Robust Risk Aggregation and Merging P-values

Ruodu Wang

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

Department of Statistics and Actuarial Science

University of Waterloo

4th European Actuarial Journal Conference

KU Leuven, Belgium September 11, 2018

Agenda

Fundamental problem in Finance/Insurance

Basic setup.

- A vector of risk factors: $\mathbf{X} = (X_1, \ldots, X_d)$
- 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)
- $\Psi(\mathbf{X}) = \sum_{i=1}^d X_i$

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

Fréchet problem

Model assumption: $X_i \sim F_i$, F_i known, $i = 1, \dots, d$.

Let $$\mathcal{S}_d = \mathcal{S}_d(F_1, \dots, F_d) = \left\{ \sum_{i=1}^d X_i : X_i \sim F_i,\ i = 1, \dots, d \right\}.$$

- Every element in $\mathcal{S}_d$ is a possible risk position
- Determination (distributions-wise) of $\mathcal{S}_d$: very challenging.
- Example:
  - $F = \mathrm{U}[0,1]$
  - Characterization of $\mathcal{S}_2(F, F)$ is an open question
  - $\mathcal{S}_3(F, F, F)$ characterized analytically recently

$$S \in \mathcal{S}_3(F, F, F) \ \Leftrightarrow \ S \le_{\mathrm{cx}} \mathrm{U}[0,3]$$

VaR and ES

Two regulatory risk measures in Basel III & IV and Solvency II

Value-at Risk (VaR)

For $p \in (0,1]$, $X \in L^0$,

$$\mathrm{VaR}_p(X) = F_X^{-1}(p) = \inf\{x \in \mathbb{R} : F_X(x) \geq p\}.$$

Expected Shortfall (ES, or TVaR, CVaR, CTE, AVaR, ...)

For $p \in (0,1)$, $X \in L^1$,

$$\mathrm{ES}_p(X) = \frac{1}{1-p}\int_p^1 \mathrm{VaR}_q(X)\mathrm{d}q \underset{(F \text{ cont.})}{=} \mathbb{E}\left[X | X > \mathrm{VaR}_p(X)\right].$$

Worst- and best-values of VaR and ES

The Fréchet problems

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

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

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

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

Basel III & IV ES calculation

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

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

where

- X_i is the total random loss from a risk class, $i = 1, \ldots, d$
  - 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

Robust risk aggregation

Calculation of $\overline{\mathrm{VaR}}_p(\mathcal{S}_d)$, $\underline{\mathrm{VaR}}_p(\mathcal{S}_d)$, and $\underline{\mathrm{ES}}_p(\mathcal{S}_d)$

- Dependence modeling, (co/counter)-monotonicity, ...
  - Dhaene, Denuit, Goovaert, Vanduffel, ...
- Risk measures
  - Artzner-Delbaen-Eber-Heath, Wang, ...
- Mass transportation
  - Rüschendorf, Embrechts, Puccetti, ...
- Joint mixability
- Computation, neural network methods, ...
- Uncertainty, ambiguity, decision making, ...

Some results in robust VaR aggregation

Some results

- Embrechts-Puccetti'06, F&S
- W.-Peng-Yang'13, F&S
- Bernard-Jiang-W.'14, IME
- Wang-W.'15, JMVA
- Embrechts-Wang-W.'15 F&S
- Jakobsons-Han-W.'16, SAJ
- Bignozzi-Mao-Wang-W.'16, Extremes
- Wang-W.'16, MOR

This talk - work with Vladimir Vovk (Royal Holloway, London)

- Vovk-W.'18

Simple VaR-ES relation

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

$$\mathrm{ES}_p^{\leftarrow}(\cdot) = \frac{1}{p}\int_0^p \mathrm{VaR}_q(\cdot)\mathrm{d}q.$$

- By definition

$$\mathrm{ES}_p^{\leftarrow} \le \mathrm{VaR}_p \le \mathrm{ES}_p.$$

- By subaddivity of ES_p and superadditivity of $\mathrm{ES}_p^{\leftarrow}$,

$$\sum_{i=1}^d \mathrm{ES}_p^{\leftarrow}(X_i) \le \mathrm{VaR}_p\left(\sum_{i=1}^d X_i\right) \le \sum_{i=1}^d \mathrm{ES}_p(X_i).$$

- Asymptotically sharp as $d \to \infty$ (Embrechts-Wang-W'15).

