Dependence in Stochastic Modeling: Financial Crisis, Strategies, Equilibria, Decisions, Transport, and Statistics

Ruodu Wang

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

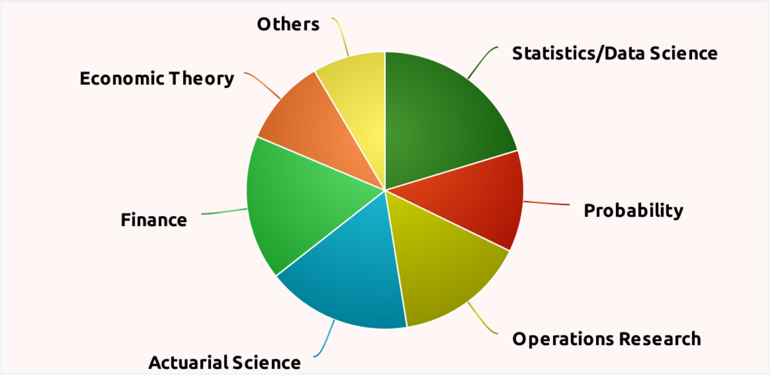
Department of Statistics and Actuarial Science
University of Waterloo

Alibaba Mathematics Colloquium, September 1, 2021

Agenda

Background

Quantitative Risk Management



Background

The University of Waterloo

- A young tech university
- One of the largest research groups in Actuarial Science/Quantitative Finance/Risk Management in the world with ≈ 20 professors
- No.1 in Actuarial Science Research worldwide by UNL ranking
- Largest Mathematics Faculty, > 8000 students, > 240 professors

For this talk, I assume

- Basic college probability theory
- Basic college statistics
- Good understanding in mathematics

A general setup

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

Assumptions

marginals may be known; dependence is unknown/arbitrary

Questions:

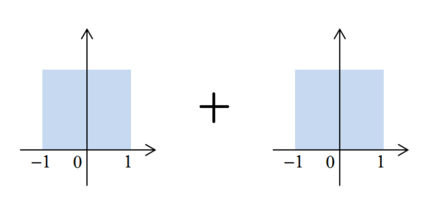
- properties of $\Psi(\mathbf{X})$ for some $\Psi : \mathbb{R}^n \to \mathbb{R}^d$
- range of $\mathbb{P}(\mathbf{X} \in A)$ for some $A \subseteq \mathbb{R}^n$
- "optimal" dependence structures of $\mathbf{X}$
- statistical decisions based on $\mathbf{X}$

Dates back to Fréchet-Hoeffding; has roots in Monge-Kantorovich

- Data scarcity; uncertainty; optimization variable; absent information; lack of models; equilibrium output

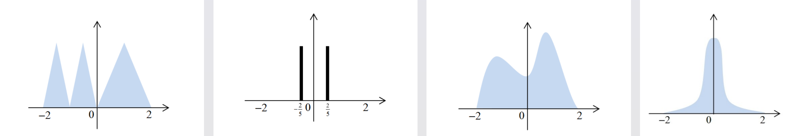
An innocent question

What is a possible distribution of $S = X_1 + X_2$ for uniformly distributed X_1 and X_2?



Obvious constraints

- $\mathbb{E}[S] = 0$
- range of $S \subseteq [-2, 2]$
- $\mathrm{Var}(S) \leq 4/3$



Financial crisis

The 2007 - 09 financial crisis:

- the worst one since the Great Depression of the 1930s
- once in 50 years event
- subprime mortgage bubble
- Key ingredients
  - a housing market at peak (2006)
  - structured financial products and derivatives
    - collateralized debt obligations (CDO)
    - credit default swaps (CDS)
  - advanced mathematical models
  - political shortsightedness and the slow reaction of regulators

CDO

A CDO repackages the cash flows from a set of assets

- Pooling the return from a set of assets (e.g. loans)
- Claims are tranched: differing priorities
- Creates new securities, of which some are less risky than the original assets, and others are riskier.

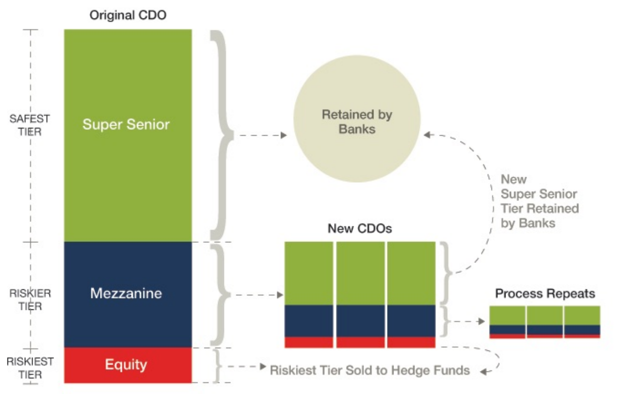
"The engine that powered the mortgage supply chain" for nonprime mortgages

- Sales of CDOs grew from $69B in 2000 to around $500B in 2006
- Between 2003 and 2007, Wall Street issued almost $700B in CDOs that included mortgage-backed securities as collateral

CDO: An example

- $X_i \geq 0$ is the random loss from a defaultable, speculative-grade bond i, $i = 1, \ldots, n$
- $X_1, \ldots, X_n$ standalone are not very attractive to investors
- The idea of CDO
  - Pool $X_1, \ldots, X_n$: let $L = \sum_{i=1}^{n} X_i$ and take some constants $K_1 < K_2$
  - Design financial products with payments Y_1, Y_2, Y_3 so that
    - $Y_1 = (L - K_2)_+$
    - $Y_2 = \min\{(L - K_1)_+, K_2 - K_1\}$
    - $Y_3 = \min\{L, K_1\}$
  - $Y_1 + Y_2 + Y_3 = L$
  - $\mathbb{P}(Y_1 > 0) = \mathbb{P}(L > K_2)$ can be very small

CDO: An example



- The one-year loss probability of senior (AAA-rated) tranches is less than 1/10,000
- Some investors are happy to hold a speculative grade bond, while others seek safer bonds.

