The algebra of integrated partial belief systems

Manuele Leonelli[1], Eva Riccomagno[2], James Q. Smith[1]

[1]Department of Statistics, The University of Warwick

[2]Dipartimento di Matematica, Universitá degli Studi di Genova

Algebraic Statistics, Genova
June 10, 2015

Research funded by the Department of Statistics, The University of Warwick, and EPRSC grant EP/K039628/1.

Modelling big systems

- Current decision support systems address complex domains: e.g. nuclear emergency management, food security;
- decision making $\Longrightarrow$ Bayesian subjective probabilities;
- single agents systems are well established, but no clear extension to multi-agent;
- distributed and exact (symbolic) computations are vital;

Notation

- random vector $\boldsymbol{Y} = (\boldsymbol{Y}_i^{\mathsf{T}})_{i \in [m]}$, $[m] = \{1, \ldots, m\}$;
- panels of experts $\{G_1, \ldots, G_m\}$, where G_i is responsible for $\boldsymbol{Y}_i$;
- decision space $d \in \mathcal{D}$;
- $\boldsymbol{\theta}_i$ parametrizes f_i the density of $\boldsymbol{Y}_i \mid (\boldsymbol{\theta}_i, d)$;
- π_i is the density over $\boldsymbol{\theta}_i \mid d$;

- d^* optimal policy maximizing the *expected utility*

$$\bar{u}(d) = \int_{\Theta} \bar{u}(d \mid \theta) \pi(\theta \mid d) \mathrm{d}\theta$$

where

$$\bar{u}(d \mid \theta) = \int_{\mathcal{Y}} u(\boldsymbol{y}, d) f(\boldsymbol{y} \mid \theta, d) \mathrm{d}\boldsymbol{y}$$

is the **conditional expected utility** (CEU).

Notation

- random vector $\boldsymbol{Y} = (\boldsymbol{Y}_i^{\mathsf{T}})_{i\in[m]}$, $[m] = \{1, \dots, m\}$;
- panels of experts $\{G_1, \dots, G_m\}$, where G_i is responsible for $\boldsymbol{Y}_i$;
- decision space $d \in \mathcal{D}$;
- $\boldsymbol{\theta}_i$ parametrizes f_i the density of $\boldsymbol{Y}_i \mid (\boldsymbol{\theta}_i, d)$;
- π_i is the density over $\boldsymbol{\theta}_i \mid d$;

- d^* optimal policy maximizing the *expected utility*

$$\bar{u}(d) = \int_{\boldsymbol{\Theta}} \bar{u}(d \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta} \mid d)\mathrm{d}\boldsymbol{\theta}$$

where

$$\bar{u}(d \mid \boldsymbol{\theta}) = \int_{\boldsymbol{\mathcal{Y}}} u(\boldsymbol{y}, d)f(\boldsymbol{y} \mid \boldsymbol{\theta}, d)\mathrm{d}\boldsymbol{y}$$

is the **conditional expected utility** (CEU).

Utility theory

Utility $u : \boldsymbol{\mathcal{Y}} \times \mathcal{D} \to \mathbb{R}$ such that

$$(\boldsymbol{y}, d) \preceq (\boldsymbol{y}', d') \iff u(\boldsymbol{y}, d) \leq u(\boldsymbol{y}', d')$$

Panel separable factorization

$$u(\boldsymbol{y}, d) = \sum_{I \in \mathcal{P}_0([m])} k_I \prod_{i \in I} u_i(\boldsymbol{y}_i, d),$$

with $\mathcal{P}_0$ the power set without empty set.

Polynomial marginal utility (univariate) of degree n_i, $\rho_{ij} \in \mathbb{R}$,

$$u(y_i, d) = \sum_{j \in [n_i]} \rho_{ij}(d) y_i^j.$$

Integrated partial belief systems

Panels agrees on:

- a decision space $\mathcal{D}$;
- a family of utility functions $\mathcal{U}$;
- a dependence structure between various functions of $\boldsymbol{Y}$, $\boldsymbol{\theta}$ and d;
- to delegate quantifications to the most informed panel.

