Combining p-values under dependence and optimal transport

Ruodu Wang

http://sas.uwaterloo.ca/~wang

Department of Statistics and Actuarial Science
University of Waterloo

Wharton Department of Statistics and Data Science
University of Pennsylvania, September 4, 2024

Agenda

This talk

Vladimir Vovk
(Royal Holloway)

Bin Wang
(CAS Beijing)

Matteo Gasparin
(Padova)

Aaditya Ramdas
(Carnegie Mellon)

- Vovk/W.'20, Combining p-values via averaging — Biometrika
- Vovk/Wang/W.'22, Admissible ways of merging p-values under arbitrary dependence — Annals of Statistics
- Gasparin/W./Ramdas'24, Combining exchangeable p-values — arXiv:2404.03484

Also based on 10+ years
of collaboration with
Paul Embrechts (ETH)

Background

Suppose that we are testing the global null hypothesis using $K \geqslant 2$ tests and obtain p-values $p_1, \ldots, p_K$. How can we combine them into a single p-value?

A question of a long history

- Tippett'31, Pearson'33, Fisher'48: assume independence
- We are primarily interested in the case of dependence, in particular, unknown dependence
- The Bonferroni correction: minimum $\times$ correction (K)
- More important is testing multiple hypotheses or selective inference, but for now let us focus on the global null

The value of no assumption

Dependence

- A set of p-values is essentially one vector: difficult to test/verify any dependence model among them
- In modern applications, multiple p-values may come from resampling, data splitting, overlapping experiments, different algorithms, ...
- Efron'10, Large-scale Inference, p50-p51:
  "*independence among the p-values ... usually an unrealistic assumption. ... even PRD* [positive regression dependence] *is unlikely to hold in practice.*"

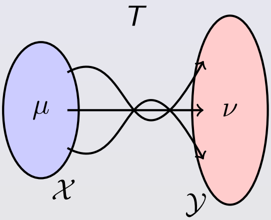
Optimal transport

The Monge (1781) problem: find a transport map $T : \mathcal{X} \to \mathcal{Y}$ that attains

$$\inf \left\{ \int_{\mathcal{X}} c(x, T(x))\, \mathrm{d}\mu(x) \;\middle|\; T_{\#}\mu = \nu \right\}$$

where

- $\mathcal{X}$ and $\mathcal{Y}$ are two Polish spaces (e.g., $\mathbb{R}^d$)
- Cost function $c : \mathcal{X} \times \mathcal{Y} \to [0, \infty]$ or $(-\infty, \infty]$
- Probabilities μ on $\mathcal{X}$ and ν pm $\mathcal{Y}$ are given
- $T_{\#}\mu = \mu \circ T^{-1}$ is the push forward of μ by T
- Such T is an optimal transport map



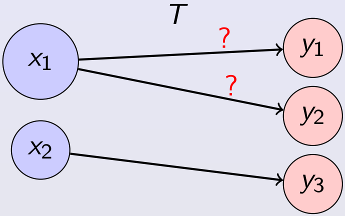
Kantorovich's formulation

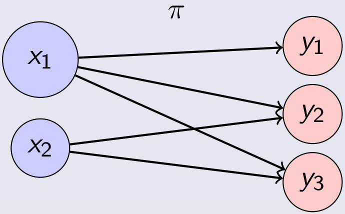
- Monge's formulation may be ill-posed (e.g., point masses)
- Kantorovich's problem: find a probability measure $\pi \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ that attains

$$\inf \left\{ \int_{\mathcal{X}\times\mathcal{Y}} c(x,y)\, \pi(\mathrm{d}x, \mathrm{d}y) \mid \pi \in \Pi(\mu,\nu) \right\},$$

  where $\Pi(\mu,\nu)$ is the set of probability measures on $\mathcal{X} \times \mathcal{Y}$ with marginals μ and ν
- Wasserstein distance is defined this way
- Linear programming

Monge and Kantorovich's formulations





Gaspard Monge
1746-1818

Leonid Kantorovich
1912-1986

Transport duality

If c is non-negative and lower semi-continuous, then duality holds

$$\min_{\pi \in \Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{Y}} c \,\mathrm{d}\pi = \sup \left(\int_{\mathcal{X}} \phi \,\mathrm{d}\mu + \int_{\mathcal{Y}} \psi \,\mathrm{d}\nu \right),$$

where the supremum runs over all pairs of bounded and continuous functions $\phi : \mathcal{X} \to \mathbb{R}$ and $\psi : \mathcal{Y} \to \mathbb{R}$ such that

$$\phi(x) + \psi(y) \leqslant c(x, y) \quad \text{for all } x, y$$

Economic interpretation

- $x \in \mathcal{X}$: the vector of characteristics of a worker
- $y \in \mathcal{Y}$: the vector of characteristics of a firm
- $g(x, y)$ the economic output (production) generated by worker x matched with firm y
- Social economic-output maximization

$$\sup \left\{ \int_{\mathcal{X}\times\mathcal{Y}} g \, \mathrm{d}\pi \mid \pi \in \Pi(\mu, \nu) \right\}$$

- Dual problem $g(x, y) \leqslant \phi(x) + \psi(y)$: social equilibrium
  - ϕ: the equilibrium wage function
  - ψ: the equilibrium profit function

Probabilistic formulation

Let L and R represent random variables on $\mathbb{R}$

- Classic optimal transport

$$\inf_{L\sim\mu, R\sim\nu} \mathbb{E}[c(L,R)]$$

- Martingale optimal transport require: $\mu \preceq_{\mathrm{cx}} \nu$

$$\inf_{L\sim\mu, R\sim\nu} \mathbb{E}[c(L,R)] : \; L = \mathbb{E}[R|L]$$

- Submartingale optimal transport require: $\mu \preceq_{\mathrm{icx}} \nu$

$$\inf_{L\sim\mu, R\sim\nu} \mathbb{E}[c(L,R)] : \; L \leqslant \mathbb{E}[R|L]$$

- Directional optimal transport require: $\mu \preceq_{\mathrm{st}} \nu$

$$\inf_{L\sim\mu, R\sim\nu} \mathbb{E}[c(L,R)] : \; L \leqslant R$$

Martingale: Beiglböck/Henry-Labordère/Penkner'13; Beiglböck/Juillet'16
Submartingale: Nutz/Stebegg'18; Directional: Nutz/W.'22

The multi-marginal formulation

Focus on $\mathbb{R}$ in what follows

- $\mu_1, \ldots, \mu_n$: n probability measures on $\mathbb{R}$
- $\Pi(\mu_1, \ldots, \mu_n)$: the set of probability measures on $\mathbb{R}^n$ with marginals $\mu_1, \ldots, \mu_n$
- $X_1, \ldots, X_n$: random variables ⇐ now probabilistic notation ...

Kantorovich's formulation:

$$\inf \left\{ \int_{\mathbb{R}^n} c(x_1, \ldots, x_n)\, \pi(\mathrm{d}x_1, \ldots, \mathrm{d}x_n) \mid \pi \in \Pi(\mu_1, \ldots, \mu_n) \right\}$$

Probabilistic formulation:

$$\inf \{\mathbb{E}[c(X_1, \ldots, X_n)] \mid X_i \sim \mu_i,\ i \in [n]\}$$

Rüschendorf'13; Pass'15

Dependence uncertainty problems

A random vector $\mathbf{X} = (X_1, \ldots, X_n)$

Assumptions

marginals may be known; dependence is unknown/arbitrary

- Motivation: data scarcity; uncertainty; absent information; lack of models; conservative decisions

Questions:

- properties of $\Psi(\mathbf{X})$ for some $\Psi : \mathbb{R}^n \to \mathbb{R}^d$
- range of $\mathbb{P}(\mathbf{X} \in A)$ for some $A \subseteq \mathbb{R}^n$
- "optimal" dependence structures of $\mathbf{X}$
- statistical decisions based on $\mathbf{X}$

Risk management and dependence uncertainty

Basic setup.