Dependence modeling and risk aggregation

The rating for CDO tranches involves calculating $\mathbb{P}(L > K)$, where

- $L = \sum_{i=1}^{n} X_i$, and X_i is the loss from a loan
- K is a constant and $K \gg \mathbb{E}[L]$
- n is large, and each X_i has a small probability of loss (default), i.e. $\mathbb{P}(X_i = 0) = 1 - \epsilon_i$ and ϵ_i is small
- ϵ_i is the default probability of loan i and it is decisive in the calculation of the interest rate or price for this loan
- ϵ_i is modelled "relatively well" using individual credit characteristics

Dependence modeling and risk aggregation

- How $X_1, \ldots, X_n$ are dependent is unknown and they are almost "uncorrelated" because they were diversified by region
- If $X_1, \ldots, X_n$ are almost independent, then the central limit theorem can be applied, and $\mathbb{P}(L > K)$ can be approximated
- The dependence structure of $(X_1, \ldots, X_n)$ matters:
  - Assume $\mathbb{P}(X_i = 1) = 0.1$, $\mathbb{P}(X_i = 0) = 0.9$, $n = 1000$, $K = 200$
  - If $X_1, \ldots, X_n$ are iid, then $\mathbb{P}(L > K) < 10^{-20}$
  - If $X_1, \ldots, X_n$ are positively dependent, then $\mathbb{P}(L > K) \approx 0.1$
  - $\sup\{\mathbb{P}(L > K) : \text{all dependence structures}\} = ?$

Financial crisis

- Classic statistics fails to apply here: no data are available for the scenario "house prices started to fall"
- The past data (the scenario "house prices are good") suggests that $X_1, \dots, X_n$ are mildly correlated or almost independent
- Substantial miscalculation of $\mathbb{P}(L > K)$ leads to unjustified high rating of CDO products $\Rightarrow$ huge model risk
- In 2007, the mortgage backed securities turned out to be highly correlated
- CDOs made up over half ($542 billion) of the nearly trillion dollars in losses suffered by financial institutions from 2007 to early 2009

Risk assessment under uncertainty

Abstract setup.

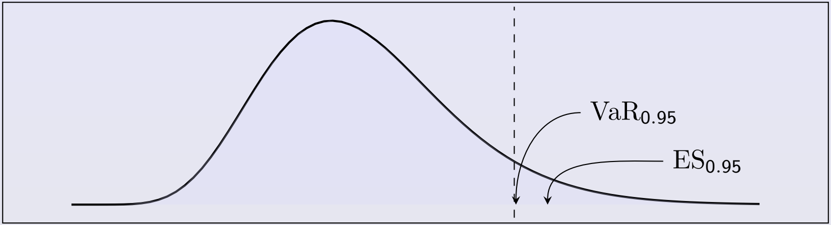
- A vector of risk factors: $\mathbf{X} = (X_1, \dots, X_n)$
- A financial position $\Psi(\mathbf{X})$
- A mapping $\rho : \mathbf{X} \to \mathbb{R}$ (a measure of risk)

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

Most practical choices:

- $\Psi(\mathbf{X}) = \sum_{i=1}^{n} X_i$
- $\rho(X) = \mathbb{P}(X > t)$, $\rho = \mathrm{VaR}_p$ or $\rho = \mathrm{ES}_p$

Risk assessment under uncertainty



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/CTE)

Risk assessment under uncertainty

- Because ES is subadditive, with unknown dependence

$$\mathrm{ES}_p\left(\sum_{i=1}^n X_n\right) \le \sum_{i=1}^n \mathrm{ES}_p(X_i)$$

- Marginal information provides bounds on the portfolio
- Worst-case ES: $\overline{\mathrm{ES}}_p = \sum_{i=1}^n \mathrm{ES}_p(X_i)$
- VaR: not subadditive
- Worst-case VaR: generally an open question for $n \geq 3$
- Similarly: bounds on $\mathbb{P}(\sum_{i=1}^n X_i > t)$

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

Equilibria

- n agents each with a preference
- Competitive equilibrium
  - Each agent i chooses a decision X_i according to some optimization of his/her own preference and constraints
  - Equilibrium: A random vector $(X_1, \ldots, X_n)$ such that no agent would be able to change positions to improve
- Cooperative (Pareto) equilibrium
  - A central planner chooses $(X_1, \ldots, X_n)$
  - Equilibrium: A random vector $(X_1, \ldots, X_n)$ that cannot be strictly improved

Welfare theorems

Under some conditions, competitive equilibrium $\iff$ cooperative equilibrium

Positive and negative dependence

Random variables $X_1, \ldots, X_n$

- Positive dependence
  - Random variables roughly move in the same direction
  - If one of them is large, then others are likely to be large
  - Example: $X_1, \ldots, X_n$ are all proportional to each other
- Independence
- Negative dependence
  - Random variables roughly move in the opposite direction
  - If one of them is large, then others are likely to be small
  - Example: $(X_1, \ldots, X_n) \sim$ Multinomial
  - Very difficult to analyze if $n \geq 3$

Risk sharing games

Risk sharing, risk exchange, and market equilibria

$$X \quad \longmapsto \quad (X_1, \dots, X_n) \;\text{ s.t. } \sum_{i=1}^{n} X_i = X$$

"Canonical form" of an equilibrium allocation?