Dual bound of Embrechts-Puccetti'06

Assume a homogeneous model $F_1 = \cdots = F_d = F$.

Derived from Theorem 4.2 of Embrechts-Puccetti'06

For a continuous cdf F and $X_1, \ldots, X_d \sim F$,

$$\mathbb{P}(X_1 + \cdots + X_d \geq ds) \leq \inf_{r \in [0,s)} \frac{\int_r^{ds-(d-1)r} (1 - F(x)) \mathrm{d}x}{s - r}.$$

- Based on Rüschendorf'82
- probability bounds $\Leftrightarrow$ VaR bounds

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

Derived from Theorem 3.4 of W.-Peng-Yang'13

Suppose that the density of F is decreasing on its support. For $p \in (0,1)$ and $X \sim F$,

$$\overline{\mathrm{VaR}}_p(\mathcal{S}_d) = d\mathbb{E}[X | F^{-1}(p+(d-1)c) \le X \le F^{-1}(1-c)],$$

where c is the smallest number in $[0, \frac{1}{d}(1-p)]$ such that

$$\int_{p+(d-1)c}^{1-c} F^{-1}(t)\mathrm{d}t \ge \frac{1-p-dc}{d}((d-1)F^{-1}(p+(d-1)c) + F^{-1}(1-c)).$$

- only decreasing density in the tail part is required
- dual form in Puccetti-Rüschendorf'13

$\underline{\mathrm{VaR}}_p$ formula of W.-Peng-Yang'13

Derived from Theorem 3.4 of W.-Peng-Yang'13

Suppose that the density of F is decreasing on its support. For $p \in (0,1)$ and $X \sim F$,

$$\underline{\mathrm{VaR}}_p(\mathcal{S}_d) = \max\{(d-1)F^{-1}(0) + F^{-1}(p), d\mathrm{ES}_p^{\leftarrow}(X)\}.$$

- An asymmetry between upper and lower bounds ...

Asymptotic equivalence of Wang-W.'15

Corollary 3.7 of Wang-W.'15

For $p \in (0,1)$ and $X \sim F$,

$$\lim_{d\to\infty} \frac{1}{d}\overline{\mathrm{VaR}}_p(\mathcal{S}_d) = \mathrm{ES}_p(X).$$

- $\overline{\mathrm{VaR}}_p(\mathcal{S}_d)/\overline{\mathrm{ES}}_p(\mathcal{S}_d) \to 1$.
- First results obtained by Puccetti-Rüschendorf'14
- Inhomogeneous portfolios: Embrechts-Wang-W.'15
- Other risk measures, rank dependent utilities: W.-Bignozzi-Tsanakas'15, Cai-Liu-W.'18, W.-Xu-Zhou'18

Aggregation ratio of Bignozzi-Mao-Wang-W.'16

Derived from Proposition 3.5 of Bignozzi-Mao-Wang-W.'16

For a Pareto distribution F with index $\beta \in (0,1)$,

$$\lim_{d\to\infty} \frac{\overline{\mathrm{VaR}}_p(\mathcal{S}_d)}{d^{1/\beta}F^{-1}(p)} = \left(\frac{1}{1-\beta}\right)^{\frac{1}{\beta}}.$$

- If $\beta \geq 1$, asymptotic equivalence implies

$$\lim_{d\to\infty} \frac{\overline{\mathrm{VaR}}_p(\mathcal{S}_d)}{dF^{-1}(p)} = \frac{\mathrm{ES}_p(X)}{\mathrm{VaR}_p(X)} \approx \frac{\beta}{\beta-1}.$$

- For MRV models (e.g. Embrechts-Lambrigger-Wüthrich'09),

$$\lim_{p\uparrow 1} \frac{\mathrm{VaR}_p(X_1+\cdots+X_d)}{d^{1/\beta}F^{-1}(p)} \leq 1.$$

$\underline{\mathrm{VaR}}_p$ formula of Jakobsons-Han-W.'16

Assume inhomogeneous models

Derived from Corollary 4.7 of Jakobsons-Han-W.'16

Suppose that the density of each $F_1, \dots, F_d$ is decreasing on its support. For $p \in (0,1)$,

$$\underline{\mathrm{VaR}}_p(\mathcal{S}_d) = \max\left\{ \max_{i=1,\dots,d} \left\{ F_i^{-1}(p) + \sum_{j\neq i} F_j^{-1}(0) \right\}, \mu_p \right\},$$

where $\mu_p = \sum_{i=1}^d \mathrm{ES}_p^{\leftarrow}(X_i)$, $X_i \sim F_i$, $i = 1, \dots, d$.

