Robust Risk Aggregation, Merging P-values, and E-values

Ruodu Wang

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

Department of Statistics and Actuarial Science
University of Waterloo

Department of Statistical Sciences, University of Toronto
Toronto, ON, Canada   December 5, 2019

Agenda

Fundamental problem in Finance/Insurance

Basic setup.

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

Fundamental problem in Finance/Insurance

Basic setup.

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

Calculate $\rho(\Psi(\mathbf{X}))$

Fundamental problem in Finance/Insurance

Basic setup.

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

Calculate $\rho(\Psi(\mathbf{X}))$

Most relevant choices:

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

Fundamental problem in Finance/Insurance

Basic setup.

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

Calculate $\rho(\Psi(\mathbf{X}))$

Most relevant choices:

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

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

Unknown dependence

Model assumption

$$X_i \sim F_i,\ F_i \text{ known with arbitrary dependence},\ i = 1, \dots, n$$

Unknown dependence

Model assumption

$$X_i \sim F_i,\ F_i \text{ known with arbitrary dependence},\ i = 1, \ldots, n$$

Main object

$$\mathcal{S}_n = \mathcal{S}_n(F_1, \ldots, F_n) = \left\{ \sum_{i=1}^{d} X_i : X_i \sim F_i,\ i = 1, \ldots, n \right\}$$

- Every element in $\mathcal{S}_n$ is a possible risk position

Unknown dependence

Model assumption

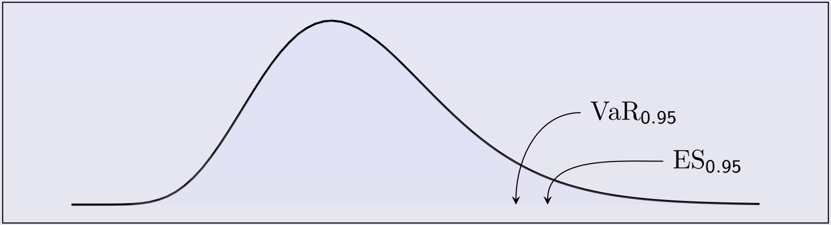
$$X_i \sim F_i,\ F_i \text{ known with arbitrary dependence},\ i = 1, \dots, n$$

Main object

$$\mathcal{S}_n = \mathcal{S}_n(F_1, \dots, F_n) = \left\{ \sum_{i=1}^{d} X_i : X_i \sim F_i,\ \ i = 1, \dots, n \right\}$$

- Every element in $\mathcal{S}_n$ is a possible risk position
- $\mathcal{D}_n = \{$distributions of elements in $\mathcal{S}_n\}$
- Determination of $\mathcal{S}_n$ and $\mathcal{D}_n$: very challenging
- Particular interest: $c \in \mathcal{S}_n$ for some $c \in \mathbb{R}$? $\Rightarrow$ joint mixability

Regulatory risk measures in Basel IV and Solvency II



Value-at-Risk (VaR), $p \in (0,1)$

$\mathrm{VaR}_p : L^0 \to \mathbb{R}$,

$$\mathrm{VaR}_p(X) = q_p(X) = \inf\{x \in \mathbb{R} : \mathbb{P}(X \le x) \ge p\}$$

(left-quantile)

Expected Shortfall (ES), $p \in (0,1)$

$\mathrm{ES}_p : L^1 \to \mathbb{R}$,

$$\mathrm{ES}_p(X) = \frac{1}{1-p}\int_p^1 \mathrm{VaR}_q(X)\mathrm{d}q$$

(also: TVaR/CVaR/AVaR)

Worst- and best-values of VaR and ES

The Fréchet problems

- For $p \in (0,1)$,

$$\overline{\mathrm{VaR}}_p(\mathcal{S}_n) = \sup\{\mathrm{VaR}_p(S) : S \in \mathcal{S}_n(F_1,\dots,F_n)\},$$

$$\underline{\mathrm{VaR}}_p(\mathcal{S}_n) = \inf\{\mathrm{VaR}_p(S) : S \in \mathcal{S}_n(F_1,\dots,F_n)\}.$$

- Same notation for ES_p

Worst- and best-values of VaR and ES

The Fréchet problems

- For $p \in (0,1)$,

$$\overline{\mathrm{VaR}}_p(\mathcal{S}_n) = \sup\{\mathrm{VaR}_p(S) : S \in \mathcal{S}_n(F_1, \dots, F_n)\},$$

$$\underline{\mathrm{VaR}}_p(\mathcal{S}_n) = \inf\{\mathrm{VaR}_p(S) : S \in \mathcal{S}_n(F_1, \dots, F_n)\}.$$

- Same notation for ES_p
- ES is subadditive: $\overline{\mathrm{ES}}_p(\mathcal{S}_n) = \sum_{i=1}^n \mathrm{ES}_p(X_i)$
- $\overline{\mathrm{VaR}}_p(\mathcal{S}_n)$, $\underline{\mathrm{VaR}}_p(\mathcal{S}_n)$, and $\underline{\mathrm{ES}}_p(\mathcal{S}_n)$: generally open questions

Basel III & IV ES calculation

In the Basel FRTB (2019) internal model approach, for market risk:

$$\text{Capital Charge} = \lambda \mathrm{ES}_p \underbrace{\left(\sum_{i=1}^{n} X_i\right)}_{\text{internal model}} + \underbrace{(1-\lambda)\sum_{i=1}^{n} \mathrm{ES}_p(X_i)}_{\overline{\mathrm{ES}}_p(\mathcal{S}_n)},$$

where

- X_i is the total random loss from a risk class, $i = 1, \dots, n$
  - commodity, equity, credit spread, interest rate, exchange
- $T = 10$-day, $p = 0.975$, $\lambda = 0.5$
- ES_p is calculated under a stressed scenario

Dependence uncertainty!

Solvency II SCR calculation

The Basic Solvency Capital Requirement set out in Article 104(1) shall be equal to the following:

$$\text{Basic SCR} = \sqrt{\sum_{i,j} \text{Corr}_{i,j} \times \text{SCR}_i \times \text{SCR}_j}$$

The factor Corr $_{i,j}$ denotes the item set out in row i and in column j of the following correlation matrix:

i \ j	Market	Default	Life	Health	Non-life
Market	1	0,25	0,25	0,25	0,25
Default	0,25	1	0,25	0,25	0,5
Life	0,25	0,25	1	0,25	0
Health	0,25	0,25	0,25	1	0
Non-life	0,25	0,5	0	0	1

Copied from Solvency II, 2009

Unknown/uncertain dependence structure

Statistical examples

- Joint model inference with additional information
- Treatment effect
- Meta-analysis

Some properties of $\mathcal{S}_n$ and $\mathcal{D}_n$

Theorem

For $\lambda \in [0,1]$ and vectors of distributions $\mathbf{F}$ and $\mathbf{G}$:

(i) $\mathcal{D}_n(\mathbf{F}) = \mathcal{D}_n(\sigma(\mathbf{F}))$ for all n-permutations σ.

(ii) $\lambda\mathcal{D}_n(\mathbf{F}) + (1-\lambda)\mathcal{D}_n(\mathbf{G}) \subset \mathcal{D}_n(\lambda\mathbf{F} + (1-\lambda)\mathbf{G})$. In particular,

   (a) $\lambda\mathcal{D}_n(\mathbf{F}) + (1-\lambda)\mathcal{D}_n(\mathbf{F}) = \mathcal{D}_n(\mathbf{F})$.

   (b) $\mathcal{D}_n(\mathbf{F}) \cap \mathcal{D}_n(\mathbf{G}) \subset \mathcal{D}_n(\lambda\mathbf{F} + (1-\lambda)\mathbf{G})$.

(iii) $\mathcal{D}_n$ is closed under weak convergence.

(iv) $\mathcal{D}_n(\mathbf{F}) \subset \mathcal{D}_n(F_A, \ldots, F_A)$ where F_A is the average of $\mathbf{F}$.