- proportional: $X_i = a_i X$ for some $\sum_{i=1}^{n} a_i = 1$?
- lottery: $X_i = \mathbb{1}_{A_i} X$ for some $\bigcup_{i=1}^{n} A_i = \Omega$?
- other forms?

Quantile-based risk sharing

utility-based	quantile-based
max $\sum_{i=1}^n \lambda_i \mathbb{E}[u_i(X_i)]$	min $\sum_{i=1}^n \lambda_i \mathrm{VaR}_{\alpha_i}(X_i)$
horizontally cut	vertically cut
$(X/n, \dots, X/n)$	$(X\mathbb{1}_{A_1}, \dots, X\mathbb{1}_{A_n})$
coinsurance	roulette
positive dependence	negative dependence

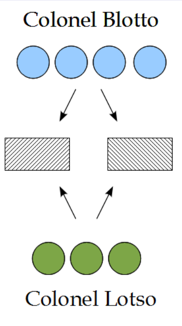
Theorem

For mixed VaR, ES, and "other similar" agents, an equilibrium allocation is extremally negatively dependent.

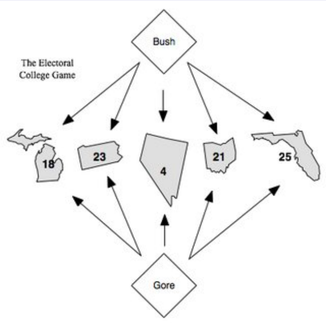
Embrechts-Liu-W., Quantile-based risk sharing.
Operations Research, 2018, Theorems 1 - 3

Colonel Blotto games

Colonel Blotto games (all-pay auctions)



- Two players
- $X_1 + \cdots + X_n = x$
- $Y_1 + \cdots + Y_n = y$
- Goal: maximize $\sum_{i=1}^{n} \mathbb{E}[f_i(X_i, Y_i)]$
  e.g. $f_i(s,t) = v_i \mathbb{1}_{\{s>t\}}$
- Nash equilibrium



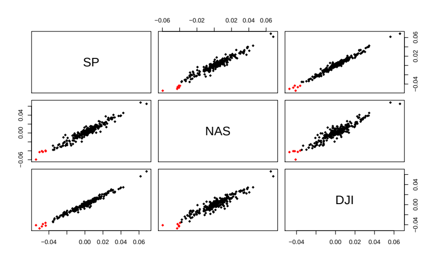
Approach:
- solve for marginals $X_1 \sim F_1, \ldots, X_n \sim F_n$
- find dependence (if possible) s.t. $X_1 + \cdots + X_n = x$
- $\Rightarrow$ Extremal negative dependence (joint mixability)

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

Axiomatic characterization of ES

Theorem

A risk measure penalizes risk concentration (a special form of positive dependence) if and only if it is an ES.



- ES is the most important risk measure in banking regulation (Basel FRTB)
- The first axiomatic characterization of ES (introduced ∼2000)

W.-Zitikis, An axiomatic foundation for the Expected Shortfall. Management Science, 2021, Theorem 1

Axiomatic characterization of ES

Axioms

M. (Monotonicity) A surely larger or equal loss leads to a larger or equal risk value, that is, $\rho(X) \le \rho(Y)$ whenever $X \le Y$.

LI. (Law-invariance) The risk value depends on the loss via its distribution, that is, $\rho(X) = \rho(Y)$ whenever $X \overset{\mathrm{d}}{=} Y$.

P. (Prudence) The risk value is not underestimated by approximations, that is, $\limsup_n \rho(\xi_n) \ge \rho(X)$ whenever $\xi_n \to X$ point-wise.

NRC. (No reward for concentration) There exists an event $A \in \mathcal{F}$ such that $\rho(X+Y) = \rho(X) + \rho(Y)$ holds for all risks X and Y sharing the tail event A.

Definition (Tail events)

A tail event of X is $A \in \mathcal{F}$ such that

a) $0 < \mathbb{P}(A) < 1$

b) $X(\omega) \ge X(\omega')$ for a.s. all $\omega \in A$ and $\omega' \in A^c$

Theorem

A functional $\rho : L^1 \to \mathbb{R}$ with $\rho(1) = 1$ satisfies Axioms **M**, **LI**, **P** and **NRC** if and only if $\rho = \mathrm{ES}_p$ for some $p \in (0,1)$.

Equivalence between risk and dependence

Theorem

Risk aversion (Rothschild-Stiglitz) $\iff$ dependence aversion

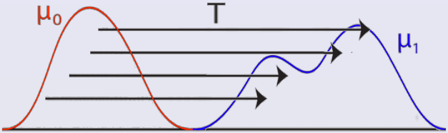
Theorem

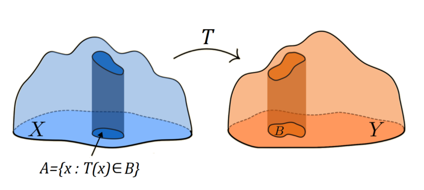
A mapping $\rho : \mathcal{X} \to \mathbb{R}$ is dependence neutral, i.e., $\rho(X+Y)$ depends only on the marginal distributions of $(X, Y) \in \mathcal{X}^2$, if and only if $\rho = f \circ \mathbb{E}$ on $\mathcal{X}$ for some $f : \mathbb{R} \to \mathbb{R}$.