- Formula for $\overline{\mathrm{VaR}}_p(\mathcal{S}_d)$ involves solving an implicit ODE

Summary on VaR aggregation

$d = 2$

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

$d \geq 3$

- solved analytically for decreasing densities
  - also solved for increasing densities by putting a negative sign
  - relies on joint mixability for monotone densities (Wang-W'16)
- generalization to other distributions is limited
- homogeneous relaxation in Bernard-Jiang-W.'14

Summary on VaR aggregation

Remarks.

- Efficient numerical algorithm: the Rearrangement Algorithm
  - Puccetti-Rüschendorf'12, Embrechts-Puccetti-Rüschendorf'13, Hofert-Memartoluie-Saunders-Wirjanto'17, Bernard-Bondarenko-Vanduffel'18, ...
- Risk aggregation with partial dependence information
  - Bernard-Vanduffel'15, Puccetti-Rüschendorf-Manko'16, Lux-Papapantoleon'17, Bernard-Rüschendorf-Vanduffel'17, Bernard-Rüschendorf-Vanduffel-W.'17, ...
- Risk aggregation with marginal and dependence uncertainty
  - Li-Shao-W.-Yang'18, Shao'18, ...
- Connection to distributionally robust optimization

Robust VaR aggregation

Some disadvantages:

- often too high risk values for practical use
- analytical results require monotone density assumptions
- assumed no dependence information
- assumed full marginal information

Are there real-world problems with these features?

P-values

STAT 101

A (true) p-value p for testing a hypothesis H_0:

- p is uniform on $[0, 1]$ under $H_0 \Leftrightarrow \mathbb{P}^{H_0}(p \le \epsilon) = \epsilon$ for $\epsilon \in [0, 1]$
- Specify a significant level α, typically $\alpha = 0.05,\ 0.01,\ 0.005$ ...
- Rejects H_0 if (realized) $p \le \alpha$
  - cannot reject H_0 if $p > \alpha$
- Probability of type I error $= \mathbb{P}^{H_0}(\text{reject } H_0) = \alpha$

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

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

- Fisher'48: assumes $p_1, \dots, p_K$ are independent

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:

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

The Bonferroni method

The Bonferroni method

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

Particular interest: the case of heavily dependent tests.

- is it more natural to look at some average of p-values?

Merging functions

In any atomless probability space $(\Omega, \mathcal{A}, \mathbb{P})$ ...

Definition 1

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

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

(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-value for all p-values $P_1, \dots, P_K$.

- 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;
- $F(U_1, \dots, U_K)$ is a p-value for all $U_1, \dots, U_K \in \mathcal{U}$;
- for all $\epsilon \in (0,1)$, $\overline{\mathbb{P}}(F \le \epsilon) \le \epsilon$, where

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

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

Precise merging functions

Definition 2

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

- A precise merging function cannot be improved

Examples.

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

Precise merging functions

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

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

$$\le \sum_{i=1}^{K} \mathbb{P}(Kp_i \le \epsilon)$$

$$= \sum_{i=1}^{K} \frac{\epsilon}{K} = \epsilon.$$

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

Asymptotically precise merging functions

Definition 3

A family of merging functions F_K on $[0,1]^K$, $K = 2, 3, \ldots$, is called asymptotically precise if, for any $a \in (0,1)$, aF_K is not a merging function for a large enough K.

- This family of merging functions cannot be improved by a constant multiplier

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

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

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

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

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

- $M_{r,K}$ increases in r
  - The constants $a_{r,K}$ 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

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

Translation to a risk aggregation problem

Lemma 1

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

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

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

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

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

The same conclusion holds if all $\le$ and $\ge$ are replaced by $=$.