- A vector of risk factors: $\mathbf{X} = (X_1, \ldots, X_n)$
- A financial position $\Psi(\mathbf{X})$
- A risk measure ρ

Key task: Calculate $\rho(\Psi(\mathbf{X}))$

Most practical choices:

- $\Psi(\mathbf{X}) = \sum_{i=1}^{n} X_i$
- $\rho = \mathrm{VaR}_p$ (quantile) or $\rho = \mathrm{ES}_p$ (tail mean)

Challenge: We need a joint model for the random vector $\mathbf{X}$

- Range of $\rho(\Psi(\mathbf{X}))$ under dependence uncertainty

Risk management and dependence uncertainty

Misspecified dependence for credit risks $\Longrightarrow$ 07–09 financial crisis

- In the current Basel FRTB internal model approach, for market risk

$$\text{Capital Charge} = \lambda \mathrm{ES}_p \begin{pmatrix} \text{internal} \\ \text{model} \end{pmatrix} + (1-\lambda)\mathrm{ES}_p \begin{pmatrix} \text{worst-case} \\ \text{dependence} \end{pmatrix}$$

- Similar considerations in other regulatory frameworks, such as Solvency II

General formulation

Minimize or maximize

$$\rho(c(X_1, \ldots, X_n)) \text{ subject to } X_i \sim \mu_i,\ i \in [n]$$

Differences from the classic OT theory:

- ρ may be non-linear, such as VaR, ES and other risk measures
- The optimal value often matters more than the optimizer
- Even the linear case $c(x_1, \ldots, x_n) = \sum_{i=1}^n c_i(x_i)$ is interesting
- Optimizers usually depend on $\mu_1, \ldots, \mu_n$ in a complicated way
- No duality formulas

[many collaborations with P. Embrechts and coauthors]

Merging functions

Let $\mathcal{H}$ be a collection of probability measures ...

Definition (p-variables and merging functions)

(i) A p-variable is a random variable P that satisfies

$$\sup_{\mathbb{P}\in\mathcal{H}} \mathbb{P}(P \leqslant \varepsilon) \leqslant \varepsilon, \quad \varepsilon \in (0,1).$$

(ii) A p-merging function is an increasing Borel function $F : [0,\infty)^K \to [0,\infty)$ such that $F(P_1,\dots,P_K)$ is a p-variable for all p-variables $P_1,\dots,P_K$.

- Controlled type I error under arbitrary dependence

Merging functions

- $\mathcal{U}$: the set of all uniform $[0,1]$ random variables under $\mathbb{P}$

For an increasing Borel $F : [0,\infty)^K \to [0,\infty)$, equivalent are:

- F is a p-merging function w.r.t. some atomless collection $\mathcal{H}$
- F is a p-merging function w.r.t. all collections $\mathcal{H}$
- fixing atomless $\mathbb{P}$, $F(\mathbf{P})$ is a p-variable for all $\mathbf{P} \in \mathcal{U}^K$
- fixing atomless $\mathbb{P}$, for all $\varepsilon \in (0,1)$, $\overline{\mathbb{P}}(F \leqslant \varepsilon) \leqslant \varepsilon$, where

$$\overline{\mathbb{P}}(F \leqslant \varepsilon) = \sup\left\{\mathbb{P}(F(\mathbf{P}) \leqslant \varepsilon) : \mathbf{P} \in \mathcal{U}^K\right\}$$

- It suffices to consider $\mathcal{H} = \{\mathbb{P}\}$ for an atomless $\mathbb{P}$ and $\mathcal{U}^K$

Precise merging functions

- Multi-marginal OT problem:

$$\sup\{\mathbb{E}[\mathbb{1}_{\{F(\mathbf{P})\leqslant\varepsilon\}}] : \mathbf{P} \in \mathcal{U}^K\}$$

- The cost function is not submodular/supermodular
- Dependence assumption is available $\Longrightarrow$ constrained OT

Definition (precise merging functions)

A p-merging function F is precise if, for all $\varepsilon \in (0,1)$, $\overline{\mathbb{P}}(F \leqslant \varepsilon) = \varepsilon$.

Existing methods

- $p_{(1)}, \dots, p_{(K)}$ are the ascending order statistics
- The Bonferroni method

$$F_B(p_1, \dots, p_K) = K p_{(1)}$$

- Order-family (O-family) Rüger'78

$$G_{k,K} = (p_1, \dots, p_K) = \frac{K}{k} p_{(k)}$$

- Hommel–Simes Hommel'83

$$H_K(p_1, \dots, p_K) = \ell_K \bigwedge_{k=1}^{K} \frac{K}{k} p_{(k)}; \qquad \ell_K = \sum_{k=1}^{K} \frac{1}{k}$$

(also appearing in the BH procedure)

Merging p-values via averaging

- Generalized mean — Kolmogorov'30

$$M_{\phi,K}(p_1,\dots,p_K) = \phi^{-1}\left(\frac{\phi(p_1)+\cdots+\phi(p_K)}{K}\right),$$

where $\phi : [0,1] \to [-\infty,\infty]$ is continuous & strictly monotone

- M-family: for $r \in \mathbb{R} \setminus \{0\}$,

$$M_{r,K}(p_1,\dots,p_K) = \left(\frac{p_1^r+\cdots+p_K^r}{K}\right)^{1/r}$$

Limiting cases: geometric, minimum, maximum

- $\phi(x) = \tan((x-\frac{1}{2})\pi)$: Cauchy combination — Liu/Xie'20

Merging p-values via averaging

The arithmetic average $M_{1,K}(p_1, \ldots, p_K) = \frac{1}{K}\sum_{k=1}^{K} p_k$ is not a p-merging function Rüschendorf'82, Meng'93

$$\overline{\mathbb{P}}(M_{1,K} \leqslant \varepsilon) = \min(2\varepsilon, 1).$$

- $\Rightarrow 2M_{1,K}$ is a precise p-merging function

Task. Find $b_{r,K} > 0$ such that (the M-family)

$$F_{r,K} = b_{r,K} M_{r,K} \text{ is precise}$$

- $M_{r,K}$ increases in $r \Longrightarrow b_{r,K}$ should decrease in r.

Translation to a risk aggregation problem

For $\alpha \in (0,1]$ define

$$q_\alpha(X) = \inf\{x \in \mathbb{R} : \mathbb{P}(X \leqslant x) \geqslant \alpha\};$$
$$\underline{q}_\alpha(F) = \inf\left\{q_\alpha(F(\mathbf{P})) : \mathbf{P} \in \mathcal{U}^K\right\}$$

Lemma 1

For $a > 0$, $r \in [-\infty, \infty]$, and $F = aM_{r,K}$, equivalent are:

(i) *F is a p-merging function, i.e., $\overline{\mathbb{P}}(F \leqslant \varepsilon) \leqslant \varepsilon$ for all $\varepsilon \in (0,1)$;*

(ii) *$\underline{q}_\varepsilon(F) \geqslant \varepsilon$ for all $\varepsilon \in (0,1)$;*

(iii) *$\overline{\mathbb{P}}(F \leqslant \varepsilon) \leqslant \varepsilon$ for some $\varepsilon \in (0,1)$;*

(iv) *$\underline{q}_\varepsilon(F) \geqslant \varepsilon$ for some $\varepsilon \in (0,1)$.*

The same conclusion holds if all $\leqslant$ and $\geqslant$ are replaced by $=$.