Definition

An IPBS is **adequate** if $\bar{u}(d)$, for each $d \in \mathcal{D}$ and $u \in \mathcal{U}$, can be computed from the beliefs of G_i, $i \in [m]$.

Integrated partial belief systems

Panels agrees on:

- a decision space $\mathcal{D}$;
- a family of utility functions $\mathcal{U}$;
- a dependence structure between various functions of $\boldsymbol{Y}$, $\boldsymbol{\theta}$ and d;
- to delegate quantifications to the most informed panel.

Definition

An IPBS is **adequate** if $\bar{u}(d)$, for each $d \in \mathcal{D}$ and $u \in \mathcal{U}$, can be computed from the beliefs of G_i, $i \in [m]$.

Algebraic expected utility

$\bar{u}(d \mid \boldsymbol{\theta})$ is called *algebraic in the panels* if, for each $d \in \mathcal{D}$, there exist $\lambda_i(\boldsymbol{\theta}_i, d)$ such that $\bar{u}(d \mid \boldsymbol{\theta})$ is a square-free polynomial q_d of the $\boldsymbol{\lambda}_i$

$$\bar{u}(d \mid \boldsymbol{\theta}) = q_d\left(\boldsymbol{\lambda}_1(\boldsymbol{\theta}_1, d), \cdots, \boldsymbol{\lambda}_m(\boldsymbol{\theta}_m, d)\right).$$

Let $\boldsymbol{\lambda}_i(\boldsymbol{\theta}_i, d) = (\lambda_{ji}(\boldsymbol{\theta}_i, d))_{j \in [s_i]}$, $\lambda_{0i}(\boldsymbol{\theta}_i, d) = 1$ and $B = \times_{i \in [m]} \{0, \ldots, s_i\}$. For a given $\boldsymbol{b} \in B$ let

$$b_{j,i} = 0 \text{ if } j \neq b_i, \qquad b_{j,i} = 1 \text{ if } j = b_i, \qquad b_{0,i} = 1.$$

Definition

$\bar{u}(d \mid \theta)$ is called **algebraic** if, for each $d \in \mathcal{D}$, q_d is a square-free polynomial of the λ_{ji} such that

$$q_d\left(\boldsymbol{\lambda}_1(\theta_1, d), \ldots, \boldsymbol{\lambda}_m(\theta_m, d)\right) = \sum_{\boldsymbol{b} \in B} k_{\boldsymbol{b},d} \lambda_{\boldsymbol{b}}(\theta, d),$$

$$\lambda_{\boldsymbol{b}}(\theta, d) = \prod_{i \in [m]} \prod_{j \in [s_i]^0} \lambda_{ji}(\theta_i, d)^{b_{j,i}}.$$

Algebraic expected utility

$\bar{u}(d \mid \boldsymbol{\theta})$ is called *algebraic in the panels* if, for each $d \in \mathcal{D}$, there exist $\lambda_i(\boldsymbol{\theta}_i, d)$ such that $\bar{u}(d \mid \boldsymbol{\theta})$ is a square-free polynomial q_d of the $\boldsymbol{\lambda}_i$

$$\bar{u}(d \mid \boldsymbol{\theta}) = q_d\left(\boldsymbol{\lambda}_1(\boldsymbol{\theta}_1, d), \cdots, \boldsymbol{\lambda}_m(\boldsymbol{\theta}_m, d)\right).$$

Let $\boldsymbol{\lambda}_i(\boldsymbol{\theta}_i, d) = (\lambda_{ji}(\boldsymbol{\theta}_i, d))_{j \in [s_i]}$, $\lambda_{0i}(\boldsymbol{\theta}_i, d) = 1$ and $B = \times_{i \in [m]} \{0, \ldots, s_i\}$. For a given $\boldsymbol{b} \in B$ let

$$b_{j,i} = 0 \text{ if } j \neq b_i, \qquad b_{j,i} = 1 \text{ if } j = b_i, \qquad b_{0,i} = 1.$$