Bernard-Jiang-W., Risk aggregation with dependence uncertainty. Insurance: Mathematics and Economics, Theorems 2.1 and 3.5

Aggregation of Cauchy random variables

Theorem

Let $c \in \mathbb{R}$. There exist standard Cauchy random variables $X_1, \dots, X_n$ such that $(X_1 + \cdots + X_n)/n = c$ if and only if

$$|c| \leq \frac{\log(n-1)}{\pi}.$$

- $\mathbb{P}((X_1 + \cdots + X_n)/n \geq \log(n-1)/\pi) = 1.$

Puccetti-Rigo-Wang-W., Centers of probability measures without the mean. Journal of Theoretical Probability, 2019, Theorem 4.2

Aggregation of uniform random variables

Theorem

For any random variable X and $n \geq 3$, there exist standard uniform random variables $X_1, \dots, X_n$ such that $(X_1 + \cdots + X_n)/n \overset{d}{=} X$ if and only if

$$X \overset{d}{=} \mathbb{E}[X_1 | \mathcal{G}] \quad \text{for some } \sigma\text{-field } \mathcal{G}.$$

- Not true for $n = 2$; $\mathcal{D}_2(F_U, F_U)$ is an open question

---

Mao-Wang-W., Sums of standard uniform random variables. Journal of Applied Probability, 2019, Theorem 5

Aggregation of normal random variables

Theorem

For $i = 1, \dots, n$, let F_i be normal (uniform, t, or normal mixture) with scale parameter $\sigma_i > 0$. There exists a constant c in $\mathcal{S}_n(F_1, \dots, F_n)$ if and only if

$$2 \bigvee_{i=1}^{n} \sigma_i \le \sum_{i=1}^{n} \sigma_i.$$

- If exists, $c = \sum_{i=1}^{n} \mu_i$

W.-Peng-Yang, Bounds for the sum of dependent risks and worst Value-at-Risk with monotone marginal densities. Finance and Stochastics, 2013, Proposition 2.4

Aggregation with decreasing densities

Theorem

For $i = 1, \ldots, n$, let F_i be a distribution with mean μ_i and decreasing density on a bounded support $[a_i, a_i + \ell_i]$. There exists a constant $c \in \mathcal{S}_n(F_1, \ldots, F_n)$ if and only if

$$2 \bigvee_{i=1}^{n} \ell_i \leq \sum_{i=1}^{n} (\mu_i - a_i) + \bigvee_{i=1}^{n} \ell_i \leq \sum_{i=1}^{n} \ell_i.$$

- If exists, $c = \sum_{i=1}^{n} \mu_i$

Wang-W., Joint mixability.
Mathematics of Operations Research, 2016, Theorem 3.2

Quantile aggregation

Theorem

Let $\alpha_1, \dots, \alpha_n, \beta_1, \dots, \beta_n > 0$ with $\gamma = \sum_{i=1}^n \alpha_i + \bigvee_{i=1}^n \beta_i < 1$, $F_1, \dots, F_n$ be any distributions, and $F \in \mathcal{D}_n(F_1, \dots, F_n)$. Then

$$F^{-1}(1-\gamma) \leq \sum_{i=1}^n \int_{\alpha_i}^{\alpha_i+\beta_i} F_i^{-1}(1-t)\mathrm{d}t.$$

- Limit case:

$$F^{-1}\left(1 - \sum_{i=1}^n \alpha_i\right) \leq \sum_{i=1}^n F_i^{-1}(1-\alpha_i).$$

Embrechts-Liu-W., Quantile-based risk sharing.
Operations Research, 2018, Theorem 1

Results on VaR (quantile) aggregation

$d = 2$

- solved analytically (Makarov'81, Rüschendorf'82)
- based on counter-monotonicity

Results on VaR (quantile) aggregation

$d = 2$

- solved analytically (Makarov'81, Rüschendorf'82)
- based on counter-monotonicity

$d \geq 3$

- dual bounds (Embrechts-Puccett'06)
- solved analytically for monotone densities

Results on VaR (quantile) aggregation

$d = 2$

- solved analytically (Makarov'81, Rüschendorf'82)
- based on counter-monotonicity

$d \geq 3$

- dual bounds (Embrechts-Puccett'06)
- solved analytically for monotone densities
  - homogeneous case (W.-Peng-Yang'13)
  - heterogeneous case (Jakobsons-Han-W.'16)
  - based on joint-mixability
- generalization to other distributions is limited

Results on VaR (quantile) aggregation

Remarks.

- Efficient numerical algorithm: the Rearrangement Algorithm
  - Puccetti-Rüschendorf'12, Embrechts-Puccetti-Rüschendorf'13, Bernard-Bondarenko-Vanduffel'18, ...

Results on VaR (quantile) aggregation

Remarks.

- Efficient numerical algorithm: the Rearrangement Algorithm
  - Puccetti-Rüschendorf'12, Embrechts-Puccetti-Rüschendorf'13, Bernard-Bondarenko-Vanduffel'18, ...
- Risk aggregation with partial dependence information
  - Puccetti-Rüschendorf-Manko'16, Bernard-Rüschendorf-Vanduffel'17, Lux-Papapantoleon'17, Bernard-Rüschendorf-Vanduffel-W.'17, ...

Results on VaR (quantile) aggregation

Remarks.

- Efficient numerical algorithm: the Rearrangement Algorithm
  - Puccetti-Rüschendorf'12, Embrechts-Puccetti-Rüschendorf'13, Bernard-Bondarenko-Vanduffel'18, ...
- Risk aggregation with partial dependence information
  - Puccetti-Rüschendorf-Manko'16, Bernard-Rüschendorf-Vanduffel'17, Lux-Papapantoleon'17, Bernard-Rüschendorf-Vanduffel-W.'17, ...
- Risk aggregation with marginal and dependence uncertainty
  - Li-Shao-W.-Yang'18, Blanchet-Murthy'18, ...

Results on VaR (quantile) aggregation

Remarks.

- Efficient numerical algorithm: the Rearrangement Algorithm
  - Puccetti-Rüschendorf'12, Embrechts-Puccetti-Rüschendorf'13, Bernard-Bondarenko-Vanduffel'18, ...
- Risk aggregation with partial dependence information
  - Puccetti-Rüschendorf-Manko'16, Bernard-Rüschendorf-Vanduffel'17, Lux-Papapantoleon'17, Bernard-Rüschendorf-Vanduffel-W.'17, ...
- Risk aggregation with marginal and dependence uncertainty
  - Li-Shao-W.-Yang'18, Blanchet-Murthy'18, ...
- Connection to distributionally robust optimization
  - Gao-Kleywegt'17, ...

Combining p-values via averaging

Based on joint work with Vladimir Vovk (CS @ Royal Holloway)

P-values

STAT 101

A p-value P for testing a hypothesis H_0:

- Uniform on $[0,1]$ under H_0 $\Leftrightarrow$ $\mathbb{P}^{H_0}(P \le \epsilon) = \epsilon$ for $\epsilon \in [0,1]$
  - $\sup_{H \in H_0} \mathbb{P}^{H}(P \le \epsilon) \le \epsilon$ in case H_0 is a set of hypotheses

P-values

STAT 101

A p-value P for testing a hypothesis H_0:

- Uniform on $[0,1]$ under $H_0 \Leftrightarrow \mathbb{P}^{H_0}(P \le \epsilon) = \epsilon$ for $\epsilon \in [0,1]$
  - $\sup_{H \in H_0} \mathbb{P}^{H}(P \le \epsilon) \le \epsilon$ in case H_0 is a set of hypotheses
- A significance level α, typically $\alpha = 0.05,\ 0.01,\ 0.005 \dots$

P-values

STAT 101

A p-value P for testing a hypothesis H_0:

- Uniform on $[0,1]$ under $H_0 \Leftrightarrow \mathbb{P}^{H_0}(P \le \epsilon) = \epsilon$ for $\epsilon \in [0,1]$
  - $\sup_{H \in H_0} \mathbb{P}^{H}(P \le \epsilon) \le \epsilon$ in case H_0 is a set of hypotheses
- A significance level α, typically $\alpha = 0.05,\ 0.01,\ 0.005 \ldots$
- Rejects H_0 if (realized) $P \le \alpha$
  - cannot reject H_0 if $P > \alpha$

P-values

STAT 101

A p-value P for testing a hypothesis H_0:

- Uniform on $[0, 1]$ under $H_0 \Leftrightarrow \mathbb{P}^{H_0}(P \le \epsilon) = \epsilon$ for $\epsilon \in [0, 1]$
  - $\sup_{H \in H_0} \mathbb{P}^{H}(P \le \epsilon) \le \epsilon$ in case H_0 is a set of hypotheses
- A significance level α, typically $\alpha = 0.05,\ 0.01,\ 0.005 \ldots$
- Rejects H_0 if (realized) $P \le \alpha$
  - cannot reject H_0 if $P > \alpha$
- Probability of type I error $= \mathbb{P}^{H_0}(\text{reject } H_0) \le \alpha$

Merging p-values

Suppose we are testing the same hypothesis using $K \geq 2$ different statistical tests and obtain p-values $p_1, \ldots, p_K$. How can we combine them into a single p-value?