Translation to a risk aggregation problem

It boils down to calculate $\underline{q}_\varepsilon(M_{r,K})$, or equivalently:

(i) for $r > 0$, aggregation of Beta risks

$$(\underline{q}_\varepsilon(M_{r,K}))^r = \inf_{U_1,\dots,U_K \in \mathcal{U}} \left\{ q_\varepsilon \left(\frac{1}{K} (U_1^r + \cdots + U_K^r) \right) \right\}$$

(ii) for $r = 0$, aggregation of exponential risks

$$\log(\underline{q}_\varepsilon(M_{r,K})) = \inf_{U_1,\dots,U_K \in \mathcal{U}} \left\{ q_\varepsilon \left(\frac{1}{K} (\log U_1 + \cdots + \log U_K) \right) \right\}$$

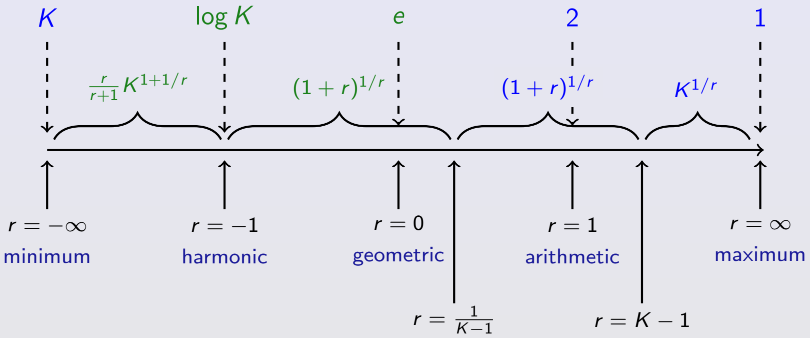
(iii) for $r < 0$, aggregation of Pareto risks

$$(\underline{q}_\varepsilon(M_{r,K}))^r = \sup_{U_1,\dots,U_K \in \mathcal{U}} \left\{ q_{1-\varepsilon} \left(\frac{1}{K} (U_1^r + \cdots + U_K^r) \right) \right\}$$

- A general tool: convolution bound Blanchet/Lam/Liu/W.'24

Summary

Constant multiplier in front of $M_{r,K}$



blue: precise; green: approximate

1. Combining p-values, optimal transport, and risk management
2. P-merging functions
3. Admissibility
4. Combining exchangeable p-values
5. External randomization
6. Simulation and summary

Admissible p-merging functions

For p-merging functions F and G:

- F dominates G if $F \leqslant G$
- F is admissible if it is not dominated by any other one
- F is symmetric if $F(\mathbf{p})$ is invariant under permutation of $\mathbf{p}$
- F is homogeneous if $F(\lambda \mathbf{p}) = \lambda F(\mathbf{p})$ for all $\lambda \in (0, 1]$ and $\mathbf{p}$ with $F(\mathbf{p}) \leqslant 1$

Properties

- Admissible $\Longrightarrow$ precise, lower semicontinuous, grounded
- Any p-merging function is dominated by an admissible one

Rejection regions of admissible p-merging functions

- The rejection region of F at level $\varepsilon \in (0,1)$:

$$R_\varepsilon(F) := \left\{\mathbf{p} \in [0,\infty)^K : F(\mathbf{p}) \leqslant \varepsilon\right\}$$

- An accending chain $\{R_\varepsilon \subseteq [0,\infty)^K : \varepsilon \in (0,1)\}$ of Borel lower sets induces a function $F : [0,\infty)^K \to [0,1]$ via

$$F(\mathbf{p}) = \inf\{\varepsilon \in (0,1) : \mathbf{p} \in R_\varepsilon\} \quad \text{with } \inf \varnothing = 1$$

- F is p-merging $\iff \mathbb{P}(\mathbf{P} \in R_\varepsilon) \leqslant \varepsilon$ for all $\varepsilon \in (0,1)$, $\mathbf{P} \in \mathcal{U}^K$
- F is homogeneous $\implies R_\varepsilon(F) = \varepsilon A$ for some $A \subseteq [0,\infty)^K$.

Rejection regions of admissible p-merging functions

Admissibility $\iff$ rejection region cannot be enlarged

- Precision of p-merging $\iff$ classic OT

$$\text{Compute} \sup_{\mathbf{P}\in\mathcal{U}^K} \mathbb{E}[\mathbb{1}_A(\mathbf{P})]$$

- Admissibility (or optimality) $\iff$ "reverse OT"

$$\text{Given} \sup_{\mathbf{P}\in\mathcal{U}^K} \mathbb{E}[\mathbb{1}_A(\mathbf{P})] \leqslant \varepsilon, \text{ find a maximal } A \subseteq [0,1]^K$$

- Such A needs to be nested
- Techniques in OT can be very helpful

One slide on e-values

Using p-to-e calibrators Vovk/W.'21

- A calibrator is a decreasing function $f : [0, \infty) \to [0, \infty]$ satisfying $f = 0$ on $(1, \infty)$ and $\int_0^1 f(x)\mathrm{d}x \leqslant 1$
- A calibrator f is admissible if it is upper semicontinuous, $f(0) = \infty$, and $\int_0^1 f(x)\mathrm{d}x = 1$

E-variables

An e-variable E for $\mathcal{H}$ is a $[0, \infty]$-valued random variable satisfying $\mathbb{E}^{\mathbb{P}}[E] \leqslant 1$ for $\mathbb{P} \in \mathcal{H}$.

- $f(P/\varepsilon)/\varepsilon$ is an e-variable for $P \in \mathcal{U}$ and $\varepsilon \in (0, 1)$
- $\mathbb{P}(E \geqslant 1/\varepsilon) \leqslant \varepsilon$ for an e-variable E (Markov's inequality)

Representation theorems

Let Δ_K be the standard K-simplex and write $\mathbf{p} = (p_1, \dots, p_K)$.

Theorem 1

For any admissible homogeneous p-merging function F, there exist $(\lambda_1, \dots, \lambda_K) \in \Delta_K$ and admissible calibrators $f_1, \dots, f_K$ such that

$$R_\varepsilon(F) = \left\{ \mathbf{p} \in [0,\infty)^K : \sum_{k=1}^K \lambda_k \frac{1}{\varepsilon} f_k\left(\frac{p_k}{\varepsilon}\right) \geqslant \frac{1}{\varepsilon} \right\}, \ \varepsilon \in (0,1). \quad \text{(H)}$$

Conversely, for any $(\lambda_1, \dots, \lambda_K) \in \Delta_K$ and calibrators $f_1, \dots, f_K$, (H) induces a homogeneous p-merging function.

Connection to e-tests

(H) is equivalent to

$$F(\mathbf{p}) = \inf\left\{\varepsilon \in (0,1) : \sum_{k=1}^{K} \lambda_k f_k\left(\frac{p_k}{\varepsilon}\right) \geqslant 1\right\}$$

- Admissible (homogeneous) ways of merging p-values must be through merging e-values and testing with Markov's inequality
- Proof is based on optimal transport duality

$$\min\left\{\sum_{k=1}^{K}\int_0^1 g_k(x)\mathrm{d}x : \bigoplus_{k=1}^{K} g_k \geqslant \mathbb{1}_{R_\varepsilon(F)}\right\} = \max_{\mathbf{P}\in\mathcal{U}} \mathbb{P}(\mathbf{P} \in R_\varepsilon(F)) = \varepsilon$$

Representation theorems

Theorem 2

For any F that is admissible within the family of homogeneous symmetric p-merging functions, there exists an admissible calibrator f such that

$$R_\varepsilon(F) = \left\{ \mathbf{p} \in [0,\infty)^K : \frac{1}{K}\sum_{k=1}^{K} f\left(\frac{p_k}{\varepsilon}\right) \geqslant 1 \right\}, \ \varepsilon \in (0,1). \quad \text{(SH)}$$

Conversely, for any calibrator f, (SH) induces a homogeneous symmetric p-merging function.

- We say f induces F if (SH) holds
- Converse: not true; calibrator: not unique

Connection to joint mixability

A necessary and sufficient condition for a calibrator f to induce a precise p-merging function via (SH) is

$$\mathbb{P}\left(\frac{1}{K}\sum_{k=1}^{K} f(P_k) = 1\right) = 1 \quad \text{for some } P_1, \dots, P_K \in \mathcal{U}. \qquad \text{(JM)}$$

- $\Longrightarrow$ Joint mixability — Wang/W.'11'16
- Difficult to check for a given f in general
- For a convex f, (JM) holds if and only if $f \leqslant K$ on $(0, 1]$
- Weaker than admissibility

Sufficient conditions for admissibility

Theorem 3

Suppose that an admissible calibrator f is strictly convex or strictly concave on $(0,1]$, $f(0+) \in (K/(K-1), K]$, and $f(1) = 0$. The p-merging function induced by f is admissible.

- Proof based on joint mixability (JM)
- Open question: can strict convexity be reduced to convexity?
- Conditions of this type are not necessary

Hommel's function

Define the Hommel* calibrator f by

$$f : x \mapsto \frac{K \mathbb{1}_{\{\ell_K x \leqslant 1\}}}{\lceil K \ell_K x \rceil}.$$

Theorem 4

The p-merging function $H_K \wedge 1$ is dominated (strictly if $K \geqslant 4$) by the p-merging function H_K^* induced by f via (SH). Moreover, H_K^* is always admissible among symmetric p-merging functions, and it is admissible if K is not a prime number.