Definition

$\bar{u}(d \mid \boldsymbol{\theta})$ is called **algebraic** if, for each $d \in \mathcal{D}$, q_d is a square-free polynomial of the λ_{ji} such that

$$q_d\left(\boldsymbol{\lambda}_1(\boldsymbol{\theta}_1, d), \ldots, \boldsymbol{\lambda}_m(\boldsymbol{\theta}_m, d)\right) = \sum_{\boldsymbol{b} \in B} k_{\boldsymbol{b},d} \lambda_{\boldsymbol{b}}(\boldsymbol{\theta}, d),$$

$$\lambda_{\boldsymbol{b}}(\boldsymbol{\theta}, d) = \prod_{i \in [m]} \prod_{j \in [s_i]^0} \lambda_{ji}(\boldsymbol{\theta}_i, d)^{b_{j,i}}.$$

Score separability

For a given $\boldsymbol{b} \in B$, let

$$\mu_{ji}(d) = \mathbb{E}\left(\lambda_{ji}(\boldsymbol{\theta}_i, d)^{b_{j,i}}\right), \qquad \boldsymbol{\mu}_i(d) = \left(\mu_{ji}(d)\right)_{j \in [s_i]}.$$

Definition

Call an IPBS **score separable** if, for all $d \in \mathcal{D}$ and all $\boldsymbol{b} \in B$ such that $k_{\boldsymbol{b},d} \neq 0$,

$$\mathbb{E}\left(\lambda_{\boldsymbol{b}}(\boldsymbol{\theta}, d)\right) = \prod_{i \in [m]} \prod_{j \in [s_i] \cup \{0\}} \mu_{ji}(d).$$

Lemma

Suppose panel G_i delivers $\boldsymbol{\mu}_i(d)$, $i \in [m]$, $d \in \mathcal{D}$. Then, assuming a CEU is algebraic, if the IPBS is score separable then it is adequate.

New independence conditions

Definition (Quasi independence)

An IPBS is called quasi independent if

$$\mathbb{E}(q_d(\boldsymbol{\lambda}_1(\boldsymbol{\theta}_1, d), \ldots, \boldsymbol{\lambda}_m(\boldsymbol{\theta}_m, d))) = q_d(\mathbb{E}(\boldsymbol{\lambda}_1(\boldsymbol{\theta}_1, d)), \ldots, \mathbb{E}(\boldsymbol{\lambda}_m(\boldsymbol{\theta}_m, d))).$$

Let $\boldsymbol{\theta} = \theta_1 \cdots \theta_n$, $\boldsymbol{a}, \boldsymbol{c} \in \mathbb{Z}^n_{\geq 0}$.

Definition (Moment independence)

θ entertains moment independence of order $\boldsymbol{c}$ if for any $\boldsymbol{a} \leq_{lex} \boldsymbol{c}$

$$\mathbb{E}(\boldsymbol{\theta}^a) = \prod_{i \in [n]} \mathbb{E}(\theta_i^{a_i}).$$

New independence conditions

Definition (Quasi independence)

An IPBS is called quasi independent if

$$\mathbb{E}(q_d(\boldsymbol{\lambda}_1(\theta_1, d), \ldots, \boldsymbol{\lambda}_m(\theta_m, d))) = q_d(\mathbb{E}(\boldsymbol{\lambda}_1(\theta_1, d)), \ldots, \mathbb{E}(\boldsymbol{\lambda}_m(\theta_m, d))).$$

Let $\boldsymbol{\theta} = \theta_1 \cdots \theta_n$, $\boldsymbol{a}, \boldsymbol{c} \in \mathbb{Z}^n_{\geq 0}$.