Merging p-values

Suppose we are testing the same hypothesis using $K \geq 2$ different statistical tests and obtain p-values $p_1, \dots, p_K$. How can we combine them into a single p-value?

Examples.

- backtesting credit risk ratings: typically 17 binomial tests
- backtesting market risk models: several quantile level tests
- meta-analysis
- genome-wide association studies (GWAS)

Meta-analysis

A typical example from meta-analysis

TABLE 1
Data on 10 Studies of Sex Differences in Conformity Using the Fictitious Norm Group Paradigm

Study	Sample size: Control n^C	Sample size: Experimental n^E	Effect size d	Student's t	Significance level p	$-2 \log p$	$\Phi^{-1}(p)$	$\log[p/(1-p)]$
1	118	136	0.35	2.78	0.0029	11.682	−2.758	−5.838
2	40	40	0.37	1.65	0.0510	5.952	−1.635	−2.923
3	61	64	−0.06	−0.33	0.6310	0.921	0.335	0.537
4	77	114	−0.30	−2.03	0.9783	0.044	2.020	3.809
5	32	32	0.70	2.80	0.0034	11.367	−2.706	−5.680
6	45	45	0.40	1.90	0.0305	6.978	−1.873	−3.458
7	30	30	0.48	1.86	0.0341	6.760	−1.824	−3.345
8	10	10	0.85	1.90	0.0367	6.608	−1.790	−3.266
9	70	71	−0.33	−1.96	0.9740	0.053	1.942	3.622
10	60	59	0.07	0.38	0.3517	2.090	−0.381	−0.612

The `sex differences` dataset, from p.35 of Hedges-Olkin'85

The Bonferroni method

A question of a long history

- Tippett'31, Pearson'33, Fisher'48: assume independence

The Bonferroni method

A question of a long history

- Tippett'31, Pearson'33, Fisher'48: assume independence

Without any assumptions on the p-values $p_1, \dots, p_K$ ...

- The Bonferroni method (Dunn'58):

$$F(p_1, \dots, p_K) = K \min(p_1, \dots, p_K).$$

The Bonferroni method

A question of a long history

- Tippett'31, Pearson'33, Fisher'48: assume independence

Without any assumptions on the p-values $p_1, \ldots, p_K$ ...

- The Bonferroni method (Dunn'58):

$$F(p_1, \ldots, p_K) = K \min(p_1, \ldots, p_K).$$

- Rüger'78:

$$F(p_1, \ldots, p_K) = \frac{K}{k} p_{(k)}.$$

In particular, 2 times the median or the maximum.

The Bonferroni method

A question of a long history

- Tippett'31, Pearson'33, Fisher'48: assume independence

Without any assumptions on the p-values $p_1, \dots, p_K$ ...

- The Bonferroni method (Dunn'58):

$$F(p_1, \dots, p_K) = K \min(p_1, \dots, p_K).$$

- Rüger'78:

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

  In particular, 2 times the median or the maximum.

- Hommel'83; Simes'86:

$$F(p_1, \dots, p_K) = \left(1 + \frac{1}{2} + \dots + \frac{1}{K}\right) \bigwedge_{k=1}^{K} \frac{K}{k} p_{(k)}.$$

The Bonferroni method

The Bonferroni method

- overly conservative ... if tests are similar
- dictated by a single experiment (contamination?)
- what if some p-values are more important (e.g. bigger experiments)?

The Bonferroni method

The Bonferroni method

- overly conservative ... if tests are similar
- dictated by a single experiment (contamination?)
- what if some p-values are more important (e.g. bigger experiments)?

Particular interest: heavily but not nicely dependent tests.

Merging functions

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

Definition (p-variables and merging functions)

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

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

Merging functions

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

Definition (p-variables and merging functions)

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

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

(ii) A merging function is an increasing Borel function $F : [0,1]^K \to [0,\infty)$ 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
- Merging functions may be applied iteratively in multiple layers

Merging functions

For an increasing Borel function $F : [0,1]^K \to [0,\infty)$, equivalent are:

- F is a merging function w.r.t. some collection $\mathcal{H}$;
- F is a merging function w.r.t. all collections $\mathcal{H}$;
- fixing $\mathbb{P}$, $F(U_1,\dots,U_K)$ is a p-variable for all $U_1,\dots,U_K \in \mathcal{U}$;
- fixing $\mathbb{P}$, for all $\epsilon \in (0,1)$, $\overline{\mathbb{P}}(F \le \epsilon) \le \epsilon$, where

$$\overline{\mathbb{P}}(F \le \epsilon) = \sup\left\{\mathbb{P}(F(U_1,\dots,U_K) \le \epsilon) \mid U_1,\dots,U_K \in \mathcal{U}\right\}.$$

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

Merging functions

For an increasing Borel function $F : [0,1]^K \to [0,\infty)$, equivalent are:

- F is a merging function w.r.t. some collection $\mathcal{H}$;
- F is a merging function w.r.t. all collections $\mathcal{H}$;
- fixing $\mathbb{P}$, $F(U_1, \dots, U_K)$ is a p-variable for all $U_1, \dots, U_K \in \mathcal{U}$;
- fixing $\mathbb{P}$, for all $\epsilon \in (0,1)$, $\overline{\mathbb{P}}(F \le \epsilon) \le \epsilon$, where

$$\overline{\mathbb{P}}(F \le \epsilon) = \sup \{\mathbb{P}(F(U_1, \dots, U_K) \le \epsilon) \mid U_1, \dots, U_K \in \mathcal{U}\}.$$

It is sufficient to consider $\mathcal{H} = \{\mathbb{P}\}$ for a generic $\mathbb{P}$

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

Precise merging functions

Definition (precise merging functions)

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

Precise merging functions

Definition (precise merging functions)

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

Examples.

- The Bonferroni method $F(p_1, \dots, p_K) = K \min(p_1, \dots, p_K)$
- $F(p_1, \dots, p_K) = \max(p_1, \dots, p_K)$
- $F(p_1, \dots, p_K) = p_1$ (trivial)

Precise merging functions

The Bonferroni method $F(p_1, \dots, p_K) = K \min(p_1, \dots, p_K)$

$$\mathbb{P}(K \min(p_1, \dots, p_K) \le \epsilon) = \mathbb{P}\left(\bigcup_{i=1}^{K} \{Kp_i \le \epsilon\}\right)$$
$$\le \sum_{i=1}^{K} \mathbb{P}\left(Kp_i \le \epsilon\right)$$
$$= \sum_{i=1}^{K} \frac{\epsilon}{K} = \epsilon.$$

The inequality is an equality if $\{Kp_i \le \epsilon\}$, $i = 1, \dots, K$ are mutually exclusive.

Merging p-values via averaging

A general notion of averaging

- Axiomatized by 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 and strictly monotonic.

Merging p-values via averaging

A general notion of averaging

- Axiomatized by 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 and strictly monotonic.