- In statistical practice one only needs to have $\overline{\mathbb{P}}(F \le \epsilon) \le \epsilon$ 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

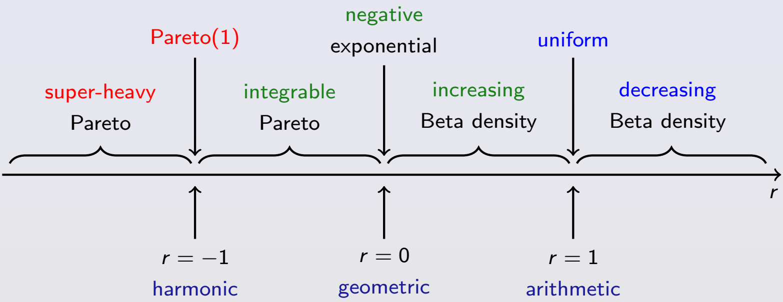
$$\ln(\underline{q}_\epsilon(M_{r,K})) = \inf_{U_1,\dots,U_K \in \mathcal{U}} \left\{ q_\epsilon \left(\frac{1}{K} (\ln U_1 + \cdots + \ln U_K) \right) \right\}$$

(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 $\ln U$) for $r \in \mathbb{R}$



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.

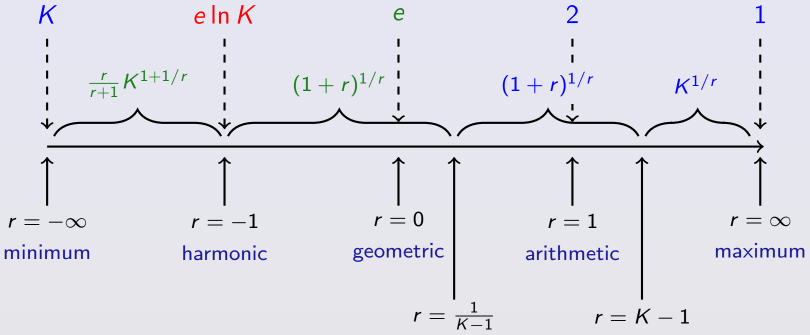
Proof.

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

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

Main results summary

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

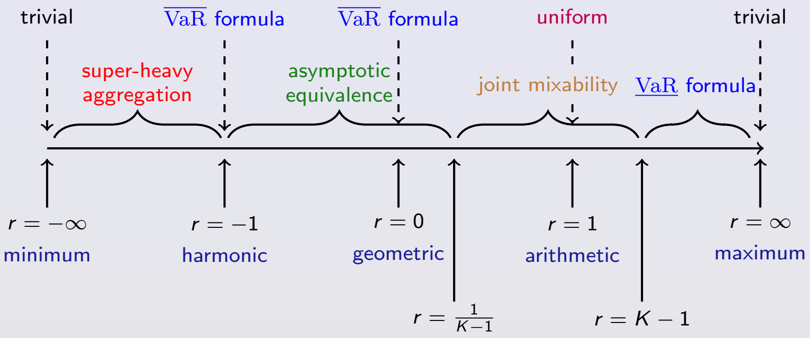


blue: precise; green: asymptotically precise; red: rough

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 2

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

Proof.

- $(r+1)^{1/r} M_{r,\mathbf{w}}$ is precise $\Leftrightarrow$ $\mathrm{VaR} = \mathrm{ES}$ $\Leftrightarrow$ joint mixability
- For $r \in (0, 1)$, use the characterization of joint mixability for monotone densities of Wang-W.'16.
- For $r \ge 1$, use the $\underline{\mathrm{VaR}}_p$ formula of Jakobsons-Han-W.'16.

Weighted averaging

The special case $r = 1$ (weighted arithmetic mean):

- $w \le 1/2$
  - no single experiment outweighs the total of all the others
  - the optimal multiplier is 2 (same as uniform weights)
- $w > 1/2$
  - a single experiment outweighs the total of all the others
  - assuming $w_1 = \max(\mathbf{w})$,

$$\frac{1}{w} M_{1,\mathbf{w}}(p_1, \dots, p_K) = p_1 + \sum_{k=2}^{K} \frac{w_k}{w} p_k$$

  - weighted adjustments are added to the p-value obtained from the most important experiment

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 $\mathbf{w} \in \Delta_K$.