W.-Wu, Dependence and risk attitudes: An equivalence. SSRN: 3707709, 2020, Theorems 1 - 2 and Proposition 3

Transport theory

- Pure mathematics
- Important applications
  - economics
  - decision theory
  - finance
  - engineering
  - operations research
  - physics
- 1 Nobel Prize laureate
- 2 Fields medalists





Monge's formulation

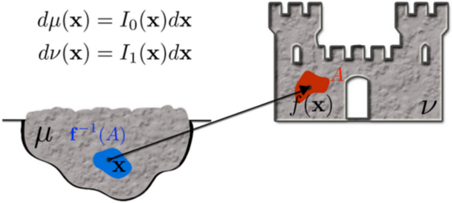
- A and B are two Radon spaces (main example: $\mathbb{R}^d$)
- Cost function $c : A \times B \to [0, \infty]$ or $(-\infty, \infty]$
- Given probability measures μ on A and ν on B
- Monge's problem: find a transport map $T : A \to B$ that attains

$$\inf \left\{ \int_A c(x, T(x))\,\mathrm{d}\mu(x) \;\middle|\; T_*(\mu) = \nu \right\},$$

where $T_*(\mu)$ is the push forward of μ by T

Monge's formulation

Gaspard Monge
1746-1818



Le mémoire sur les déblais et les remblais

(The note on land excavation and infill)

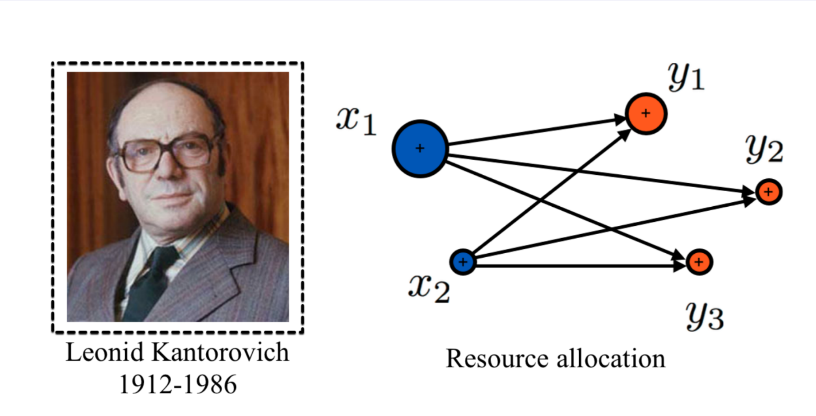
Kantorovich's formulation

- Monge's formulation may be ill-posed (e.g., point masses)
- Kantorovich's problem: find a probability measure P on $A \times B$ that attains

$$\inf \left\{ \int_{A\times B} c(x,y)\,\mathrm{d}P(x,y) \mid P \in \Gamma(\mu,\nu) \right\},$$

  where $\Gamma(\mu,\nu)$ is the set of probability measures on $A \times B$ with marginals μ and ν.
- $A \times B = \mathbb{R} \times \mathbb{R}$: copulas and dependence
- Discrete version: linear programming

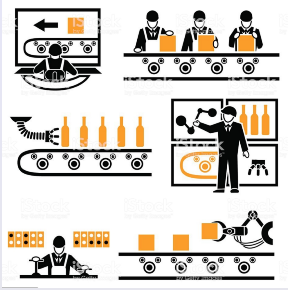
Kantorovich's formulation



Scheduling

$$\text{Minimize } \max(\Psi(\mathbf{X}))$$
$$\text{Minimize } \mathrm{Var}(\Psi(\mathbf{X}))$$
$$\text{s.t. } X_i \sim F_i,\ i = 1, \ldots, n$$

(an NP-hard problem)



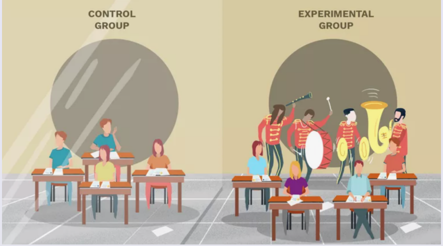
$$\begin{bmatrix} 44 & 10 & 24 \\ 66 & 32 & 37 \\ 67 & 48 & 41 \\ 71 & 57 & 43 \\ 87 & 60 & 83 \end{bmatrix} \begin{matrix} 78 \\ 135 \\ 156 \\ 171 \\ 230 \end{matrix} \Longrightarrow \begin{bmatrix} 87 & 10 & 43 \\ 71 & 60 & 24 \\ 67 & 48 & 41 \\ 44 & 32 & 83 \\ 66 & 57 & 37 \end{bmatrix} \begin{matrix} 140 \\ 155 \\ 156 \\ 159 \\ 160 \end{matrix}$$

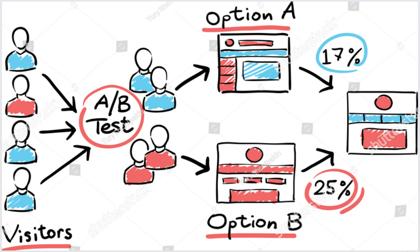
Treatment effect analysis

score X (control)

score Y (experimental)

- Marginals of (X, Y): ✓
- Effect measurement $\mathbb{E}[Y - X]$: ✓
- $\mathrm{Var}(Y - X)$: ?
- Dependence of (X, Y): unidentifiable (Neyman'23)





Meta analysis

- A (large) set of p-values is only one vector: little hope to test/verify the dependence model
- 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.*"
- Need procedures which work on arbitrarily dependent p-values
- Complicated/strange dependence arises when tests statistics are generated by some adaptive procedure
  - selective inference
  - multi-armed bandit problems

Merging p-values in multiple hypothesis testing

- $P_1, \ldots, P_K$: p-values (satisfying $\mathbb{P}(P_k \le \epsilon) \le \epsilon$)
- arbitrarily dependent
- P-merging function F:

  $\mathbb{P}(F(P_1, \ldots, P_K) \le \epsilon) \le \epsilon$ for all $(P_1, \ldots, P_K)$ and ϵ
- Find $a_{r,K}$ such that $a_{r,K} M_{r,K}$ is a p-merging function
  - Generalized average $M_{r,K}(\mathbf{p}) = (\frac{p_1^r + \cdots + p_K^r}{K})^{1/r}$

Theorem

$a_{1,K} = 2$ (arithmetic) $\quad a_{0,K} \sim e$ (geometric)

$a_{-1,K} \sim \log K$ (harmonic) $\quad a_{-\infty,K} = K$ (Bonferroni)

Vovk-W., Combining p-values via averaging. Biometrika, 2020, Theorems 1 - 2

P-hacking

Typical scientific research

- Group A tests a medication; gets "promising but not conclusive" results
- Group B continues with new data; even more promising
- Group C continues with new data ...
- Sweep all data together to recalculate p-value $\Rightarrow$ p-hacking

What is an e-value?

- A hypothesis $\mathcal{H}$: a set of probability measures

Definition (e-values and p-values)

(1) An e-value for testing $\mathcal{H}$ is a non-negative extended random variable $E : \Omega \to [0, \infty]$ that satisfies $\sup_{H \in \mathcal{H}} \int E \, \mathrm{d}H \leq 1$.

(2) A p-value for testing $\mathcal{H}$ is a random variable $P : \Omega \to [0, \infty)$ that satisfies $\sup_{H \in \mathcal{H}} H(P \leq \alpha) \leq \alpha$ for all $\alpha \in (0, 1)$.

- For simple hypothesis $\{\mathbb{P}\}$: non-negative E with mean ≤ 1
- P-test: p(data) small $\Longrightarrow$ reject
- E-test: e(data) large $\Longrightarrow$ reject

Vovk-W., E-values: Calibration, combination, and applications. Annals of Statistics, 2021

E-values, test supermartingales and betting scores

- A test supermartingale: a supermartingale $X = (X_t)$ (i.e., $\mathbb{E}[X_{t+1}|X_t] \le X_t$) under the null with $X_0 = 1$
- Optional validity (Doob's optional stopping theorem):

$$X_\tau \text{ is an e-value for any stopping time } \tau$$

- Retrospective validity (Ville's inequality):

$$\mathbb{P}\left(\sup_{t\ge 0} X_t \ge \frac{1}{\alpha}\right) \le \alpha \implies \inf_{t\ge 0} X_t^{-1} \text{ is a p-value}$$

- Bayes factors and likelihood ratios:

$$e(\text{data}) = \frac{\Pr(\text{data} \mid \mathbb{Q})}{\Pr(\text{data} \mid \mathbb{P})}$$

- Betting scores (Shafer-Vovk'19, Shafer'21)

An analogy of p-values and e-values

	requirement	specific interpretation	representative forms	keyword
p-value P	$\mathbb{P}(P \le \alpha) \le \alpha$ for $\alpha \in (0,1)$	probability of a more extreme observation	$\mathbb{P}(T' \le T(\mathbf{X}) \vert \mathbf{X})$	(conditional) probability
e-value E	$\mathbb{E}^{\mathbb{P}}[E] \le 1$ and $E \ge 0$	likelihood ratios, stopped martingales, and betting scores	$\mathbb{E}^{\mathbb{P}}\left[\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} \middle\vert \mathbf{X}\right]$ $\mathbb{E}^{\mathbb{P}}[M_\tau \vert \mathbf{X}]$	(conditional) expectation

An analogy of p-variables and e-variables for a simple hypothesis $\{\mathbb{P}\}$

- $\mathbf{X}$ is data
- $T(\mathbf{X})$ is any test statistic
- T' is an independent copy of $T(\mathbf{X})$ under $\mathbb{P}$
- $\mathbb{Q}$ is any probability measure
- M is a test supermartingale under $\mathbb{P}$ and τ a stopping time

Advantages of e-values

- Validity for arbitrary dependence $\Rightarrow$ expectation
- Validity for optional stopping times $\Rightarrow$ martingale

E-values are a useful tool even if one is only interested in p-values

- Easy to combine
- Flexible to stop/continue (online testing; unfixed sample size)
- Non-asymptotic and often model-free

Vovk-Wang-W., Admissible ways of merging p-values under arbitrary dependence. Annals of Statistics, 2021, Theorem 5.1

Example: Multi-armed bandit problems

- K arms
- H_k: arm k has mean reward at most 1
- Strategy (k_t): at time t, pull arm k_t, get iid reward $X_{k_t,t} \geq 0$
  - optimized strategy
- Aim: quickly detect arms with mean > 1
  - or maximize profit, minimize regret, etc ...
- Running reward: $M_{k,t} = \prod_{j=1}^{t} X_{k,j} \mathbb{1}_{\{k_j=k\}}$
- Complicated dependence due to exploration/exploitation
- $M_{1,\tau}, \ldots, M_{K,\tau}$ are e-values for any stopping time τ

Selective inference

Basic framework

- K hypotheses $H_1, \ldots, H_K$
- $\mathcal{K} = \{1, \ldots, K\}$
- H_k is null if $\mathbb{P} \in H_k$
- $\mathcal{N} \subseteq \mathcal{K}$: the set of (unknown) indices of null hypotheses
- $K_0 = |\mathcal{N}|$; if $K_0/K \approx 1$ then the signals are sparse

Examples

- Drug experiments; brain imaging; investment opportunities; A/B tests; genome-wide association studies

Selective inference

For a testing procedure $\mathcal{D} : [0,1]^K \to 2^{\mathcal{K}}$ or $[0,\infty]^K \to 2^{\mathcal{K}}$:

- $R_{\mathcal{D}}$: number of total discoveries ($R_{\mathcal{D}} = |\mathcal{D}|$)
- $F_{\mathcal{D}}$: number of false discoveries ($F_{\mathcal{D}} = |\mathcal{D} \cap \mathcal{N}|$)
- False discovery proportion (FDP): $F_{\mathcal{D}}/R_{\mathcal{D}}$ with $0/0 = 0$
- Benjamini-Hochberg'95: control the FDR $\mathbb{E}[F_{\mathcal{D}}/R_{\mathcal{D}}] \leq \alpha$

BH procedure

BH procedure

The Benjamini-Hochberg (BH) procedure $\mathcal{D}(\alpha)$ rejects hypotheses with the smallest k^* p-values, where

$$k^* = \max\left\{k \in \mathcal{K} : \frac{Kp_{(k)}}{k} \le \alpha\right\}.$$

	FDR	dependence
BH'95 BY'01	$\frac{K_0}{K}\alpha$	independence PRDS
BY'01	$\ell_K \frac{K_0}{K}\alpha$	arbitrary

$\ell_K = \sum_{j=1}^{K} j^{-1} \approx \log K$. PRDS: positive regression dependence on a subset, e.g., jointly Gaussian test statistics with correlations ≥ 0

E-BH procedure

- $e_{[1]} \geq \cdots \geq e_{[K]}$: order statistics of arbitrary e-values

E-BH procedure

The e-BH procedure $\mathcal{G}(\alpha) : [0,\infty]^K \to 2^{\mathcal{K}}$ for $\alpha > 0$ rejects hypotheses with the largest k^* e-values, where

$$k^* = \max\left\{k \in \mathcal{K} : \frac{k e_{[k]}}{K} \geq \frac{1}{\alpha}\right\}.$$

Theorem

The e-BH procedure always has FDR at most $K_0\alpha/K$.

W.-Ramdas, False discovery rate control with e-values. arXiv: 2009.02824, 2020, Theorem 5.1

Combination and multiple testing

	arbitrarily dependent	optimality	sequential/independent	optimality
p-values $P_1, \dots, P_K$	Bonferroni robust averaging many others	NO	Fisher Simes many others	NO
e-values $E_1, \dots, E_K$	arithmetic mean	YES	product martingale merging	weakly NO

	FDR	dependence
BH procedure	$\frac{K_0}{K}\alpha$	independence/PRDS
BY procedure	$\ell_K \frac{K_0}{K}\alpha$	arbitrary
e-BH procedure	$\frac{K_0}{K}\alpha$	arbitrary

FDR procedures ($K_0 = \#$nulls, $K = \#$hypotheses, $\ell_K = \sum_{k=1}^{K} k^{-1} \approx \log K$)

Thank you

Thank you for your attention!

A multi-armed bandit problem

Problem setting

- K arms each with a reward $X^k \geq 0$
- Pulling arm k produces an iid sample $(X_1^k, X_2^k, \dots)$ from X^k
- Null hypotheses: $\mathbb{E}[X_k] \leq 1,\ k \in \mathcal{K}$
- Arms have to be pulled in order and previous arms cannot be revisited
- An arm can be pulled at most n times (budget)
- Goal: detect non-null arms as quickly as possible
- Example: investment opportunities; medical experiment

A multi-armed bandit problem

The e-value $e_{k,j}$ and the p-value $p_{k,j}$ are realized by, respectively,

$$E_{k,j} := \prod_{i=1}^{j} X_i^k \quad \text{and} \quad P_{k,j} := \left(\max_{i=1,\dots,j} E_{k,i} \right)^{-1} \quad (p \le 1/e)$$

Algorithm

- Select a p- or e-testing procedure $\mathcal{D}$ and start with $\mathbf{e} = \mathbf{p} = \mathbf{1}$
- For arm k, stop at T_k such that either $\mathcal{D}$ produces a new discovery or $T_k = n$
- Update e-values or p-values and move to arm $k + 1$

The final e-variables E_k and p-variables P_k are obtained by

$$E_k = E_{k,T_k} \quad \text{and} \quad P_k = P_{k,T_k}, \quad k = 1, \dots, K.$$

A multi-armed bandit problem

Table: Conditions for the validity of the testing algorithm

	AD data across arms	AD stopping rules T_k	FDR guarantee in our experiments
e-BH	YES	YES	valid at level $\alpha K_0/K$
BH	NO	NO	not valid
BY	YES	YES	valid at level $\alpha K_0/K$
cBH	NO	YES	valid at level $\alpha K_0/K$

Consider BH, e-BH, BY and compliant BH (cBH) procedures

- BY: $\mathcal{D}(\alpha_1)$ where $\alpha_1 \ell_K = \alpha$ (Benjamini-Yekutieli'01)
- cBH: $\mathcal{D}(\alpha_2)$ where $\alpha_2(1 + \log(1/\alpha_2)) = \alpha$ (Su'18)

A multi-armed bandit problem

Data generating process

- More promising arms come first: arm k is non-null with probability $\theta(K-k+1)/(K+1)$, $\theta \in [0,1]$
- The expected number of non-nulls in this setting is $\theta/2$
- $s_k \sim \mathrm{Expo}(\mu)$ is the strength of signal for arm k
- Conditional on s_k,

$$X_1^k, \ldots, X_n^k \overset{\text{iid}}{\sim} X^k = \exp\left(Z^k + s_k \mathbb{1}_{\{k \in \mathcal{K} \setminus \mathcal{N}\}} - 1/2\right)$$

  where $Z^1, \ldots, Z^K$ are iid standard normal.
- Set $\alpha = 0.05$ and $\theta = 0.5$

A multi-armed bandit problem

Table: $R = \#\{\text{rejected hypothesis}\}$, $B\% = \%(\text{unused budget})$, $\text{TD} = \#\{\text{true discoveries}\}$. Each number is computed over an average of 500 trials. Default values: $K = 500$, $n = 50$ and $\mu = 1$.