- Most common forms, for $r\in\mathbb{R}\setminus\{0\}$,

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

Merging p-values via averaging

A general notion of averaging

- Axiomatized by 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 and strictly monotonic.

- Most common forms, for $r\in\mathbb{R}\setminus\{0\}$,

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

- $\phi(x)=\tan((x-\frac{1}{2})\pi)$: Cauchy combination test (Liu-Xie'19)

Merging p-values via averaging

Special cases:

- Arithmetic: $M_{1,K}(p_1, \ldots, p_K) = \frac{1}{K}\sum_{k=1}^{K} p_k$
- Harmonic: $M_{-1,K}(p_1, \ldots, p_K) = \left(\frac{1}{K}\sum_{k=1}^{K} \frac{1}{p_k}\right)^{-1}$
- Quadratic: $M_{2,K}(p_1, \ldots, p_K) = \sqrt{\frac{1}{K}\sum_{k=1}^{K} p_k^2}$

Merging p-values via averaging

Special cases:

- Arithmetic: $M_{1,K}(p_1, \ldots, p_K) = \frac{1}{K} \sum_{k=1}^{K} p_k$
- Harmonic: $M_{-1,K}(p_1, \ldots, p_K) = \left(\frac{1}{K} \sum_{k=1}^{K} \frac{1}{p_k} \right)^{-1}$
- Quadratic: $M_{2,K}(p_1, \ldots, p_K) = \sqrt{\frac{1}{K} \sum_{k=1}^{K} p_k^2}$

Limiting cases:

- Geometric: $M_{0,K}(p_1, \ldots, p_K) = \left(\prod_{k=1}^{K} p_k \right)^{1/K}$
- Maximum: $M_{\infty,K}(p_1, \ldots, p_K) = \max(p_1, \ldots, p_K)$
- Minimum: $M_{-\infty,K}(p_1, \ldots, p_K) = \min(p_1, \ldots, p_K)$

Merging p-values via averaging

Special cases:

- Arithmetic: $M_{1,K}(p_1,\dots,p_K) = \frac{1}{K}\sum_{k=1}^{K} p_k$
- Harmonic: $M_{-1,K}(p_1,\dots,p_K) = \left(\frac{1}{K}\sum_{k=1}^{K}\frac{1}{p_k}\right)^{-1}$
- Quadratic: $M_{2,K}(p_1,\dots,p_K) = \sqrt{\frac{1}{K}\sum_{k=1}^{K} p_k^2}$

Limiting cases:

- Geometric: $M_{0,K}(p_1,\dots,p_K) = \left(\prod_{k=1}^{K} p_k\right)^{1/K}$
- Maximum: $M_{\infty,K}(p_1,\dots,p_K) = \max(p_1,\dots,p_K)$
- Minimum: $M_{-\infty,K}(p_1,\dots,p_K) = \min(p_1,\dots,p_K)$

The cases $r \in \{-1, 0, 1\}$ are known as Platonic means.

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 merging function (Rüschendorf'82, Meng'93):

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

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 merging function (Rüschendorf'82, Meng'93):

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

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

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 merging function (Rüschendorf'82, Meng'93):

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

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

Task. Find $b_{r,K} > 0$ such that $b_{r,K} M_{r,K}$ is a precise merging function

Merging p-values via averaging

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

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

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

Task. Find $b_{r,K} > 0$ such that $b_{r,K} M_{r,K}$ is a precise merging function

- $M_{r,K}$ increases in r
  - The constants $b_{r,K}$ should decrease in r.

Translation to a risk aggregation problem

For $\alpha \in (0,1]$ and a random variable X, define

$$q_\alpha(X) = \inf\{x \in \mathbb{R} : \mathbb{P}(X \le x) \ge \alpha\} = \mathrm{VaR}_\alpha(X).$$

and for a function $F : [0,1]^K \to [0,\infty)$, define

$$\underline{q}_\alpha(F) = \inf\left\{q_\alpha(F(U_1,\dots,U_K)) \mid U_1,\dots,U_K \in \mathcal{U}\right\}.$$

Translation to a risk aggregation problem

Lemma

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

(i) F is a merging function, i.e. $\overline{\mathbb{P}}(F \leq \epsilon) \leq \epsilon$ for all $\epsilon \in (0, 1)$;

(ii) $\underline{q}_\epsilon(F) \geq \epsilon$ for all $\epsilon \in (0, 1)$;

(iii) $\overline{\mathbb{P}}(F \leq \epsilon) \leq \epsilon$ for some $\epsilon \in (0, 1)$;

(iv) $\underline{q}_\epsilon(F) \geq \epsilon$ for some $\epsilon \in (0, 1)$.

The same conclusion holds if all $\leq$ and $\geq$ are replaced by $=$.

- In statistical practice one only needs to have $\overline{\mathbb{P}}(F \leq \epsilon) \leq \epsilon$ for a specific ϵ, e.g. 0.05, 0.01, ...

Translation to a risk aggregation problem

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

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

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

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

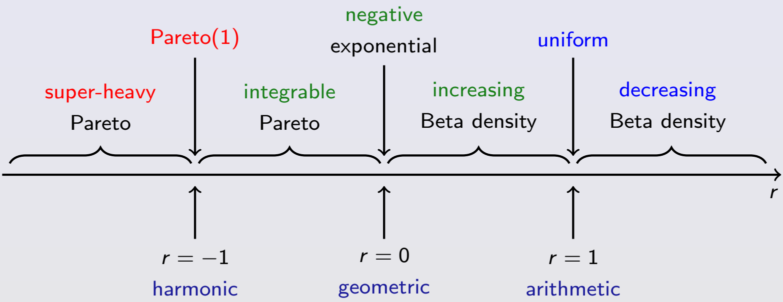
$$\log(\underline{q}_\epsilon(M_{r,K})) = \inf_{U_1,\dots,U_K\in\mathcal{U}} \left\{ q_\epsilon \left(\frac{1}{K} (\log U_1 + \cdots + \log U_K) \right) \right\}$$

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

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

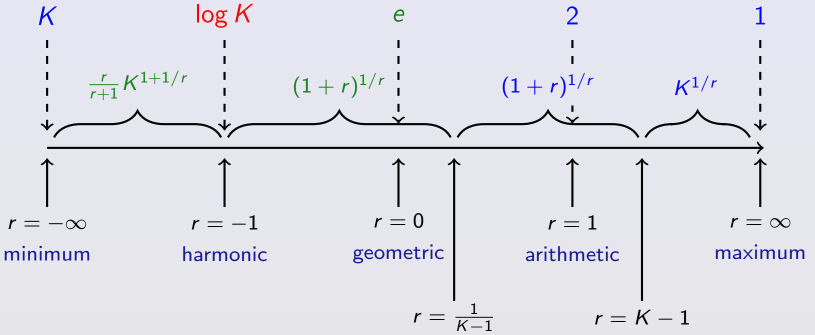
Translation to a risk aggregation problem

Breakdown of U^r (or $\log U$) for $r \in \mathbb{R}$



Main results summary

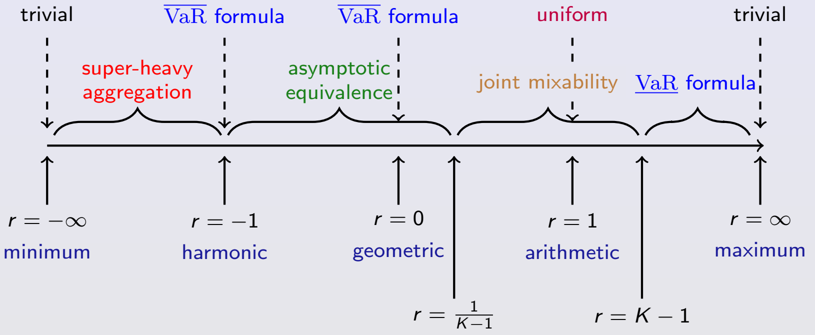
Constant multiplier in front of $M_{r,K}$



blue: precise; green: asymptotically precise; red: limit

Main results summary

Methodology breakdown ▸ details



purple: Rüshcendorf'82; blue: W.-Peng-Yang'13; brown: Wang-W'11
green: Wang-W.'15; red: Bignozzi-Mao-Wang-W.'16

Weighted averaging

Consider weighted averaging functions

$$M_{\phi,\mathbf{w}}(p_1, \dots, p_K) = \phi^{-1}\left(w_1\phi(p_1) + \cdots + w_K\phi(p_K)\right),$$

and in particular,

$$M_{r,\mathbf{w}}(p_1, \dots, p_K) = \left(w_1 p_1^r + \cdots + w_K p_K^r\right)^{1/r},$$

where $\mathbf{w} = (w_1, \dots, w_K) \in \Delta_K$.