Definition (Moment independence)

$\boldsymbol{\theta}$ entertains moment independence of order $\boldsymbol{c}$ if for any $\boldsymbol{a} \leq_{lex} \boldsymbol{c}$

$$\mathbb{E}(\boldsymbol{\theta}^{a}) = \prod_{i \in [n]} \mathbb{E}(\theta_i^{a_i}).$$

Moment independence

Consider two parameters θ_1 and θ_2 and suppose moment independence of order $(2, 2)$.

$$\mathbb{E}(\theta_1^2\theta_2^2) = \mathbb{E}(\theta_1^2)\mathbb{E}(\theta_2^2) = \mathbb{E}(\theta_1)^2\mathbb{E}(\theta_2)^2 + \mathbb{V}(\theta_1)\mathbb{E}(\theta_2)^2 + \mathbb{E}(\theta_1)^2\mathbb{V}(\theta_2) + \mathbb{V}(\theta_1)\mathbb{V}(\theta_2)$$

If $\theta_1 \perp\!\!\!\perp \theta_2$,

$$\begin{aligned}
\mathbb{E}(\theta_1^2\theta_2^2) &= \mathbb{E}(\theta_1\theta_2)^2 + \mathbb{V}(\theta_1\theta_2) \\
&= \mathbb{E}(\theta_1\mathbb{E}(\theta_2))^2 + \mathbb{V}(\theta_1\mathbb{E}(\theta_2)) + \mathbb{E}(\theta_1^2\mathbb{V}(\theta_2)) \\
&= \mathbb{E}(\theta_1)^2\mathbb{E}(\theta_2)^2 + \mathbb{V}(\theta_1)\mathbb{E}(\theta_2)^2 + \mathbb{E}(\theta_1^2)\mathbb{V}(\theta_2) \\
&= \mathbb{E}(\theta_1)^2\mathbb{E}(\theta_2)^2 + \mathbb{V}(\theta_1)\mathbb{E}(\theta_2)^2 + \mathbb{E}(\theta_1)^2\mathbb{V}(\theta_2) + \mathbb{V}(\theta_1)\mathbb{V}(\theta_2)
\end{aligned}$$

Moment independence

Consider two parameters θ_1 and θ_2 and suppose moment independence of order $(2, 2)$.

$$
\begin{aligned}
\mathbb{E}(\theta_1^2\theta_2^2) = \mathbb{E}(\theta_1^2)\mathbb{E}(\theta_2^2) = \mathbb{E}(\theta_1)^2\mathbb{E}(\theta_2)^2 + \mathbb{V}(\theta_1)\mathbb{E}(\theta_2)^2 \\
+ \mathbb{E}(\theta_1)^2\mathbb{V}(\theta_2) + \mathbb{V}(\theta_1)\mathbb{V}(\theta_2)
\end{aligned}
$$

If $\theta_1 \perp\!\!\!\perp \theta_2$,

$$
\begin{aligned}
\mathbb{E}(\theta_1^2\theta_2^2) &= \mathbb{E}(\theta_1\theta_2)^2 + \mathbb{V}(\theta_1\theta_2) \\
&= \mathbb{E}(\theta_1\mathbb{E}(\theta_2))^2 + \mathbb{V}(\theta_1\mathbb{E}(\theta_2)) + \mathbb{E}(\theta_1^2\mathbb{V}(\theta_2)) \\
&= \mathbb{E}(\theta_1)^2\mathbb{E}(\theta_2)^2 + \mathbb{V}(\theta_1)\mathbb{E}(\theta_2)^2 + \mathbb{E}(\theta_1^2)\mathbb{V}(\theta_2) \\
&= \mathbb{E}(\theta_1)^2\mathbb{E}(\theta_2)^2 + \mathbb{V}(\theta_1)\mathbb{E}(\theta_2)^2 + \mathbb{E}(\theta_1)^2\mathbb{V}(\theta_2) + \mathbb{V}(\theta_1)\mathbb{V}(\theta_2)
\end{aligned}
$$

Some results

Theorem

Under quasi independence, an algebraic CEU is score separable.

Corollary

Let

- $\lambda_{ji}(\boldsymbol{\theta}_i, d) = \theta_i^{\boldsymbol{a}_{ji}}$, $\boldsymbol{a}_{ji} \in \mathbb{Z}_{\geq 0}^{s_i}$;
- $\boldsymbol{a}_i^* = (a_{ji}^*)_{j \in [s_i]}$, *where* $a_{ji}^* = \max\{a_{ji} \mid j \in [s_i]\}$;
- $\boldsymbol{\theta} = (\boldsymbol{\theta}_i^{\mathsf{T}})$ *moment independent of order* $\boldsymbol{a}^* = (\boldsymbol{a}_i^{*\mathsf{T}})$;

an algebraic CEU is score separable.