(a) Default

	R	$B\%$	TD	FDP%
e-BH	74.4	11.42	73.2	1.58
BH	77.0	11.44	75.3	2.13
BY	70.6	10.06	70.4	0.31
cBH	71.1	10.16	70.8	0.36

(b) $K = 2000$

	R	$B\%$	TD	FDP%
e-BH	297.6	11.39	293.2	1.48
BH	307.8	11.41	301.4	2.07
BY	281.2	9.95	280.4	0.26
cBH	284.5	10.15	283.5	0.36

(c) $n = 10$

	R	$B\%$	TD	FDP%
e-BH	47.7	3.99	47.3	0.83
BH	49.3	4.01	48.7	1.06
BY	38.4	2.77	38.4	0.08
cBH	39.2	2.85	39.2	0.11

(d) $n = 100$

	R	$B\%$	TD	FDP%
e-BH	79.1	13.48	77.9	1.50
BH	81.3	13.50	79.5	2.13
BY	76.4	12.36	76.1	0.35
cBH	76.7	12.44	76.4	0.41

(e) $\mu = 0.5$

	R	$B\%$	TD	FDP%
e-BH	43.5	5.77	42.9	1.54
BH	46.3	5.80	45.3	2.13
BY	39.6	4.66	39.5	0.27
cBH	40.1	4.74	40.0	0.35

(f) $\mu = 2$

	R	$B\%$	TD	FDP%
e-BH	97.4	16.46	95.9	1.54
BH	99.3	16.47	97.2	2.07
BY	94.3	15.23	94.1	0.29
cBH	94.6	15.32	94.3	0.35

Calibration and combination

- Admissible p-to-e calibrators
  - Power calibrators: $f_\kappa(p) = \kappa p^{\kappa-1}$ for $\kappa \in (0,1)$
  - Shafer's: $f(p) = p^{-1/2} - 1$
  - Averaging f_κ: $\int_0^1 \kappa p^{\kappa-1} \mathrm{d}\kappa = \frac{1-p+p\ln p}{p(-\ln p)^2}$
- The only admissible e-to-p calibrator: $e \to (1/e) \wedge 1$
- Very roughly: $p \sim 1/e$

E-merging functions

E-merging functions

- arithmetic average M_K: arbitrary dependence
- product P_K: independence

Theorem 1

Suppose that F is a symmetric e-merging function. Then $F \leq \lambda + (1-\lambda)M_K$ for some $\lambda \in [0,1]$, and F is admissible if and only if $F = \lambda + (1-\lambda)M_K$ with $\lambda = F(\mathbf{0})$.

Vovk-W., E-values: Calibration, combination, and applications. Annals of Statistics, 2021, Theorem 3.2

Connection to p-merging

Theorem 2

For any admissible p-merging function F and $\epsilon \in (0,1)$, there exist $(\lambda_1, \dots, \lambda_K) \in \Delta_K$ (standard symplex) and admissible calibrators $f_1, \dots, f_K$ s.t.

$$F(\mathbf{p}) \le \epsilon \iff \sum_{k=1}^{K} \lambda_k f_k(p_k) \ge \frac{1}{\epsilon}.$$

If F is symmetric, then there exists an admissible calibrator f s.t.

$$F(\mathbf{p}) \le \epsilon \iff \frac{1}{K} \sum_{k=1}^{K} f(p_k) \ge \frac{1}{\epsilon}.$$

Vovk-Wang-W., Admissible ways of merging p-values under arbitrary dependence. Annals of Statistics, 2021, Theorem 5.1

Compliant procedures

An e-testing procedure $\mathcal{G}$ is said to be compliant at level $\alpha \in (0,1)$ if every rejected e-value e_k satisfies

$$e_k \geq \frac{K}{\alpha R_{\mathcal{G}}}.$$

- The base e-BH procedure is compliant and it dominates all other compliant procedures

Compliant procedures

Proposition 1

Any compliant e-testing procedure at level α has FDR at most $\alpha K_0/K$ for arbitrary configurations of e-values.

Proof. Let $\mathcal{G}$ be a compliant e-testing procedure. The FDP of $\mathcal{G}$ satisfies

$$\frac{F_{\mathcal{G}}}{R_{\mathcal{G}}} = \frac{|\mathcal{G}(\mathbf{E}) \cap \mathcal{N}|}{R_{\mathcal{G}} \vee 1} = \sum_{k \in \mathcal{N}} \frac{\mathbb{1}_{\{k \in \mathcal{G}(\mathbf{E})\}}}{R_{\mathcal{G}} \vee 1} \le \sum_{k \in \mathcal{N}} \frac{\mathbb{1}_{\{k \in \mathcal{G}(\mathbf{E})\}} \alpha E_k}{K} \le \sum_{k \in \mathcal{N}} \frac{\alpha E_k}{K},$$

where the first inequality is due to compliance. As $\mathbb{E}[E_k] \le 1$ for $k \in \mathcal{N}$, we have

$$\mathbb{E}\left[\frac{F_{\mathcal{G}}}{R_{\mathcal{G}}}\right] \le \sum_{k \in \mathcal{N}} \mathbb{E}\left[\frac{\alpha E_k}{K}\right] \le \frac{\alpha K_0}{K}.$$

Compliant procedures

- General compliant p-testing procedures do not have this property even if p-values are independent
- For independent p-values, a compliant p-testing procedure at α has a weaker FDR guarantee $\alpha(1 + \log(1/\alpha)) > \alpha$ (Su'18)

Compliance is useful in

- data-driven structured settings
- post-selection testing
- group testing
- multi-armed bandit problems

Boosting

For each $k \in \mathcal{K}$, take a boosting factor $b_k \geq 1$ such that

$$\max_{x \in K/\mathcal{K}} x\mathbb{P}(\alpha b_k E_k \geq x) \leq \alpha \qquad \text{if e-values are PRDS}$$

$$\mathbb{E}[T(\alpha b_k E_k)] \leq \alpha \qquad \text{otherwise (AD)}$$

and let $e'_k = b_k e_k$.