- A deeper conjecture: $\mathcal{S}_d(F_1, \dots, F_d) \subset \mathcal{S}_d(\hat{F}, \dots, \hat{F})$ where $\hat{F}^{-1} = \frac{1}{d}\sum_{i=1}^d F_i^{-1}$.
- Bernard-Jiang-W.'14: $\mathcal{S}_d(F_1, \dots, F_d) \subset \mathcal{S}_d(\bar{F}, \dots, \bar{F})$ where $\bar{F} = \frac{1}{d}\sum_{i=1}^d F_i$.

Simulation

A simulation study for the performance of $a_{r,K} M_{r,K}$

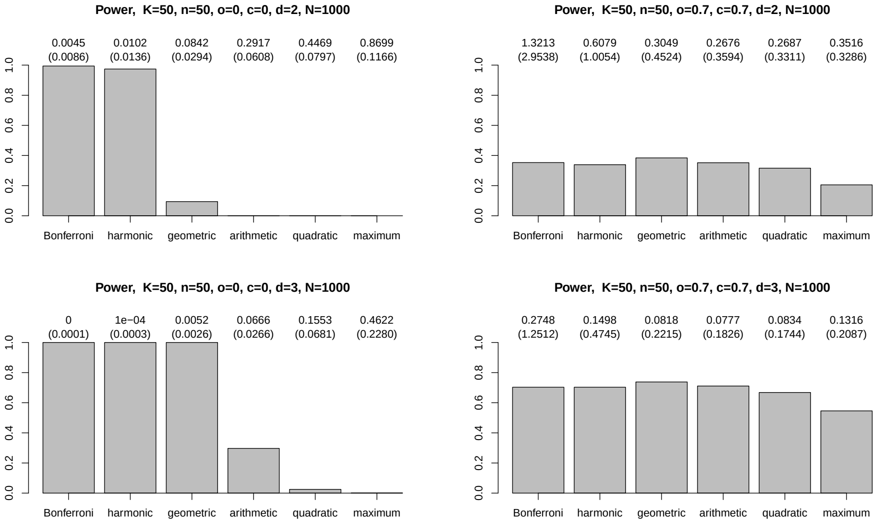
- 6 merging functions: $r \in \{-\infty, -1, 0, 1, 2, \infty\}$
  - Bonferroni (minimum), harmonic, geometric, arithmetic, quadratic, maximum
- $a_{-1,K} = 1.83 \ln K$ for $K = 20$ and $a_{-1,K} = 1.7 \ln K$ for $K \geq 50$
- $a_{0,K} = e$
- others are precise

Simulation

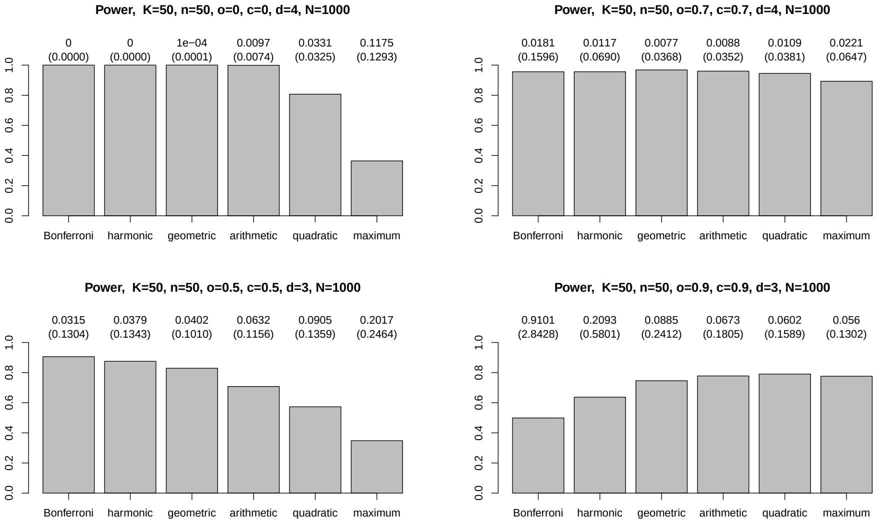
Simulation setup

- K tests for whether a given set of data has zero mean
  - n iid normal observations with mean $d/\sqrt{n}$ and known variance 1
  - two-sided z-tests
- o = %-overlap of observations between any two tests
- c = correlation of non-overlapping observations between tests
- $\alpha = 0.05$, significance level
- the procedure is repeated N times

