (1-\lambda)+\lambda x$ for some $x < 1$ or (b) $g(y) > (1-\lambda) + \lambda y$ for some $y > 1$ (or both). Since $g(y) \le y$ for $y > 1$ and $g(x) \le 1$ for $x < 1$, we know that $\lambda = 1$ implies (a) and $\lambda = 0$ implies (b).

We claim that there exists $\lambda_0 \in (0,1)$ for which both (a) and (b) happen. To show this claim, let

$$\Lambda_0 = \{\lambda \in [0,1] : g(y) > (1-\lambda) + \lambda y \text{ for some } y > 1\};$$
$$\Lambda_1 = \{\lambda \in [0,1] : g(x) > (1-\lambda) + \lambda x \text{ for some } x < 1\}.$$

Clearly, the above arguments show $\Lambda_0 \cup \Lambda_1 = [0,1]$, $0 \in \Lambda_0$, and $1 \in \Lambda_1$. Moreover, since the function $\lambda \mapsto (1-\lambda) + \lambda x$ is monotone for either $x < 1$ or $x > 1$, we know that both $\Lambda_0$ and $\Lambda_1$ are intervals. Let $\lambda_* = \sup \Lambda_0$ and $\lambda^* = \inf \Lambda_1$. We will argue $\lambda_* \notin \Lambda_0$ and $\lambda^* \notin \Lambda_1$. If $\lambda_* \in \Lambda_0$, then there exists $y > 1$ such that $g(y) > (1-\lambda_*) + \lambda_* y$. By continuity, there exists $\hat{\lambda}_* > \lambda_*$ such that $g(y) > (1-\hat{\lambda}_*) + \hat{\lambda}_* y$, showing that $\hat{\lambda}_* \in \Lambda_0$, contradicting the definition of $\lambda_*$. Therefore, $\lambda_* \notin \Lambda_0$. Similarly, $\lambda^* \notin \Lambda_1$, following the same argument. If $\lambda_* = \lambda^*$, then this point is not contained in $\Lambda_0 \cup \Lambda_1$, a contradiction to $\Lambda_0 \cup \Lambda_1 = [0,1]$. Hence, it must be $\lambda_* > \lambda^*$, which implies that $\Lambda_0 \cap \Lambda_1$ is not empty.

Let $x_0 < 1$ and $y_0 > 1$ be such that

$$g(x_0) > 1 - \lambda_0 + \lambda_0 x_0 \quad \text{and} \quad g(y_0) > 1 - \lambda_0 + \lambda_0 y_0.$$

Let $X$ be distributed as $\mathbb{P}(X = y_0) = (1-x_0)/(y_0 - x_0)$ and $\mathbb{P}(X = x_0) = (y_0 - 1)/(y_0 - x_0)$, which clearly satisfies $\mathbb{E}[X] = 1$ and is binary. Moreover,

$$\begin{aligned}\mathbb{E}[g(X)] &= \frac{1-x_0}{y_0-x_0} g(y_0) + \frac{y_0-1}{y_0-x_0} g(x_0) \\ &> \frac{1-x_0}{y_0-x_0}\left(1-\lambda_0+\lambda_0 y_0\right) + \frac{y_0-1}{y_0-x_0}\left(1-\lambda_0+\lambda_0 x_0\right) = 1.\end{aligned}$$

This yields a contradiction. □

# 5 Merging functions of subclasses of e-values

Theorem 1 focuses on merging arbitrary e-variables. In an application, if some information on the e-variables is available, one may consider merging functions that are valid for a subclass of e-variables, but not necessarily valid for arbitrary e-variables. For many subclasses, every weighted average $M_{\boldsymbol{\lambda}}$ for $\boldsymbol{\lambda} \in \Delta_{K+1}$ remains admissible, but they may not be the only admissible choices.

Formally, let $\mathcal{E}_K$ be a set of $K$-dimensional vectors of e-variables. A *merging function of the subclass* $\mathcal{E}_K$ is an increasing Borel function $F : \mathbb{R}_+^K \to \mathbb{R}_+$ such that $F(E_1, \dots, E_K)$ is an e-variable for any $(E_1, \dots, E_K) \in \mathcal{E}_K$. A merging function $F$ of $\mathcal{E}_K$ is *admissible* if for any merging function $G$ of $\mathcal{E}_K$, $G \ge F$

implies $G = F$. The cases of independent and sequential e-variables are studied by Vovk and Wang (2021, 2024). We consider four other subclasses below.