- Intuitively, the weights reflect the prior importance of the p-values.

$\Delta_K = \left\{(w_1, \dots, w_K) \in [0,1]^K \mid w_1 + \cdots + w_K = 1\right\}$ is the standard K-simplex.

Weighted averaging

Proposition

For $\mathbf{w} = (w_1, \ldots, w_K) \in \Delta_K$, $w = \max(\mathbf{w})$ and $r \in (-1, \infty)$,

(i) $(r+1)^{1/r} M_{r,\mathbf{w}}$ is a merging function;

(ii) $(r+1)^{1/r} M_{r,\mathbf{w}}$ is precise $\Leftrightarrow$ $w \le 1/2$ and $r \in [\frac{w}{1-w}, \frac{1-w}{w}]$;

(iii) if $r \in [1, \infty)$, $\min(r+1, \frac{1}{w})^{1/r} M_{r,\mathbf{w}}$ is a precise merging function.

Weighted averaging

Conjecture

For $a > 0$ and any r and K, if $aM_{r,K}$ is a merging function, then $aM_{r,\mathbf{w}}$ is also a merging function for all $\mathbf{w} \in \Delta_K$.

(Proof available for $r \leq -1$ and $r \geq 1/(K-1)$)

Weighted averaging

Conjecture

For $a > 0$ and any r and K, if $aM_{r,K}$ is a merging function, then $aM_{r,\mathbf{w}}$ is also a merging function for all $\mathbf{w} \in \Delta_K$.

(Proof available for $r \le -1$ and $r \ge 1/(K-1)$)

A deeper conjecture: under some conditions

$$\mathcal{D}_n(F_1, \dots, F_n) \subset \mathcal{D}_n(F_H, \dots, F_H), \quad \text{where} \quad F_H^{-1} = \frac{1}{n}\sum_{i=1}^n F_i^{-1}.$$

Bernard-Jiang-W.'14, Theorem 3.5:

$$\mathcal{D}_n(F_1, \dots, F_n) \subset \mathcal{D}_n(F_A, \dots, F_A), \quad \text{where} \quad F_A = \frac{1}{n}\sum_{i=1}^n F_i.$$

Robust p-merging: admissibility

Admissibility and domination structure

- A merging function F dominates another merging function G if $F \leq G$.
- A merging function is admissible if it is not dominated by any other merging functions.
  - We also consider admissibility within a family
- For $r \in [-\infty, \infty]$ and $K \geq 2$, $b_{r,K}$ is the constant such that $b_{r,K} M_{r,K}$ is a precise merging function.
- We write $F_{r,K} = b_{r,K} M_{r,K}$.

Robust p-merging: admissibility

Lemma

(i) If $r < s$, then $b_{s,K} \leq b_{r,K}$.

(ii) If $r < s$ and $rs > 0$, then $b_{r,K} K^{-1/r} \leq b_{s,K} K^{-1/s}$.

- For $r < s$ and $rs > 0$,

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

$\Rightarrow$ continuity of $b_{r,K}$ for $r \in [-\infty, 0) \cup (0, \infty]$.

Robust p-merging: admissibility

Proposition

For $r < s$ and $K \geq 2$, the following statements hold.

(i) $F_{r,K}$ dominates $F_{s,K}$ if and only if $b_{r,K} = b_{s,K}$.

(ii) If $rs > 0$, then $F_{s,K}$ dominates $F_{r,K}$ if and only if $b_{r,K}K^{-1/r} = b_{s,K}K^{-1/s}$.

(iii) If $rs \leq 0$, then $F_{s,K}$ does not dominate $F_{r,K}$.

- Both (i) and (ii) may happen in some cases; $F_{r,K}$ is not necessarily admissible even within the family $(F_{r,K})_{r\in[-\infty,\infty]}$.
- Example. $F_{1,2}(p_1, p_2) = p_1 + p_2$ is dominated by every other member of the family, although it is precise.

Robust p-merging: admissibility

Theorem

(i) All admissible merging functions are precise.

(ii) $F_{-\infty,K}$ is admissible among all merging functions.

(iii) $F_{\infty,K}$ is admissible among all symmetric and continuous merging functions.

(iv) $F_{1,K}$ is admissible within the family $(F_{r,K})_{r\in[-\infty,\infty]}$ for $K \geq 3$.

(v) The merging functions $F_{r,K}$ and $F_{s,K}$ do not dominate each other for $r \neq s$ and K large enough.

Robust p-merging: efficiency

Among (admissible) merging methods for $r \in [-\infty, \infty]$:

- Which method is the most efficient? In which situation?
- Requires the distributions of p-values under alternative hypotheses
  - p-values from different experiments tend to be highly heterogeneous
  - impossible to make inference of their dependence structure
  - an adaptive learning method is difficult to design
- $\Rightarrow$ this relies on prior or side information
- Some results on correlated z-tests are obtained

details

Data-driven choices

General form: for some $r_1, \ldots, r_m \in [-\infty, \infty]$,

$$F(p_1, \ldots, p_K) = b \sum_{i=1}^{m} F_{r_i,K}(p_1, \ldots, p_K) \mathbb{1}_{A_i}(p_1, \ldots, p_K)$$

- $(A_1, \ldots, A_m)$ is a partition of $[0, 1]^K$
- $b > 0$ is a constant so that F is a valid merging function

Special case:

$$F(p_1, \ldots, p_K) = b \min_{i=1,\ldots,m} F_{r_i,K}(p_1, \ldots, p_K).$$

- b: the price to pay to exploit the power of different methods
- $b = m$ is always valid (finding optimal $b \Rightarrow$ open question)

Compound methods

Consider the compound Bonferroni-arithmetic (BA) method

$$F_K^{\mathrm{BA}} = 2\min\left(KM_{-\infty,K}, 2M_{1,K}\right)$$

and the compound Bonferroni-geometric (BG) method

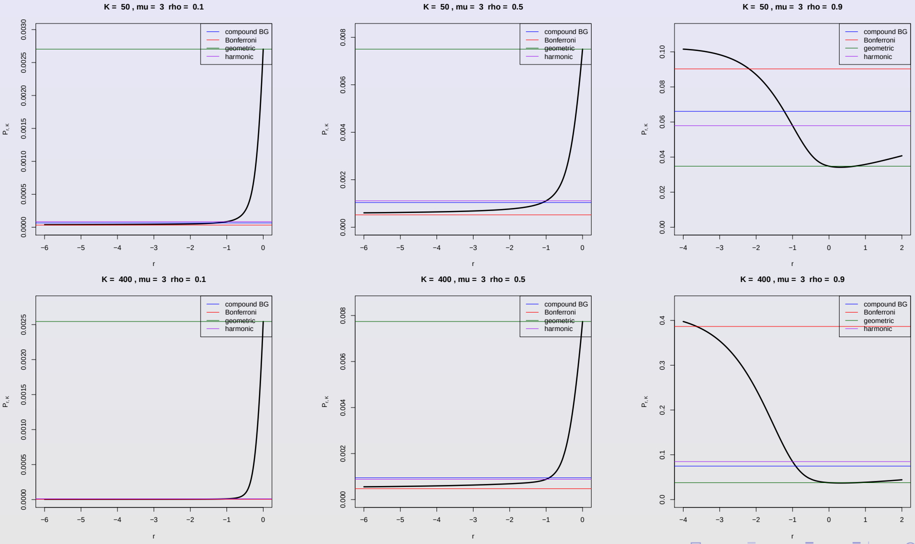
$$F_K^{\mathrm{BG}} = 2\min\left(KM_{-\infty,K}, eM_{0,K}\right)$$

Proposition

Both families of merging functions F_K^{BA} and F_K^{BG}, $K = 2, 3, \dots$ are asymptotically precise.