- Primality appears in the proof due to factoring the set $[K]$
- H_K^* is not admissible for $K = 2, 3$ (we guess also 5)

Hommel's function and the O-family

- The Bonferroni method is admissible
- Members $G_{k,K}$ of the O-family are admissible after truncation at 1 except for $k = K$
- Members $F_{r,K}$ of the M-family are not admissible except for $r = -\infty$
- $F_{r,K}$ can be strictly improved to $F^*_{r,K}$
- $F^*_{-1,K}$ is similar to H^*_K

Combining exchangeable p-values

Definition 1 (Exchangeability)

The random vector $(P_1, \dots, P_K)$ is exchangeable if

$$(P_1, \dots, P_K) \overset{\mathrm{d}}{=} (P_{\sigma(1)}, \dots, P_{\sigma(K)})$$

for all K-permutations σ.

Examples

- Split conformal prediction
- Median-of-means methods
- Overlapping sample

Exchangeable/uniformly-randomized Markov inequalities

Proposition 1 (Ramdas/Manole'24 STS)

Let $X_1, X_2, \ldots$ form an exchangeable sequence of non-negative random variables. Then, for any $a > 0$,

$$\mathbb{P}\left(\exists k \geqslant 1 : \frac{1}{k}\sum_{i=1}^{k} X_i \geqslant \frac{1}{a}\right) \leqslant a\mathbb{E}[X_1].$$

Proof.

- Let $S_k = \sum_{i=1}^{k} X_i$
- $\mathbb{E}[S_k|S_{k+1}] = kS_{k+1}/(k+1) \implies (S_k/k)_{k\in[K]}$ is a backward martingale
- Ville's inequality gives the desired statement

Combining exchangeable p-values

Theorem 5

Let f be a calibrator, and $\mathbf{P} = (P_1, \dots, P_K) \in \mathcal{U}^K$ be exchangeable. For each $\alpha \in (0,1)$, we have

$$\mathbb{P}\left(\exists k \leqslant K : \frac{1}{k}\sum_{i=1}^{k} f\left(\frac{P_i}{\alpha}\right) \geqslant 1\right) \leqslant \alpha.$$

Combining exchangeable p-values

- For a calibrator f and $\alpha \in (0,1)$, let

$$R_\alpha = \left\{ \mathbf{p} \in [0,1]^K : \frac{1}{k}\sum_{i=1}^{k} f\left(\frac{p_i}{\alpha}\right) \geqslant 1 \text{ holds for some } k \leqslant K \right\}$$

and

$$\begin{aligned} F(\mathbf{p}) &= \inf\{\alpha \in (0,1) : \mathbf{p} \in R_\alpha\} \\ &= \inf \left\{ \alpha \in (0,1) : \bigvee_{k=1}^{K} \left(\frac{1}{k}\sum_{i=1}^{k} f\left(\frac{p_i}{\alpha}\right)\right) \geqslant 1 \right\}, \qquad \text{(Ex)} \end{aligned}$$

with the convention $\inf \varnothing = 1$

- (Ex) is always smaller or equal than the deterministic version

Combining exchangeable p-values

Definition 2

An ex-p-merging function is an increasing Borel function $F : [0,1]^K \to [0,1]$ such that $\mathbb{P}(F(\mathbf{P}) \leqslant \alpha) \leqslant \alpha$ for all $\alpha \in (0,1)$ and $\mathbf{P} \in \mathcal{U}^K$ that is exchangeable.

Theorem 6

If f is a calibrator and $\mathbf{P} \in \mathcal{U}^K$ is exchangeable, then F in (Ex) is a homogeneous ex-p-merging function, and dominating (H).

Combining exchangeable p-values

Proposition 2

A symmetric ex-p-merging function is necessarily a p-merging function.

- Hence, for an ex-p-merging function to strictly dominate an admissible p-merging function, it cannot be symmetric

Proposition 3

Suppose that f is a convex admissible calibrator with $f(0+) \leqslant K$ and $f(1) = 0$, and F is in (Ex). For $\alpha \in (0,1)$, there exists an exchangeable $\mathbf{P} \in \mathcal{U}^K$ such that $\mathbb{P}(F(\mathbf{P}) \leqslant \alpha) = \alpha$.

- Again relying on (JM)

Combining exchangeable p-values

Theorem 7

Let $P_1, P_2, \ldots \in \mathcal{U}^\infty$ be an infinite exchangeable sequence and let F be the function defined in (Ex). Then,

$$\mathbb{P}(\exists k \geqslant 1 : F(\mathbf{P}_k) \leqslant \alpha) \leqslant \alpha,$$

where $\mathbf{P}_k = (P_1, \ldots, P_k)$.

External randomization

Proposition 4

Let X be a non-negative random variable independent of $U \in \mathcal{U}$. Then, for any $a > 0$,

$$\mathbb{P}(X \geqslant U/a) \leqslant a\mathbb{E}[X].$$

External randomization

$$\mathcal{U}^K \otimes \mathcal{U} = \{(\mathbf{P}, U) : \mathbf{P} \in \mathcal{U}^K \text{ and } U \in \mathcal{U} \text{ are independent}\}$$

Theorem 8

Let $f_1, \ldots, f_K$ be calibrators and $(P_1, \ldots, P_K, U) \in \mathcal{U}^K \otimes \mathcal{U}$. For each $\alpha \in (0,1)$ and $(\lambda_1, \ldots, \lambda_K) \in \Delta_K$, we have

$$\mathbb{P}\left(\sum_{k=1}^K \lambda_k f_k\left(\frac{P_k}{\alpha}\right) \geqslant U\right) \leqslant \alpha.$$

If $f_1, \ldots, f_K$ are admissible and $\mathbb{P}(\sum_{k=1}^K \lambda_k f_k(P_k/\alpha) \leqslant 1) = 1$, then

$$\mathbb{P}\left(\sum_{k=1}^K \lambda_k f_k\left(\frac{P_k}{\beta}\right) \geqslant U\right) = \beta \quad \text{for all } \beta \in (0, \alpha].$$

External randomization

- Let $f_1, \ldots, f_K$ be calibrators and $(\lambda_1, \ldots, \lambda_K) \in \Delta_K$
- For $\alpha \in (0,1)$, let

$$R_\alpha = \left\{ (\mathbf{p}, u) \in [0,1]^{K+1} : \sum_{k=1}^{K} \lambda_k f_k \left(\frac{p_k}{\alpha} \right) \geqslant u \right\}$$

where we set $f_k(p_k/u) = 0$ if $u = 0$, and

$$\begin{aligned} F(\mathbf{p}, u) &= \inf\{\alpha \in (0,1) : (\mathbf{p}, u) \in R_\alpha\} \\ &= \inf \left\{ \alpha \in (0,1) : \sum_{k=1}^{K} \lambda_k f_k \left(\frac{p_k}{\alpha} \right) \geqslant u \right\}, \qquad \text{(U)} \end{aligned}$$

with the convention $\inf \varnothing = 1$ and $0 \times \infty = \infty$

External randomization

Definition 3

A randomized p-merging function is an increasing Borel function $F : [0,1]^{K+1} \to [0,1]$ such that $\mathbb{P}(F(\mathbf{P}, U) \leqslant \alpha) \leqslant \alpha$ for all $\alpha \in (0,1)$ and $(\mathbf{P}, U) \in \mathcal{U}^K \otimes \mathcal{U}$.

Theorem 9

If $f_1, \dots, f_K$ are calibrators and $(\lambda_1, \dots, \lambda_K) \in \Delta_K$, then F in (U) is a homogenous randomized p-merging function. Moreover, F is lower semicontinuous.