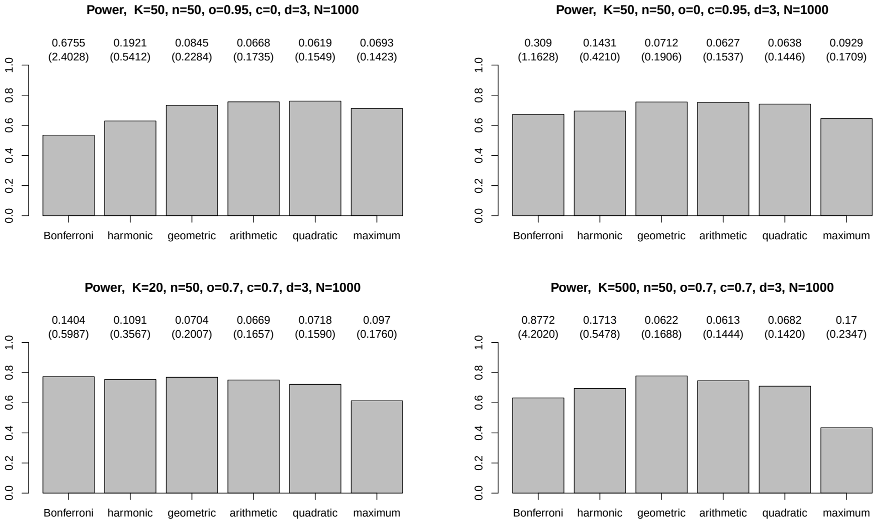
Simulation


Power, K=50, n=50, o=0, c=0, d=2, N=1000
0.0045 (0.0086)
0.0102 (0.0136)
0.0842 (0.0294)
0.2917 (0.0608)
0.4469 (0.0797)
0.8699 (0.1166)
0.0 0.2 0.4 0.6 0.8 1.0
Bonferroni harmonic geometric arithmetic quadratic maximum
Power, K=50, n=50, o=0.7, c=0.7, d=2, N=1000
1.3213 (2.9538)
0.6079 (1.0054)
0.3049 (0.4524)
0.2676 (0.3594)
0.2687 (0.3311)
0.3516 (0.3286)
0.0 0.2 0.4 0.6 0.8 1.0
Bonferroni harmonic geometric arithmetic quadratic maximum
Power, K=50, n=50, o=0, c=0, d=3, N=1000
0 (0.0001)
1e−04 (0.0003)
0.0052 (0.0026)
0.0666 (0.0266)
0.1553 (0.0681)
0.4622 (0.2280)
0.0 0.2 0.4 0.6 0.8 1.0
Bonferroni harmonic geometric arithmetic quadratic maximum
Power, K=50, n=50, o=0.7, c=0.7, d=3, N=1000
0.2748 (1.2512)
0.1498 (0.4745)
0.0818 (0.2215)
0.0777 (0.1826)
0.0834 (0.1744)
0.1316 (0.2087)
0.0 0.2 0.4 0.6 0.8 1.0
Bonferroni harmonic geometric arithmetic quadratic maximum