- The price to pay for exploiting the power of Bonferroni and arithmetic/geometric methods is precisely a factor of 2.

Simulation: $\mathbb{E}[P_{r,K}]$ for finite K



Efficiency: a rule of thumb

- stronger dependence $\Rightarrow$ higher r
- independence $\Rightarrow r \le -1$
- finite K: Bonferroni performs well for small to moderate ρ
- mixed-merging: the compound BG method performs very well for unknown dependence

E-values

E-value: non-negative random variable E with mean 1.

- Related to Bayesian factor:

$$E(\text{Obs.}) = \frac{\Pr(\text{Obs.} \mid \mathbb{Q})}{\Pr(\text{Obs.} \mid \mathbb{P})}.$$

- $E(\text{Obs.})$ very large $\Rightarrow$ reject
- Alternative to p-values
- Also related to the algorithmic theory of randomness of Kolmogorov'65, 68

E-values, robust e-merging, and calibrators

Again, let $\mathcal{H}$ be a collection of atomless probability measures ...

Definition (e-variables, e-merging functions, and calibrators)

(i) An e-variable is a non-negative random variable E that satisfies $\sup_{\mathbb{P}\in\mathcal{H}} \int E \,\mathrm{d}\mathbb{P} \le 1$.

(ii) An e-merging function is an increasing Borel function $F : [0,\infty]^K \to [0,\infty]$ such that $F(E_1,\dots,E_K)$ is an e-variable for all e-variables $E_1,\dots,E_K$.

(iii) A p-to-e calibrator is a decreasing function $f : [0,1] \to [0,\infty]$ such that $f(P)$ is an e-variable for all p-variables P.

(iv) An e-to-p calibrator is a decreasing function $g : [0,\infty] \to [0,1]$ such that $g(E)$ is an p-variable for all e-variables E.

Characterization of calibrators

Proposition (Shafer-Shen-Vereshchagin-Vovk'11)

A decreasing function $f : [0,1] \to [0,\infty]$ is a p-to-e calibrator if and only if $\int_0^1 f(x)\mathrm{d}x \le 1$. It is admissible if and only if f is upper semicontinuous, $f(0) = \infty$, and $\int_0^1 f(x)\mathrm{d}x = 1$.

Proposition

The function $f : [0,\infty] \to [0,1]$ defined by $f(t) = \min(1, 1/t)$ is an e-to-p calibrator. It is the only admissible e-to-p calibrator.

- $1/e$ is a p-value for any e-value e
- $\kappa p^{\kappa-1}$ is a p-value for any p-value p and $\kappa \in (0,1)$
- In the algorithmic theory of randomness, roughly $p \sim 1/e$

Characterization of e-merging functions

Proposition

A symmetric e-merging function F satisfying $F(0, \dots, 0) = 0$ is admissible if and only if it is the arithmetic mean.

- Admissibility of p-merging functions is quite complicated
- Similar for p-to-e merging and e-to-p merging functions

Conjecture

F is an admissible e-merging function if and only if $\mathbb{E}[F(E_1, \dots, E_K)] = 1$ for e-variables $E_1, \dots, E_K$ with mean 1.

("if" part is true; "only-if" part is true for symmetric functions)

Test supermartingales

Another important e-merging function is

$$F(e_1, \ldots, e_K) = \prod_{k=1}^{K} e_k,$$

valid for independent e-values.

- E-values $e_1, \ldots, e_K$ are obtained by laboratories $1, \ldots, K$
- Laboratory k makes sure that its result e_k is a valid e-value given the previous results $e_1, \ldots, e_{k-1}$
- $\mathbb{E}[E_k \mid E_1, \ldots, E_{k-1}] \le 1$ for all $k \in \{1, \ldots, K\}$
- $\prod_{k=1}^{K} E_k$ is a test supermartingale and is an e-variable

Concluding remarks

Statistical questions

- Power analysis for other classic statistical models
- Adaptive selection of the merging function
- Relation between prior knowledge of dependence and the optimal choice of merging functions
- Domination structure of other merging methods
- The price of robustness for different methods
- Admissibility of of p-to-e and e-to-p merging functions
- Choice of $\mathbf{w}$: robustness-power tradeoff (e.g., entropy regularized choice?)

Open questions on risk aggregation

Mathematical questions on robust risk aggregation:

- Characterization of $\mathcal{S}_n$, $\mathcal{D}_n$ and joint mixability
- Analytical formulas for $\overline{\mathrm{VaR}}_p$, $\underline{\mathrm{VaR}}_p$ and $\underline{\mathrm{ES}}_p$
- Aggregation of random vectors
- Partial information on dependence
- RDU and CPT risk aggregation
- Other aggregation functionals

Open questions on risk aggregation

A few concrete mathematical questions:

- For a given F, determine whether $F \in \mathcal{D}_2(\mathrm{U}[0,1], \mathrm{U}[0,1])$?
- For a given correlation matrix Σ and $F_1, \dots, F_n$, determine whether

$$\mathcal{V}_\Sigma = \{\mathbf{X} : \mathrm{Corr}(\mathbf{X}) = \Sigma,\ X_i \sim F_i,\ i = 1, \dots, n\}$$

  is empty?
- If $\mathcal{V}_\Sigma \neq \varnothing$, what are the values of

$$\sup\{\mathrm{VaR}_p(S) : \mathbf{X} \in \mathcal{V}_\Sigma\} \text{ and } \sup\{\mathrm{ES}_p(S) : \mathbf{X} \in \mathcal{V}_\Sigma\}?$$

  Here $S = X_1 + \cdots + X_n$.

Thank you

Thank you for your kind attention

Based on

- Vovk-W., Combining p-values via averaging. Biometrika, 2019. SSRN: 3166304
- Vovk-W., Admissibility of p-value merging methods. Working paper, 2019.
- Vovk-W., Combining and calibrating e-values. Working paper, 2019.

References I

- Bernard, C., Jiang, X. and Wang, R. (2014). Risk aggregation with dependence uncertainty. *Insurance: Mathematics and Economics*, **54**, 93–108.
- Bignozzi, V., Mao, T., Wang, B. and Wang, R. (2016). Diversification limit of quantiles under dependence uncertainty. *Extremes*, **19**(2), 143–170.
- Embrechts, P. and Puccetti, G. (2006). Bounds for functions of dependent risks. *Finance and Stochastics*, **10**, 341–352.
- Jakobsons, E., Han, X. and Wang, R. (2016). General convex order on risk aggregation. *Scandinavian Actuarial Journal*, **2016**(8), 713–740.
- Wang, B. and Wang, R. (2015). Extreme negative dependence and risk aggregation. *Journal of Multivariate Analysis*, **136**, 12–25.
- Wang, B. and Wang, R. (2016). Joint mixability. *Mathematics of Operations Research*, **41**(3), 808–826.
- Wang, R., Peng, L. and Yang, J. (2013). Bounds for the sum of dependent risks and worst Value-at-Risk with monotone marginal densities. *Finance and Stochastics*, **17**(2), 395–417.

References II

- Dunn, O. J. (1958). Estimation of the means for dependent variables. *Annals of Mathematical Statistics*. **29**(4), 1095–1111.
- Fisher, R. A. (1948). Combining independent tests of significance. *American Statistician*, **2**, 30.
- Hedges, L. V. and Olkin, I. (1985). *Statistical Methods for Meta-Analysis*. Orlando, FL: Academic Press.
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, **6**, 65–70.
- Mattner, L. (2012). Combining individually valid and arbitrarily dependent P-variables. In *Abstract Book of the Tenth German Probability and Statistics Days*, p. 104. Institut für Mathematik, Johannes Gutenberg-Universität Mainz.
- Meng, X.-L. (1993). Posterior predictive p-values. *Annals of Statistics*, **22**, 1142–1160.
- Rüschendorf, L. (1982). Random variables with maximum sums. *Advances in Applied Probability*, **14**(3), 623–632.