We first make a general observation. Let $\mathbf{E}$ be a vector of e-variables with mean 1. Clearly, $\mathbb{E}[M_{\boldsymbol{\lambda}}(\mathbf{E})] = 1$ for $\boldsymbol{\lambda} \in \Delta_{K+1}$. Since $M_{\boldsymbol{\lambda}}$ is continuous, any $G \geq M_{\boldsymbol{\lambda}}$ with $G \neq M_{\boldsymbol{\lambda}}$ satisfies $\mathbb{E}[G(\mathbf{E})] > 1$ if $\mathbf{E}$ has full support in $\mathbb{R}_+^K$, and $G$ cannot be a valid merging function of $\{\mathbf{E}\}$. Therefore, if the subclass $\mathcal{E}_K$ contains an element of full support in $\mathbb{R}_+^K$, then $M_{\boldsymbol{\lambda}}$ is admissible.

(a) **Fixed Marginals.** Let $\mathcal{E}_K$ be the set of vectors of e-variables with given marginals $\mu_1, \dots, \mu_K$. If all marginals have full support in $\mathbb{R}_+$, then $M_{\boldsymbol{\lambda}}$ is admissible. To see other admissible merging functions of $\mathcal{E}_K$, suppose that the marginals $\mu_1, \dots, \mu_K$ are continuous with full support on $\mathbb{R}_+$ and let $g_1, \dots, g_K$ be the survival functions of $\mu_1, \dots, \mu_K$. Let $f_1, \dots, f_K$ be admissible calibrators; that is, decreasing upper semicontinuous functions $f_k : [0, 1] \to [0, \infty]$ satisfying $f_k(0) = \infty$ and $\int_0^1 f_k(x)\,\mathrm{d}x = 1$. For each $k \in [K]$ and $E_k$ distributed as $\mu_k$, $g_k(E_k)$ is uniformly distributed on $[0, 1]$, and $f_k(g_k(E_k))$ is an e-variable. Therefore,

$$(e_1, \dots, e_K) \mapsto M_{\boldsymbol{\lambda}}(f_1 \circ g_1(e_1), \dots, f_K \circ g_K(e_K))$$

is an admissible merging function of $\mathcal{E}_K$, following the general observation above.

(b) **Bounded second moments.** Let $\mathcal{E}_K$ be the set of vectors $(E_1, \dots, E_K)$ of e-variables with second moments $\mathbb{E}[X_i X_j]$ bounded above by $\sigma_{ij} \in [0, \infty)$ for $i, j \in [K]$. In addition to the weighted averages, the functions $(e_1, \dots, e_K) \mapsto e_i e_j / \sigma_{ij}$ and their mixtures are merging functions of $\mathcal{E}_K$. As a special case, if elements of $\mathcal{E}_K$ have nonpositive bivariate correlation coefficients (a form of negative dependence), then the functions $(e_1, \dots, e_K) \mapsto e_i e_j$ with $i \neq j$ are merging functions of $\mathcal{E}_K$.

(c) **Identical e-variables.** Let $\mathcal{E}_K$ be the set of vectors of identical e-variables. Clearly, the function $V_K : \mathbf{e} \mapsto \max(\mathbf{e})$ is a merging function of the subclass $\mathcal{E}_K$. Moreover, all admissible merging functions in this setting have the form $\lambda + (1 - \lambda)V_K$ for $\lambda \in [0, 1]$. To see this, first note that for any merging function $F$ of $\mathcal{E}_K$, the function $g : e \mapsto F(e, \dots, e)$ satisfies the condition in Lemma 2, and hence for some $\lambda \in [0, 1]$ $g(e) \leq \lambda + (1 - \lambda)e$ for all $e \in \mathbb{R}_+$. It follows that $F(\mathbf{e}) \leq g(\max(\mathbf{e})) \leq \lambda + (1 - \lambda)\max(\mathbf{e})$ for all $\mathbf{e} \in \mathbb{R}_+^K$. Note that when applied to a vector $\mathbf{E}$ of identical e-variables, $V_K(\mathbf{E}) = M_{\boldsymbol{\lambda}}(\mathbf{E})$ for any $\boldsymbol{\lambda}$ with the final component 0, so using $\lambda + (1 - \lambda)V_K$ is the same as using weighted averages.