External randomization

Remarks:

- If P_1 is independent of $(P_2, \dots, P_K)$, then we can use $U = P_1$ and merge $(P_2, \dots, P_K)$
- Merging asymptotic p-values: if $(\mathbf{P}_n)_{n\in\mathbb{N}} \xrightarrow{\text{law}} \mathbf{P}$, then

$$\limsup_{n\to\infty} \mathbb{P}(\mathbf{P}_n \in R_\alpha) \leqslant \mathbb{P}(\mathbf{P} \in R_\alpha) \leqslant \alpha$$

because the rejection sets in (H), (Ex), and (U) are closed

Summary

Rule	Arbitrary dependence	Exchangeable	Randomized
Rüger's	$\frac{K}{k} p_{(k)}$	$\frac{K}{k} \bigwedge_{m=1}^{K} p^m_{(\lambda_m)}$	$\frac{K}{k} p_{(\lceil uk \rceil)}$
Arithmetic	$2A(\mathbf{p})$	$2 \bigwedge_{m=1}^{K} A(\mathbf{p}_m)$	$\frac{2}{2-u} A(\mathbf{p})$
Geometric	$eG(\mathbf{p})$	$e \bigwedge_{m=1}^{K} G(\mathbf{p}_m)$	$e^u G(\mathbf{p})$
Harmonic	$(T_K + 1)H(\mathbf{p})$	$(T_K + 1) \bigwedge_{m=1}^{K} H(\mathbf{p}_m)$	$(T_K u + 1)H(\mathbf{p})$

- $\mathbf{p} = (p_1, \dots, p_K)$
- $\mathbf{p}_m$: the vector containing the first m values of $\mathbf{p}$
- $p_{(k)}$ is the k-th smallest value of $\mathbf{p}$
- $p^m_{(\lambda_m)}$ is the $\lambda_m = \lceil m \frac{k}{K} \rceil$ ordered value of $\mathbf{p}_m$
- A, G and H: the arithmetic, the geometric, and the harmonic mean
- $T_K = \log K + \log\log K + 1,\ K \geqslant 2.$

Simulation results

- One-side z-test of $\mu = 0$ against the alternative $\mu > 0$
- $Z, Z_1, \ldots, Z_K \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$
- P-values are given by

$$X_k = \rho Z + \sqrt{1-\rho^2} Z_k - \mu, \quad P_k = \Phi(X_k),$$

  where Φ is standard normal cdf
- $\mu \geqslant 0$ and $\rho \in [0,1]$ are constants
- $P_1, \ldots, P_K$ are exchangeable
- $\alpha = 0.05$
- $B = 10,000$ replications

Simulation results

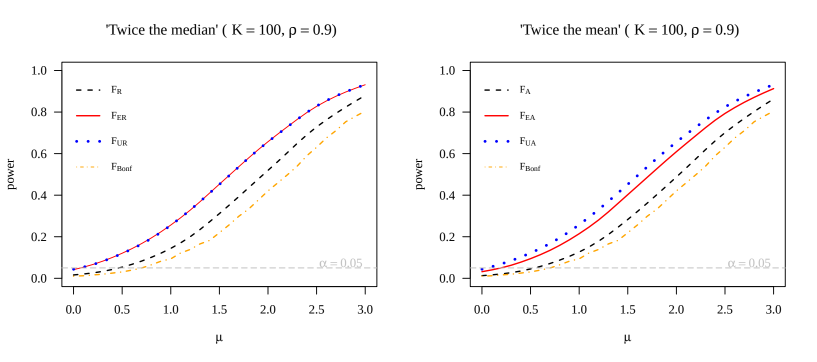


Figure: Empirical power of different combination rules, $\rho = 0.9$

Simulation results

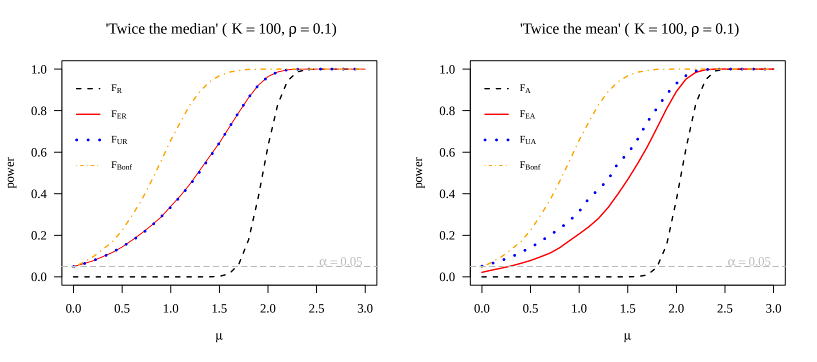


Figure: Empirical power of different combination rules, $\rho = 0.1$

Simulation results

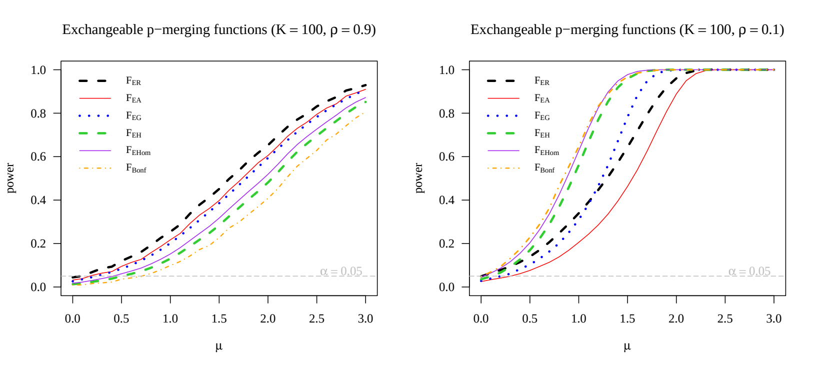


Figure: Combination of p-values using different exchangeable p-merging functions. The performance of the different exchangeable p-merging functions is almost reversed in the two situations.

Simulation results

- $X_{ij} \overset{\text{iid}}{\sim} \mathcal{N}(\mu, 1)$ for $i = 0, 1, \dots, K$ and $j = 1, \dots, n_i$
- X_{0j}, $j = 1, \dots, n_0$: common sample
- Sample average

$$\bar{X}_i = \frac{1}{\sqrt{n_i}} \sum_{j=1}^{n_i} X_{ij} \sim \mathcal{N}(\sqrt{n_i}\mu, 1), \quad i = 0, 1, \dots, K$$

- Test statistics and p-values, exchangeable under the null

$$T_i = \frac{\bar{X}_i + \bar{X}_0}{\sqrt{2}} \text{ and } P_i = 2 \min \{\Phi(T_i), \Phi(-T_i)\}, \quad i = 1, \dots, K$$

- Order $P_1, \dots, P_K$ by $n_1, \dots, n_K$

Simulation results

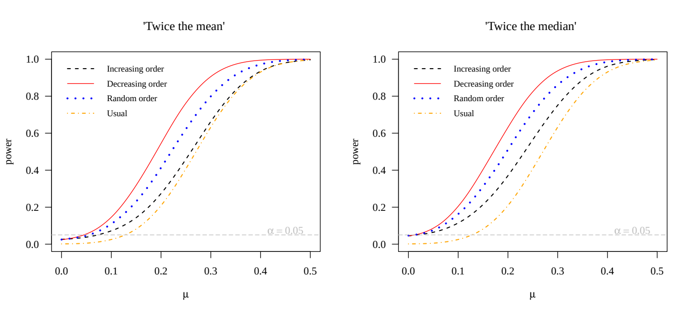


Figure: Combination of p-values using different ex-p-merging functions and different ordering based on the number of observations, where $K = 10$ and $n_i = 10 \times i$. The ex-p-merging rules are more powerful if p-values are ordered in decreasing order with respect to the number of observations.