Simulation


Power, K=50, n=50, o=0, c=0, d=4, N=1000
0 (0.0000)
0 (0.0000)
1e−04 (0.0001)
0.0097 (0.0074)
0.0331 (0.0325)
0.1175 (0.1293)
0.0 0.2 0.4 0.6 0.8 1.0
Bonferroni harmonic geometric arithmetic quadratic maximum
Power, K=50, n=50, o=0.7, c=0.7, d=4, N=1000
0.0181 (0.1596)
0.0117 (0.0690)
0.0077 (0.0368)
0.0088 (0.0352)
0.0109 (0.0381)
0.0221 (0.0647)
0.0 0.2 0.4 0.6 0.8 1.0
Bonferroni harmonic geometric arithmetic quadratic maximum
Power, K=50, n=50, o=0.5, c=0.5, d=3, N=1000
0.0315 (0.1304)
0.0379 (0.1343)
0.0402 (0.1010)
0.0632 (0.1156)
0.0905 (0.1359)
0.2017 (0.2464)
0.0 0.2 0.4 0.6 0.8 1.0
Bonferroni harmonic geometric arithmetic quadratic maximum
Power, K=50, n=50, o=0.9, c=0.9, d=3, N=1000
0.9101 (2.8428)
0.2093 (0.5801)
0.0885 (0.2412)
0.0673 (0.1805)
0.0602 (0.1589)
0.056 (0.1302)
0.0 0.2 0.4 0.6 0.8 1.0
Bonferroni harmonic geometric arithmetic quadratic maximum

Simulation


Power, K=50, n=50, o=0.95, c=0, d=3, N=1000
0.6755 (2.4028)
0.1921 (0.5412)
0.0845 (0.2284)
0.0668 (0.1735)
0.0619 (0.1549)
0.0693 (0.1423)
0.0 0.2 0.4 0.6 0.8 1.0
Bonferroni harmonic geometric arithmetic quadratic maximum
Power, K=50, n=50, o=0, c=0.95, d=3, N=1000
0.309 (1.1628)
0.1431 (0.4210)
0.0712 (0.1906)
0.0627 (0.1537)
0.0638 (0.1446)
0.0929 (0.1709)
0.0 0.2 0.4 0.6 0.8 1.0
Bonferroni harmonic geometric arithmetic quadratic maximum
Power, K=20, n=50, o=0.7, c=0.7, d=3, N=1000
0.1404 (0.5987)
0.1091 (0.3567)
0.0704 (0.2007)
0.0669 (0.1657)
0.0718 (0.1590)
0.097 (0.1760)
0.0 0.2 0.4 0.6 0.8 1.0
Bonferroni harmonic geometric arithmetic quadratic maximum
Power, K=500, n=50, o=0.7, c=0.7, d=3, N=1000
0.8772 (4.2020)
0.1713 (0.5478)
0.0622 (0.1688)
0.0613 (0.1444)
0.0682 (0.1420)
0.17 (0.2347)
0.0 0.2 0.4 0.6 0.8 1.0
Bonferroni harmonic geometric arithmetic quadratic maximum

Simulation

Observations.

- The Bonferroni method
  - performs very well in cases of no or moderate correlation
  - often has a large standard deviation
  - is able to utilize information from independent tests
- The arithmetic method
  - performs very well in cases of heavy overlap or correlation
  - has a relatively stable average p-value
- The geometric method performs quite nicely in most cases
- The harmonic method is between Bonferroni and geometric
- The maximum and quadratic methods are not remarkable

Concluding remarks

Further directions

- Power analysis for classic statistical models
- Adaptive selection of the merging function
- Relation between dependence and the choice of merging functions

Open questions on risk aggregation

Mathematical questions on robust risk aggregation:

- Characterization of $\mathcal{S}_d$ 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 questions:

- For a given F, determine whether $F \in \mathcal{S}(\mathrm{U}[0,1], \mathrm{U}[0,1])$?
- For given correlation matrix Σ and $F_1, \dots, F_d$, is the set

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

  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 $S = X_1 + \cdots + X_d$.

Thank you

Thank you for your kind attention

The manuscript is available at

- SSRN: https://ssrn.com/abstract=3166304
- arXiv: https://arxiv.org/abs/1212.4966

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.

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

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

Proposition 3

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: $KM_{-\infty,K}$.

back

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

Proposition 4

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 5

For each $K \in \{2,3,\dots\}$, $a_K M_{0,K}$ 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

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

over $c \in (0,1/K)$. Moreover, $a_K \le e$ 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 6

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

Proof.

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

Harmonic averaging

Proposition 7

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

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

Then $a_K M_{-1,K}$ *is a precise merging function. Moreover,* $a_K / \ln K \to 1$ *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 $a_K / \ln K$ for the harmonic mean

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

back