References III

Glenn Shafer, G., Alexander Shen, Nikolai Vereshchagin, and Vladimir Vovk.
Test martingales, Bayes factors, and p-values.
Statistical Science, 26:84–101, 2011.

Andrei N. Kolmogorov.
Three approaches to the quantitative definition of information.
Problems of Information Transmission, 1:1–7, 1965.

Andrei N. Kolmogorov.
Logical basis for information theory and probability theory.
IEEE Transactions on Information Theory, IT-14:662–664, 1968.

Analysis for the `sex differences` data

For the `sex differences` dataset, the combined p-values are (compared with 0.05 significance level; weighted by sample size)

- Bonferroni: 0.029 (significant)
- harmonic: 0.045 (weighted 0.041) (significant)
- geometric: 0.157 (weighted 0.198) (not significant)
- arithmetic: 0.613 (weighted 0.793) (not significant)

Analysis for the passive smoking data

For the passive smoking dataset (Hartung-Knapp-Sinha'08, Table 3.1, p.31, $K = 19$), the combined p-values are (compared with 0.05 significance level)

- Bonferroni: 0.051 (not significant)
- harmonic: 0.126 (not significant)
- geometric: 0.254 (not significant)
- arithmetic: 0.449 (not significant)

The integrable case: $r > -1$

Proposition 1

For $r \in (-1, \infty]$, $(r+1)^{1/r} M_{r,K}$, $K \in \{2, 3, \ldots\}$, is a family of merging functions and it is asymptotically precise.

$(1+r)^{1/r}|_{r=0} = e$.

The integrable case: $r > -1$

Proposition 1

For $r \in (-1, \infty]$, $(r+1)^{1/r} M_{r,K}$, $K \in \{2, 3, \dots\}$, is a family of merging functions and it is asymptotically precise.

Proof.

- $r > 0$, $q_\epsilon(\sum_{k=1}^K U_k^r) \ge \sum_{k=1}^K \mathrm{ES}_\epsilon^{\leftarrow}(U_k^r) = K \frac{1}{r+1} \epsilon^r$
- $r = 0$, $q_\epsilon(\sum_{k=1}^K \log U_k) \ge \sum_{k=1}^K \mathrm{ES}_\epsilon^{\leftarrow}(\log U_k) = K(\log \epsilon + 1)$
- $r < 0$, $q_{1-\epsilon}(\sum_{k=1}^K U_k^r) \le \sum_{k=1}^K \mathrm{ES}_{1-\epsilon}(U_k^r) = K \frac{1}{r+1} \epsilon^r$
- In all cases, $\underline{q}_\epsilon((r+1)^{1/r} M_{r,K}) \ge \epsilon$
- Use the VaR/ES asymptotic equivalence of Wang-W.'15. back

$(1+r)^{1/r}|_{r=0} = e$.

The non-integrable Pareto case: $r < -1$

No VaR/ES asymptotic equivalence for $r \leq -1$.

The non-integrable Pareto case: $r < -1$

No VaR/ES asymptotic equivalence for $r \leq -1$.

Proposition 2

For $r \in (-\infty, -1)$, $\frac{r}{r+1} K^{1+1/r} M_{r,K}$, $K \in \{2, 3, \dots\}$, is a family of merging functions and it is asymptotically precise.

The non-integrable Pareto case: $r < -1$

No VaR/ES asymptotic equivalence for $r \le -1$.

Proposition 2

For $r \in (-\infty, -1)$, $\frac{r}{r+1} K^{1+1/r} M_{r,K}$, $K \in \{2, 3, \dots\}$, is a family of merging functions and it is asymptotically precise.

Proof.

- To show $\frac{r}{r+1} K^{1+1/r} M_{r,K}$ is a merging function, directly apply the dual bound of Embrechts-Puccetti'06.
- To show the asymptotic precision, use the aggregation ratio of Bignozzi-Mao-Wang-W.'16 for super-heavy Pareto risks.

Letting $r \to -\infty$ one recovers the Bonferroni method: $KM_{-\infty,K}$.

Precise results for the Beta case: $r \geq 1/(K-1)$

Proposition 3

For $K \in \{2, 3, \ldots\}$ *and* $r \in (-1, \infty)$,

(i) $(r+1)^{1/r} M_{r,K}$ *is a precise merging function* $\Leftrightarrow$ $r \in [\frac{1}{K-1}, K-1]$.

(ii) *If* $r \geq K-1$, $K^{1/r} M_{r,K}$ *is a precise merging function.*

Precise results for the Beta case: $r \geq 1/(K-1)$

Proposition 3

For $K \in \{2, 3, \dots\}$ and $r \in (-1, \infty)$,

(i) $(r+1)^{1/r} M_{r,K}$ *is a precise merging function* $\Leftrightarrow$ $r \in [\frac{1}{K-1}, K-1]$.

(ii) *If* $r \geq K-1$, $K^{1/r} M_{r,K}$ *is a precise merging function.*

Proof.

- $r \geq 1$, U^r has a decreasing density
- $r \in [\frac{1}{K-1}, 1]$, U^r has an increasing density
- The $\overline{\mathrm{VaR}}_p$ and $\underline{\mathrm{VaR}}_p$ formulas of W.-Peng-Yang'13 give the precise value of $\underline{q}_\epsilon(M_{r,K})$

Precise results for the Beta case: $r \geq 1/(K-1)$

Examples.

- $\min(r+1, K)^{1/r} M_{r,K}$ is precise for $r \geq 1/(K-1)$.
- The arithmetic average times 2 is precise for $K \geq 2$
- The quadratic average times $\sqrt{3}$ is precise for $K \geq 3$
- Letting $r \to \infty$, the maximum $M_{\infty,K}$ is precise

Geometric averaging

Proposition 4

For each $K \in \{2, 3, \ldots\}$, $a_K M_{0,K}$ *is a precise merging function, where*

$$a_K = \frac{1}{c_K} \exp\left(-(K-1)(1-Kc_K)\right)$$

and c_K *is the unique solution to the equation*

$$\log(1/c - (K-1)) = K - K^2 c$$

over $c \in (0, 1/K)$. *Moreover,* $a_K \leq e$ *and* $a_K \to e$ *as* $K \to \infty$.

Geometric averaging

Proposition 4

For each $K \in \{2, 3, \ldots\}$, $a_K M_{0,K}$ is a precise merging function, where

$$a_K = \frac{1}{c_K} \exp\left(-(K-1)(1-Kc_K)\right)$$

and c_K is the unique solution to the equation

$$\log(1/c - (K-1)) = K - K^2 c$$

over $c \in (0, 1/K)$. Moreover, $a_K \leq e$ and $a_K \to e$ as $K \to \infty$.

Proof.

- Obtained from the $\overline{\mathrm{VaR}}_p$ formula of W.-Peng-Yang'13.

Geometric averaging

Table: Numeric values of a_K/e for the geometric mean

K	a_K/e	K	a_K/e	K	a_K/e
2	0.7357589	5	0.9925858	10	0.9999545
3	0.9286392	6	0.9974005	15	0.9999997
4	0.9779033	7	0.9990669	20	1.0000000

- In practice, use $a_K \approx e$ for $K \geq 5$
- $eM_{0,K}$ is always a merging function (noted by Mattner'12)

Harmonic averaging

Proposition 5

For $K > 2$, $(e \log K) M_{-1,K}$ is a merging function.

Harmonic averaging

Proposition 5

For $K > 2$, $(e \log K) M_{-1,K}$ is a merging function.

Proof.