Summary

Take home these

$$e \bigwedge_{m=1}^{K} G(\mathbf{p}_m); \qquad (\log K + \log\log K + 2) \bigwedge_{m=1}^{K} H(\mathbf{p}_m)$$

$$\frac{K}{k} p_{(\lceil uk \rceil)}; \qquad e^u G(\mathbf{p})$$

- Merging p-values under unknown dependence $\Longrightarrow$ merging e-values
- Exchangeability allows for processing p-values sequentially
- Randomization improves power
- Dependence problems can be solved with OT theory
- Many open questions!
- Write me if you have a problem whose challenge lies in dependence

Thank you

Thank you for your kind attention

Vladimir Vovk

Bin Wang

Matteo Gasparin

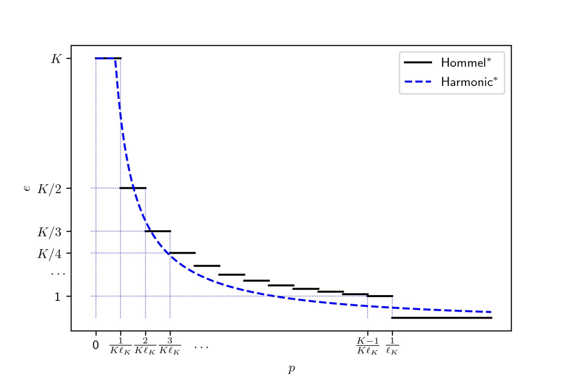
Aaditya Ramdas

Paul Embrechts

Hommel's function and the O-family

Define the Hommel* calibrator f by

$$f : x \mapsto \frac{K \mathbb{1}_{\{\ell_K x \leqslant 1\}}}{\lceil K \ell_K x \rceil}.$$



Hommel's function and the O-family

Theorem 10

The p-merging function $H_K \wedge 1$ is dominated (strictly if $K \geqslant 4$) by the p-merging function H_K^* induced by the Hommel calibrator f,*

$$R_\varepsilon(H_K^*) = \varepsilon \left\{ \mathbf{p} \in [0,\infty)^K : \frac{1}{K} \sum_{k=1}^{K} f(p_k) \geqslant 1 \right\}, \quad \varepsilon \in (0,1).$$

Moreover, H_K^* is always admissible among symmetric p-merging functions, and it is admissible if K is not a prime number.

Hommel's function and the O-family

Proof sketch.

- Recall: $H_K(\mathbf{p}) = \ell_K \bigwedge_{k=1}^K \frac{K}{k} p_{(k)}$ where $\ell_K = \sum_{i=1}^K \frac{1}{k}$.
- Induced by the calibrator $f \implies H_K^*$ is a p-merging function.
- Verify $H_K \geqslant H_K^*$: $H_K(\mathbf{p}) \leqslant \varepsilon \implies$ there exists m such that $K\ell_K p_{(m)} \leqslant \varepsilon$ $\implies \#\{k : K\ell_K p_k/m \leqslant \varepsilon\} \geqslant m \implies$

$$\sum_{k=1}^K \frac{\mathbb{1}_{\{\ell_K p_k \leqslant \varepsilon\}}}{\lceil K\ell_K p_i/\varepsilon \rceil} \geqslant \sum_{k=1}^K \frac{1}{m} \mathbb{1}_{\{K\ell_K p_k/\varepsilon \leqslant m\}} = \frac{1}{m} \#\{k : K\ell_K p_k/m \leqslant \varepsilon\} \geqslant 1$$

$\implies R_\varepsilon(H_K) \subseteq R_\varepsilon(H_K^*) \implies H_K \geqslant H_K^*$.

- Check $H_K = H_K^*$ if and only if $K \leqslant 3$.

Hommel's function and the O-family

Proof sketch (continued).

- Suppose H_K^* is not admissible among symmetric p-merging functions.
- There exists a calibrator g satisfying

$$\left\{\mathbf{p}\in[0,\infty)^K : \frac{1}{K}\sum_{k=1}^K f(p_k)\geqslant 1\right\} \subsetneq \left\{\mathbf{p}\in[0,\infty)^K : \frac{1}{K}\sum_{k=1}^K g(p_k)\geqslant 1\right\}.$$

- Denote by $\tau := 1/(K\ell_K)$. For $x\in(0,K\tau]$, set $p_1=\cdots=p_m=x$ and $p_{m+1}=\cdots=p_K>1$, where $m:=\lceil \tau x\rceil$.
- $f(x)=K/m \implies \sum_{k=1}^K f(p_k)=K \implies K\leqslant\sum_{k=1}^K g(p_k)=mg(x)\implies g(x)\geqslant K/m=f(x)$.
- $\int_0^{K\tau} g(x)\mathrm{d}x \geqslant \int_0^{K\tau} f(x)\mathrm{d}x = 1 \implies g=f$ almost everywhere on $[0,1]$.
- f is left-continuous $\implies g\leqslant f$, a contradiction.
- The admissibility statement for non-prime K is much more complicated.

Hommel's function and the O-family

- $S_K \leqslant H_K^* \leqslant H_K \implies 1/\ell_K \leqslant H_K^*/H_K \leqslant 1$
- H_K^* may not be admissible for a prime K

Example 1

In case $K = 2$, $H_2^* = H_2 : (p_1, p_2) \mapsto (3p_{(1)}) \wedge (\frac{3}{2}p_{(2)})$ is strictly dominated by $F : (p_1, p_2) \mapsto (3p_1) \wedge (\frac{3}{2}p_2)$, which is a (non-symmetric) p-merging function because

$$\mathbb{P}(F(P_1, P_2) \leqslant \alpha) \leqslant \mathbb{P}\left(P_1 \leqslant \frac{1}{3}\alpha\right) + \mathbb{P}\left(P_2 \leqslant \frac{2}{3}\alpha\right) \leqslant \alpha.$$

Hommel's function and the O-family

Theorem 11

The p-merging function $\mathbf{p} \mapsto G_{k,K}(\mathbf{p}) \wedge \mathbb{1}_{\{\min(\mathbf{p})>0\}}$ *is admissible for* $k = 1, \ldots, K-1$*, and it is admissible among symmetric p-merging functions for* $k = K$*.*

The M-family

- $F_{r,K} = (b_{r,K} M_{r,K}) \wedge 1$
- For $r \neq \{-1, 0\}$ and $r < 1/(K-1)$, denote by c_r the unique number $c \in (0, 1/K)$ solving the equation

$$\frac{(K-1)(1-(K-1)c)^r + c^r}{K} = \frac{(1-(K-1)c)^{r+1} - c^{r+1}}{(r+1)(1-Kc)}$$

- c_{-1} and c_0 are limits of c_r
- Set $c_r := 0$ for $r \geqslant 1/(K-1)$
- Write $d_r := 1 - (K-1)c_r$

The M-family

Proposition 5

For $K \geqslant 3$ and $r \in (-\infty, \frac{1}{K-1})$,

$$b_{r,K} = 1/M_{r,K}(c_r, d_r, \ldots, d_r).$$

- If $r < s$ and $rs > 0$, then

$$K^{1/s-1/r} b_{r,K} \leqslant b_{s,K} \leqslant b_{r,K}$$

The M-family

For $r < 0$:

- Rejection region

$$R_\varepsilon(F_{r,K}) = \varepsilon \left\{ \mathbf{p} \in [0,\infty)^K : \frac{\sum_{k=1}^K p_k^r}{c_r^r + (K-1)d_r^r} \geqslant 1 \right\}$$

$$= \varepsilon \left\{ \mathbf{p} \in [0,\infty)^K : \sum_{k=1}^K \frac{p_k^r - d_r^r}{c_r^r - d_r^r} \geqslant 1 \right\}.$$

- Define a calibrator

$$f_r : x \mapsto K \left(\frac{x^r - d_r^r}{c_r^r - d_r^r} \wedge 1 \right)_+ .$$

- f_r is strictly convex on $[c_r, d_r]$.

The M-family

Let $F^*_{r,K}$ be the p-merging function induced by f_r, i.e.,

$$R_\varepsilon(F^*_{r,K}) = \varepsilon \left\{ \mathbf{p} \in [0,\infty)^K : \sum_{k=1}^K \left(\frac{p_k^r - d_r^r}{c_r^r - d_r^r} \right)_+ \geqslant 1 \right\}, \ \varepsilon \in (0,1).$$

- $R_\varepsilon(F_{r,K}) \subset R_\varepsilon(F^*_{r,K})$
- $F^*_{r,K}$ is admissible

The M-family

Theorem 12

For $K \geqslant 3$ and $r \in (-\infty, K-1)$, $F_{r,K}$ is strictly dominated by the p-merging function $F^*_{r,K}$ defined via, for $\mathbf{p} \in (0,\infty)^K$ and $\varepsilon \in (0,1)$,

$$F^*_{r,K}(\mathbf{p}) \leqslant \varepsilon \iff F_{r,K}(\mathbf{p} \wedge (\varepsilon d_r \mathbf{1})) \leqslant \varepsilon \quad \text{or} \quad \min(\mathbf{p}) = 0.$$

Moreover, $F^*_{r,K}$ is admissible unless $r = 1$.