(d) **Exchangeable e-variables.** Let $\mathcal{E}_K$ be the set of vectors of exchangeable e-variables. For $(E_1, \dots, E_K) \in \mathcal{E}_K$, let $\bar{E}_k = (E_1 + \cdots + E_k)/k$ for $k \in [K]$. By the randomized Markov inequality in Ramdas and Wang (2024, Theorem 4.5),

$$\mathbb{E}\left[\mathbb{1}_{\{\bigvee_{k \in [K]} \bar{E}_k \geq \beta\}}\right] \leq 1/\beta \quad \text{for all } \beta > 1.$$

Hence, $F_\beta : (e_1, \dots, e_K) \mapsto \beta \mathbb{1}_{\{\bigvee_{k \in [K]} \bar{e}_k \geq \beta\}}$ for $\beta > 1$ is an e-merging function of $\mathcal{E}_K$. It is straightforward to check that neither $F_\beta \geq M_{\boldsymbol{\lambda}}$ nor $F_\beta \leq M_{\boldsymbol{\lambda}}$ for any $\beta > 0$ and $\boldsymbol{\lambda} \in \Delta_{K+1}$. It remains unclear whether $F_\beta$ is admissible.

## Acknowledgements

The author would like to thank the editors, two anonymous referees for useful comments on the paper, and Jose Blanchet, Qinyu Wu, Vladimir Vovk, and Zhenyuan Zhang for helpful discussions on the main result. The work is financially supported by the Natural Sciences and Engineering Research Council of Canada (RGPIN-2024-03728, CRC-2022-00141).

# References

Barber, R. F. and Candès, E. (2015). Controlling the false discovery rate via knockoffs. *Annals of Statistics*, **43**(5), 2055–2085.

Fan, K. (1953). Minimax theorems. *Proceedings of the National Academy of Sciences*, **39**(1), 42–47.

Grünwald, P. (2024). Beyond Neyman–Pearson: E-values enable hypothesis testing with a data-driven alpha. *Proceedings of the National Academy of Sciences*, **121**(39), e2302098121.

Grünwald, P., de Heide, R. and Koolen, W. M. (2024). Safe testing. *Journal of the Royal Statistical Society Series B*, forthcoming.

Kellerer, H. G. (1984). Duality theorems for marginal problems. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **67**, 399–432.

Koning, N. (2024). Continuous testing: Unifying tests and e-values. *arXiv*: 2409.05654v6.

Lee, J. and Ren, Z. (2024). Boosting e-BH via conditional calibration. *arXiv*: 2404.17562.

Rachev, S. T., and Rüschendorf, L. (1998). *Mass Transportation Problems: Volume I: Theory.* Springer Science & Business Media.

Ramdas, A., Grünwald, P., Vovk, V. and Shafer, G. (2023). Game-theoretic statistics and safe anytime-valid inference. *Statistical Science*, **38**(4), 576–601.

Ramdas, A. and Wang, R. (2024). Hypothesis testing with e-values. *arXiv*: 2410.23614v2.

Ren, Z. and Barber, R. F. (2024). Derandomized knockoffs: Leveraging e-values for false discovery rate control. *Journal of the Royal Statistical Society Series B*, **86**(1), 122–154.

Shafer, G. (2021). The language of betting as a strategy for statistical and scientific communication. *Journal of the Royal Statistical Society, Series A*, 184(2), 407–431.

Sion, M. (1958). On general minimax theorems. *Pacific Journal of Mathematics*, **8**(1), 171–176.

Villani, C. (2009). *Optimal Transport: Old and New*. Springer-Verlag, Berlin.

Vovk, V., Wang, B. and Wang, R. (2022). Admissible ways of merging p-values under arbitrary dependence. *Annals of Statistics*, **50**(1), 351–375.

Vovk, V. and Wang, R. (2021). E-values: Calibration, combination, and applications. *Annals of Statistics*, **49**(3), 1736–1754.

Vovk, V. and Wang, R. (2023). Confidence and discoveries with e-values. *Statistical Science*, **38**(2), 329–354.

Vovk, V. and Wang, R. (2024). Merging sequential e-values via martingales. *Electronic Journal of Statistics*, **18**(1), 1185–1205.

Wang, Q., Wang, R. and Ziegel, J. (2025). E-backtesting. *Management Science*, forthcoming.

Wang, R. and Ramdas, A. (2022). False discovery rate control with e-values. *Journal of the Royal Statistical Society Series B*, **84**(3), 822–852.

Wasserman, L., Ramdas, A. and Balakrishnan, S. (2020). Universal inference. *Proceedings of the National Academy of Sciences*, **117**(29), 16880–16890.