- $\mathbb{E}$ and $\mathbb{P}$ are computed under the null distribution of E_k
- Composite null: require for all probability measures in H_k
- $b_k = 1$ is always valid
- Non-linear boosting is also possible
- $\mathbf{e}'$ may not have the same order as $\mathbf{e}$.

E-BH procedure

Example.

- For $\lambda \in (0,1)$
$$E_k = \lambda P_k^{\lambda-1},$$
where P_k is standard uniform if $k \in \mathcal{N}$
- $y_\alpha \le (\lambda^\lambda \alpha)^{1/(1-\lambda)}$
- $\lambda = 1/2 \implies y_\alpha \le \alpha^2/2$
- $\alpha = 0.05,\ \lambda = 1/2$
  - $b_k \approx 6.32$ (AD)
  - $b_k \approx 8.94$ (PRDS)

E-BH procedure

Example.

- For $\delta > 0$,

$$E_k = e^{\delta X_k - \delta^2/2},$$

where X_k is standard normal if $k \in \mathcal{N}$

- $\alpha = 0.05$, $\delta = 3$
  - $b \approx 1.37$ (AD)
  - $b \approx 7.88$ (PRDS)

Correlated z-tests

- $X_k \sim \mathrm{N}(0,1)$ if $k \in \mathcal{N}$
- $X_k \sim \mathrm{N}(\delta,1)$ if $k \notin \mathcal{N}$, $\delta < 0$
- $X_1, \ldots, X_K$ are jointly Gaussian
- E-values from likelihood ratios

$$E_k = \exp(\delta X_k - \delta^2/2)$$

- P-values from Neyman-Pearson tests

$$P_k = \Phi(X_k)$$

- Set $\delta = -3$

Correlated z-tests

Table: Simulation results for correlated z-tests, where $\rho_{i,j}$ is the correlation between two test statistics X_i and X_j for $i \neq j$. Each cell gives the number of rejections and, in parentheses, the realized FDP (in %). Each number is computed over an average of 1,000 trials.

(a) Independent and positively correlated tests, $K = 1000$, $K_0 = 800$

	$\rho_{ij} = 0$			$\rho_{ij} = 0.5$		
	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 2\%$	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 2\%$
BH	177.3 (8.01)	148.7 (4.07)	115.0 (1.63)	180.0 (7.00)	144.8 (3.64)	109.8 (1.50)
e-BH PRDS	171.8 (7.07)	147.6 (3.95)	114.6 (1.62)	170.2 (5.71)	142.5 (3.35)	108.0 (1.50)
BY	101.1 (1.10)	78.8 (0.57)	53.2 (0.22)	96.6 (1.03)	76.7 (0.50)	55.0 (0.20)
e-BH AD	109.4 (1.41)	85.4 (0.68)	54.6 (0.24)	103.1 (1.32)	81.4 (0.70)	56.6 (0.28)
base e-BH	97.5 (1.00)	70.6 (0.43)	36.9 (0.11)	91.9 (0.97)	69.1 (0.45)	43.6 (0.16)

Correlated z-tests

(b) Independent tests with large number of hypotheses

	$K = 20,000,\ K_0 = 10,000$			$K = 20,000,\ K_0 = 19,000$		
	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 2\%$	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 2\%$
BH	9567 (5.00)	8564 (2.49)	7164 (1.00)	681.3 (9.58)	520.2 (4.79)	357.7 (1.93)
e-BH PRDS	9092 (3.60)	8330 (2.13)	7124 (0.98)	681.3 (9.58)	509.3 (4.54)	312.1 (1.40)
BY	5956 (0.48)	4818 (0.24)	3417 (0.10)	254.1 (0.89)	177.6 (0.46)	103.1 (0.19)
e-BH AD	6811 (0.80)	5809 (0.44)	4384 (0.18)	271.0 (1.02)	159.5 (0.39)	51.4 (0.07)
base e-BH	6426 (0.64)	5234 (0.31)	3509 (0.10)	224.8 (0.69)	109.2 (0.21)	16.4 (0.01)

(c) Negatively correlated tests, $K = 1000,\ K_0 = 800$.

	$\rho_{ij} = -1/(K-1)$			$\rho_{ij} = -0.5\mathbb{1}_{\{\lvert i-j\rvert=1\}}$		
	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 2\%$	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 2\%$
BH	177.7 (8.14)	149.0 (4.09)	115.2 (1.61)	177.2 (8.10)	148.8 (4.00)	115.3 (1.62)
e-BH PRDS	172.0 (7.13)	147.9 (3.98)	114.9 (1.59)	171.5 (7.13)	147.7 (3.89)	114.9 (1.61)
BY	101.2 (1.08)	78.8 (0.52)	53.3 (0.20)	101.3 (1.11)	78.8 (0.56)	53.2 (0.22)
e-BH AD	109.7 (1.38)	85.5 (0.65)	54.6 (0.22)	109.8 (1.40)	85.6 (0.69)	54.6 (0.24)
base e-BH	97.8 (0.98)	70.7 (0.40)	37.2 (0.11)	97.6 (0.99)	70.7 (0.41)	36.7 (0.12)