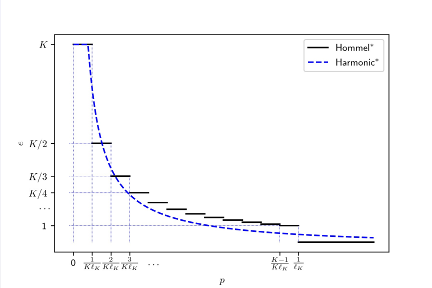
The M-family

Recall

$$f_{-1} : x \mapsto K \left(\frac{x^{-1} - d_{-1}^{-1}}{c_{-1}^{-1} - d_{-1}^{-1}} \wedge 1 \right)_+ ;$$

the Hommel* calibrator

$$f : x \mapsto \frac{K \mathbb{1}_{\{\ell_K x \leqslant 1\}}}{\lceil K \ell_K x \rceil}.$$



- When taking values in $(0, K)$:

$$f_{-1}(x) = a/x - b \quad \text{vs} \quad f(y) = a'/\lceil b'y \rceil$$

The M-family

Proposition 6

For $K \geqslant 3$ *and* $\mathbf{p} \in [0,\infty)^K$, *we have, if* $r \in (-\infty, 1/(K-1))$,

$$F^*_{r,K}(\mathbf{p}) = \left(\bigwedge_{m=1}^{K} \frac{M_{r,m}(p_{(1)},\dots,p_{(m)})}{M_{r,m}(c_r, d_r, \dots, d_r)} \right) \wedge 1,$$

and, if $r \in [1/(K-1), K-1)$, *with the convention* $\cdot/0 = \infty$,

$$F^*_{r,K}(\mathbf{p}) = \left(\bigwedge_{m=1}^{K} \frac{M_{r,m}(p_{(1)},\dots,p_{(m)})}{\left(1 - \frac{rK}{(r+1)m}\right)_+} \right) \wedge \mathbb{1}_{\{p_{(1)}>0\}}.$$

The M-family

Proposition 7

For $r < s$, $K \geqslant 2$ and $a, b > 0$, the following statements hold.

(i) $aM_{r,K} \leqslant bM_{s,K}$ *if and only if* $a \leqslant b$.

(ii) $bM_{s,K} \leqslant aM_{r,K}$ *if and only if* $rs > 0$ *and* $aK^{-1/r} \geqslant bK^{-1/s}$.

Proposition 8

Suppose $r \neq s$. *If* $K = 2$, $F_{r,K} \geqslant F_{s,K}$ *if and only if* $1 \leqslant r < s$ *or* $s < r \leqslant 1$. *If* $K \geqslant 3$, $F_{r,K} \geqslant F_{s,K}$ *if and only if* $K - 1 \leqslant r < s$.

Magnitude of improvement

Proposition 9

For $K \geqslant 3$, we have

$$\inf_{\mathbf{p}>\mathbf{0}} \frac{F^*_{1,K}(\mathbf{p})}{F_{1,K}(\mathbf{p})} = \inf_{\mathbf{p}>\mathbf{0}} \frac{F^*_{0,K}(\mathbf{p})}{F_{0,K}(\mathbf{p})} = 0, \quad \inf_{\mathbf{p}>\mathbf{0}} \frac{F^*_{-1,K}(\mathbf{p})}{F_{-1,K}(\mathbf{p})} = 1-(K-1)c_{-1},$$

and

$$\min_{\mathbf{p}>\mathbf{0}} \frac{H^*_K(\mathbf{p})}{H_K(\mathbf{p})} = \min\left\{ t>0 : \sum_{k=1}^{K} \frac{\mathbb{1}_{\{t \geqslant k/K\}}}{\lceil k/t \rceil} \geqslant 1 \right\} =: \gamma_K.$$

Moreover, $c_{-1} \sim 1/(K \log K)$ and $\gamma_K \sim 1/\log K$ as $K \to \infty$.

Magnitude of improvement

- $F^*_{-1,K}$ improves $F_{-1,K}$ only by a factor $1 - 1/\log K \sim 1$
- H^*_K can improve H_K by a significant factor of $1/\log K$
- $H^*_K(\mathbf{p})/H_K(\mathbf{p}) = \gamma_K$ is attained by $\mathbf{p} = (\alpha, 2\alpha, \cdots, K\alpha)$ for $\alpha \in (0, 1/K\ell_K]$.
- Since $H_K = \ell_K S_K$ and

$$\gamma_K \sim 1/\log K \sim 1/\ell_K,$$

H^*_K performs similarly to the Simes function S_K for some values of $\mathbf{p}$ above

Simes function

Proof sketch.

- Take any symmetric p-merging function F and $\mathbf{p} = (p_1, \dots, p_K)$
- Let $\alpha := S_K(\mathbf{p})/K \implies p_{(k)} \geqslant k\alpha$ for each k
- Symmetry and monotonicity of $F \implies$

$$F(\mathbf{p}) = F(p_{(1)}, \dots, p_{(K)}) \geqslant F(\alpha, 2\alpha, \dots, K\alpha) =: \beta$$

- Let Π be the set of all permutations of $(\alpha, 2\alpha, \dots, K\alpha)$, and $\mu = \mathrm{U}(\Pi)$
- Take $(P_1, \dots, P_K) \sim K\alpha\mu + (1 - K\alpha)\delta_{(1,\dots,1)}$
- For each k, $P_k \sim \sum_{k=1}^K \alpha\delta_{k\alpha} + (1 - K\alpha)\delta_1 \implies P_k$ is a p-variable
- F is a p-merging function $\implies$

$$\beta \geqslant \mathbb{P}(F(P_1, \dots, P_K) \leqslant \beta) \geqslant \mathbb{P}((P_1, \dots, P_K) \in \Pi) = K\alpha$$

- $F(\mathbf{p}) \geqslant K\alpha = S_K(\mathbf{p}) \implies S_K$ dominates all symmetric p-merging functions
- $S_K = \bigwedge_{k=1}^K G_{k,K}$

Representation theorems

Proof sketch.

- For decreasing functions $g_1, \dots, g_K$ on $[0, \infty)$, denote by

$$\left(\bigoplus_{k=1}^{K} g_k\right)(x_1, \dots, x_K) := \sum_{k=1}^{K} g_k(x_k)$$

- Classic duality ($R_\varepsilon(F)$ is closed and F is precise)

$$\min\left\{\sum_{k=1}^{K}\int_0^1 g_k(x)\mathrm{d}x : \bigoplus_{k=1}^{K} g_k \geqslant \mathbb{1}_{R_\varepsilon(F)}\right\} = \max_{\mathbf{P}\in\mathcal{U}} \mathbb{P}(\mathbf{P} \in R_\varepsilon(F)) = \varepsilon$$

- Take $(g_1^\varepsilon, \dots, g_K^\varepsilon)$ such that $\bigoplus_{k=1}^K g_k^\varepsilon \geqslant \mathbb{1}_{R_\varepsilon(F)}$ and $\sum_{k=1}^K \int_0^1 g_k^\varepsilon(x)\mathrm{d}x = \varepsilon$
- Choose g_k^ε to be non-negative and left-continuous
- Monotonicity

$$\max_{\mathbf{P}\in\mathcal{U}^K} \mathbb{P}(\mathbf{P} \in R_\varepsilon(F)) = \varepsilon \implies \max_{\mathbf{P}\in\mathcal{U}^K} \mathbb{P}(\varepsilon\mathbf{P} \in R_\varepsilon(F)) = 1$$

Representation theorems

Proof sketch (continued).