- For a given $K > 2$, $e \log K = \min_{r<-1} \frac{r}{r+1} K^{1+1/r}$
- $(e \log K) M_{r,K}$ is a merging function for some $r < -1$
- $M_{-1,K} \geq M_{r,K}$ for $r < -1$

Harmonic averaging

Proposition 6

Set $a_K = \frac{(y_K+K)^2}{(y_K+1)K}$, $K > 2$, *where* y_K *is the unique solution to the equation*

$$y^2 = K((y+1)\log(y+1) - y), \qquad y \in (0,\infty).$$

Then $a_K M_{-1,K}$ *is a precise merging function. Moreover,* $a_K / \log K \to 1$ *as* $K \to \infty$.

Harmonic averaging

Proposition 6

Set $a_K = \frac{(y_K+K)^2}{(y_K+1)K}$, $K > 2$, *where* y_K *is the unique solution to the equation*

$$y^2 = K((y+1)\log(y+1) - y), \qquad y \in (0, \infty).$$

Then $a_K M_{-1,K}$ *is a precise merging function. Moreover,* $a_K / \log K \to 1$ *as* $K \to \infty$.

Proof.

- Again obtained from the $\overline{\mathrm{VaR}}_p$ formula of W.-Peng-Yang'13.

Harmonic averaging

Table: Numeric values of $a_K / \log K$ for the harmonic mean

K	$a_K / \log K$	K	$a_K / \log K$	K	$a_K / \log K$
3	2.499192	10	1.980287	100	1.619631
4	2.321831	20	1.828861	200	1.561359
5	2.214749	50	1.693497	400	1.514096

- The rate of convergence $a_K / \log K \to 1$ is very slow

Harmonic averaging

Table: Numeric values of $a_K / \log K$ for the harmonic mean

K	$a_K / \log K$	K	$a_K / \log K$	K	$a_K / \log K$
3	2.499192	10	1.980287	100	1.619631
4	2.321831	20	1.828861	200	1.561359
5	2.214749	50	1.693497	400	1.514096

- The rate of convergence $a_K / \log K \to 1$ is very slow
- Suggestions:
  - for $K \geq 3$, use $(2.5 \log K) M_{-1,K}$
  - for $K \geq 10$, use $(2 \log K) M_{-1,K}$
  - for $K \geq 50$, use $(1.7 \log K) M_{-1,K}$

General formulas

Proposition 7

For $K > 2$ *and* $r \in (-\infty, \frac{1}{K-1}) \setminus \{-1, 0\}$, *set*

$$b_{r,K} := \left(\frac{K}{(K-1)(1-(K-1)c^*)^r + c^{*r}} \right)^{1/r},$$

where c^* *is the unique solution* $c \in (0, 1/K)$ *to the equation*

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

Then $b_{r,K} M_{r,K}$ *is a precise p-merging function.*

back

Efficiency: iid

Assume an iid setting under the true nature (different from H_0):

- $p_1, \ldots, p_K$ are iid Q-distributed
- Let

$$\Pi = \Pi(Q) = \sup \left\{ m \in [0, \infty) : \int p^{-m} Q(\mathrm{d}p) < \infty \right\}$$

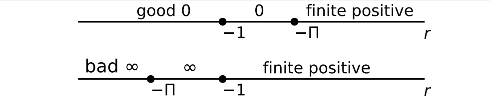
- $\Pi(\mathrm{U}[0,1]) = 1$ (under H_0)
- Write $P_{r,K} = a_{r,K} M_{r,K}$
- Consider $K \to \infty$

Note: we are not interested in the iid case

back

Efficiency: iid

Some results:



The combined p-value for different r in the cases $\Pi < 1$ (top) and $\Pi > 1$ (bottom).

- If $\Pi < 1$, then $r \in [-\infty, -1]$ has the best rate of convergence to zero $P_{r,K} \approx cK^{1-1/\Pi}$
- If $\Pi > 1$, then $r \in [-\infty, -\Pi]$ has the worst rate of convergence to infinity $P_{r,K} \approx cK^{1-1/\Pi}$
- Usually $\Pi \leq 1$ which indicates some power

Efficiency: dependence

Suppose that $p_1, \ldots, p_K$ comes from an exchangeable distribution.

- By de Finetti's Theorem, there is some latent random variable Z, and $p_1, \ldots, p_K$ are iid conditional on Z
- Let Q_z be the conditional distribution of p_1 given $Z = z$
- The power of the merging methods depends on $\Pi(Q_z)$
- It may happen that $\Pi(Q) \le 1$ but $\Pi(Q_z) > 1$ for all z (e.g. identical p-values)

Efficiency: dependence

Dependent one-sided z-tests

- $X_1, \dots, X_K$ are jointly normal, $X_k \sim \mathrm{N}(-\mu, 1)$ where $\mu \geq 0$ and $\mathrm{Cov}(X_k, X_j) = \rho \in [0, 1]$ for $k \neq j$
- H_0 is $\mu = 0$
- p-values are $p_k = \Phi(X_k)$ where Φ is the standard normal cdf
- $\rho = 0$ means iid tests; $\rho = 1$ means identical tests
- $\Pi(Q) = 1$ and $\Pi(Q_z) = 1/(1 - \rho^2) \geq 1$ for all z

Efficiency: dependence

Proposition 8

Assume $\rho > 0$ and $\mu > 0$ in the above model. As $K \to \infty$, if $r \leq -1$, then $P_{r,K} \to \infty$; if $r > -1$, then $\mathbb{E}[P_{r,K}] \to A(r, \mu, \rho)$ which is

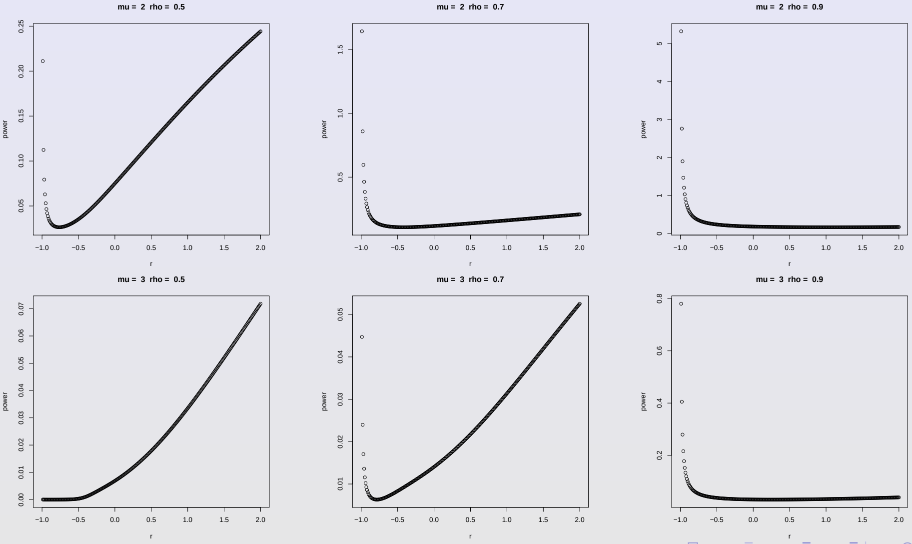
$$(1+r)^{1/r}\mathbb{E}\left[\left(\mathbb{E}\left[\left(\Phi\left(\sqrt{1-\rho^2}W + \rho Z - \mu\right)\right)^r |Z\right]\right)^{1/r}\right] < \infty,$$

where Z and W are iid standard normal random variables.

Remark.

- For $r < -1$, $P_{r,K} \approx cK^{\rho^2}$ (grows very slow)

Simulation: $A(r, \mu, \rho)$



Numerical results

	$\rho = 0.3$	$\rho = 0.5$	$\rho = 0.7$	$\rho = 0.9$	$\rho = 0.95$
$\mu = 1$	-0.880	-0.559	0.314	≥ 2	≥ 2
$\mu = 2$	-0.849	-0.769	-0.418	1.037	≥ 2
$\mu = 3$	-0.880	-0.910	-0.789	0.244	1.207
$\mu = 4$	-0.890	-0.870	-0.779	-0.077	0.555
$\mu = 5$	-0.900	-0.880	-0.839	-0.478	0.064

Table: r^* which minimizes $A(r, \mu, \rho)$ for different values of μ, ρ. Red choices lead to insignificant p-values for $\alpha = 0.05$

back