- Using duality again

$$\min\left\{\sum_{k=1}^K \frac{1}{\varepsilon}\int_0^\varepsilon g_k(x)\mathrm{d}x : \bigoplus_{k=1}^K g_k \geqslant \mathbb{1}_{R_\varepsilon(F)}\right\} = 1$$

$\Longrightarrow \sum_{k=1}^K \int_0^\varepsilon g_k^\varepsilon(x)\mathrm{d}x \geqslant \varepsilon \Longrightarrow g_k^\varepsilon(x) = 0$ for $x > \varepsilon$

- Define the set $A_\varepsilon := \{\mathbf{p} \in [0,\infty)^K : \sum_{k=1}^K g_k^\varepsilon(p_k) \geqslant 1\}$
- $\bigoplus_{k=1}^K g_k^\varepsilon \geqslant \mathbb{1}_{R_\varepsilon(F)} \Longrightarrow R_\varepsilon(F) \subseteq A_\varepsilon$
- By Markov's inequality,

$$\sup_{\mathbf{P}\in\mathcal{U}^K} \mathbb{P}\left(\bigoplus_{k=1}^K g_k^\varepsilon(\mathbf{P}) \geqslant 1\right) \leqslant \sup_{P\in\mathcal{U}} \sum_{k=1}^K \mathbb{E}[g_k^\varepsilon(P)] = \varepsilon$$

- Define a function F' with rejection region $R_\delta(F') = \delta\varepsilon^{-1}A_\varepsilon$ for $\delta \in (0,1)$
- F' is a valid homogeneous p-merging function and $F' \leqslant F$

Representation theorems

Proof sketch (continued).

- Admissibility of $F \implies F = F'$, thus

$$R_\varepsilon(F) = A_\varepsilon = \varepsilon\left\{\mathbf{p} \in [0,\infty)^K : \sum_{k=1}^{K} g_k^\varepsilon(\varepsilon p_k) \geqslant 1\right\} \quad \text{for each } \varepsilon \in (0,1)$$

- $\varepsilon^{-1} R_\varepsilon(F) = \varepsilon^{-1} A_\varepsilon$ does not depend on $\varepsilon \in (0,1)$
- For a fixed $\varepsilon \in (0,1)$ and each k, let $\lambda_k := \varepsilon^{-1} \int_0^\varepsilon g^\varepsilon(x)\mathrm{d}x$ and $f_k : (0,\infty) \to \mathbb{R}$, $x \mapsto g_k^\varepsilon(\varepsilon x)/\lambda_k$ (if $\lambda_k = 0$, then let $f_k := 1$), and further set $f_k(0) = \infty$
- For each k with $\lambda_k \neq 0$,

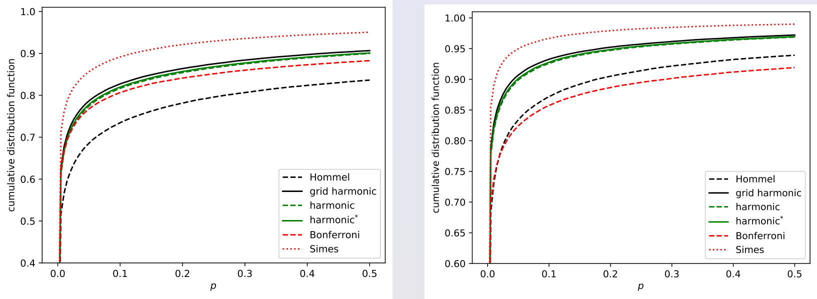
$$\int_0^1 f_k(x)\mathrm{d}x = \frac{\int_0^1 \varepsilon g_k^\varepsilon(\varepsilon x)\mathrm{d}x}{\int_0^1 g_k^\varepsilon(x)\mathrm{d}x} = \frac{\int_0^\varepsilon g_k^\varepsilon(x)\mathrm{d}x}{\int_0^1 g_k^\varepsilon(x)\mathrm{d}x} = 1$$

  $\implies f_k$ is an admissible calibrator
- Converse statement: Markov's inequality

Simulation results

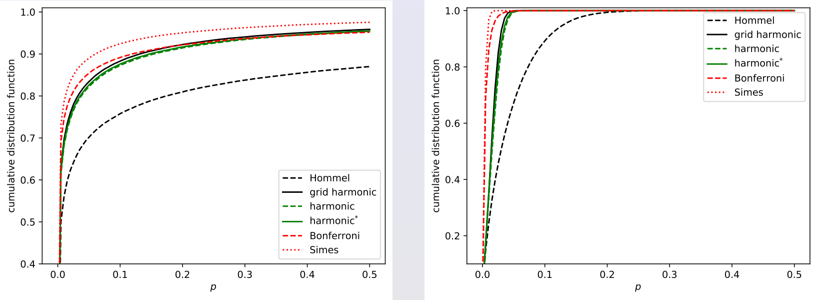
- Correlated z-tests
- $K = 10^6$ observations from $\mathrm{N}(\mu, 1)$
- Pairwise correlation: 0.9
- Last observation: -0.9 correlation with all others
- $\mu = 0$ for null and $\mu = -5$ for alternative
- K_1 observations are drawn from the alternative; the rest from the null
- P-values are $\Phi(x)$
- $F_{-\infty,K}$ (Bonferroni); H_K (Hommel); $F_{-1,K}$ (harmonic); $F^*_{-1,K}$ (harmonic*'); H^*_K (grid harmonic'); S_K (Simes),

Simulation results



$K_1 = 10^3$ (left panel) and $K_1 = 10^4$ (right panel)

Simulation results



$K_1 = 10^5$ with correlation 0.5 (left panel) and 0 (right panel) in place of 0.9

Simulation results

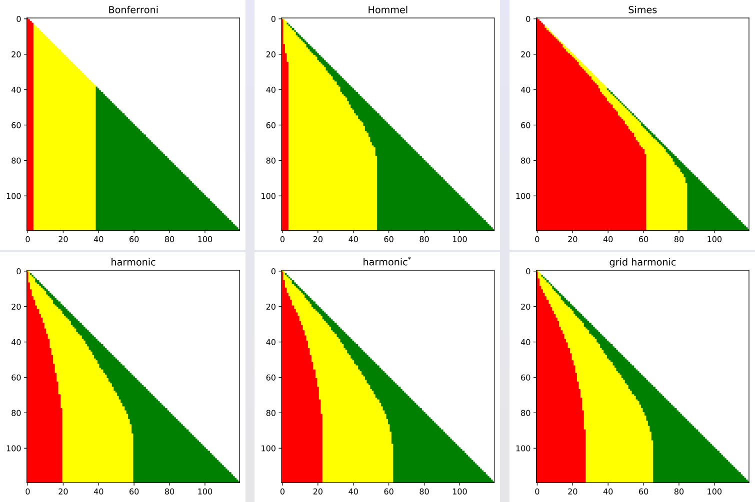
- GWGS discovery matrix Genovese/Wasserman'04; Goeman/Solari'11
- $\mathrm{DM}_{i,j}$: a p-value for testing "there are less than j true discoveries among the i rejected hypotheses"
- $\mathcal{N} \subseteq [K]$: nulls
- Jointly validity: for each $\alpha \in (0,1)$,

$$\mathbb{P}\left(\exists (i,j) \in D_\alpha : \#(R_i \setminus \mathcal{N}) < j\right) \leqslant \alpha$$

where R_i is the set of i hypotheses with smallest p-values and

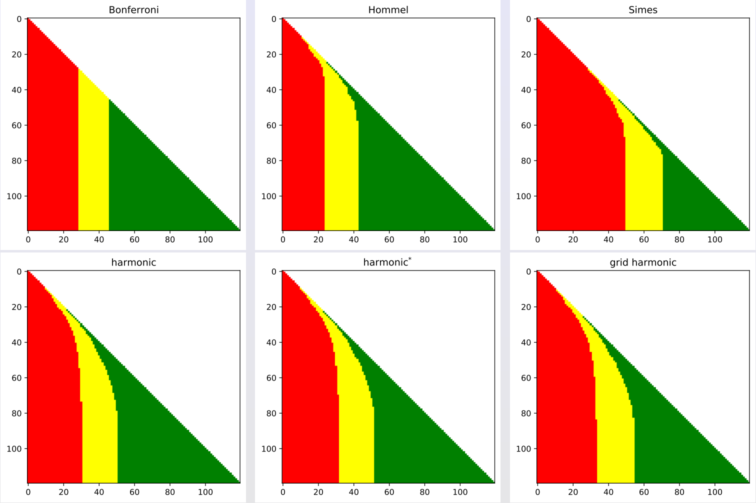
$$D_\alpha = \{(i,j) : \mathrm{DM}_{i,j} \leqslant \alpha\}$$

Simulation results



GWGS discovery matrices with correlation 0.9 and significance levels 1% and 5%

Simulation results



GWGS discovery matrices with correlation 0.5 and significance levels 1% and 5%