Polynomial SEMs

A polynomial structural equation model (SEM) over $\boldsymbol{Y} = (Y_i)_{i\in[m]}$ is defined as

$$Y_i = \sum_{\boldsymbol{a}_i \in A_i} \theta_{i\boldsymbol{a}_i} \boldsymbol{Y}_{[i-1]}^{\boldsymbol{a}_i} + \varepsilon_i,$$

where $A_i \subset \mathbb{Z}_{\geq 0}^{i-1}$, $\varepsilon_i \sim (0, \psi_i)$, $\boldsymbol{Y}_{[i-1]} = Y_1 \cdots Y_{i-1}$.

Theorem

Assume

- *a polynomial SEM;*
- *a panel separable utility;*
- *a marginal polynomial utility;*

The IPBS is score separable under quasi independence.

Polynomial SEMs

A polynomial structural equation model (SEM) over $\boldsymbol{Y} = (Y_i)_{i\in[m]}$ is defined as

$$Y_i = \sum_{\boldsymbol{a}_i \in A_i} \theta_{i\boldsymbol{a}_i} \boldsymbol{Y}_{[i-1]}^{\boldsymbol{a}_i} + \varepsilon_i,$$

where $A_i \subset \mathbb{Z}_{\geq 0}^{i-1}$, $\varepsilon_i \sim (0, \psi_i)$, $\boldsymbol{Y}_{[i-1]} = Y_1 \cdots Y_{i-1}$.

Theorem

Assume

- *a polynomial SEM;*
- *a panel separable utility;*
- *a marginal polynomial utility;*

The IPBS is score separable under quasi independence.

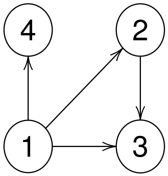
Bayesian networks

Definition (Linear SEM)

A BN over a DAG $\mathcal{G}$, $V(\mathcal{G}) = \{1, \ldots, m\}$, is a linear SEM if

$$Y_i = \theta_{0i} + \sum_{j \in \Pi_i} \theta_{ji} Y_j + \varepsilon_i,$$

Π_i parent set of Y_i, $\varepsilon_i \sim (0, \psi_i)$ and $\theta_{0i}, \theta_{ji} \in \mathbb{R}$.



$$Y_1 = \theta_{01} + \varepsilon_1$$
$$Y_2 = \theta_{02} + \theta_{12} Y_1 + \varepsilon_2$$
$$Y_3 = \theta_{03} + \theta_{13} Y_1 + \theta_{23} Y_2 + \varepsilon_3$$
$$Y_4 = \theta_{04} + \varepsilon_4$$

Rooted paths

A rooted path P from i_1 to j_m is a sequence

$$(i_1, (i_1, j_1), \ldots, (i_k, j_k), (i_{k+1}, j_{k+1}), \ldots, (i_m, j_m)),$$

where $j_k = i_{k+1}$. $\vec{P}_i$ is the set of rooted paths ending in i.

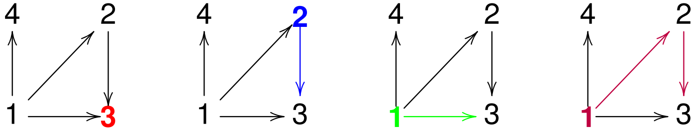
$$\vec{P}_3 = \{(3), (2, (2, 3)), (1, (1, 3)), (1, (1, 2), (2, 3))\}$$

Rooted paths

A rooted path P from i_1 to j_m is a sequence

$$(i_1, (i_1, j_1), \ldots, (i_k, j_k), (i_{k+1}, j_{k+1}), \ldots, (i_m, j_m)),$$

where $j_k = i_{k+1}$. $\vec{P}_i$ is the set of rooted paths ending in i.



$$\vec{P}_3 = \{(3), (2, (2, 3)), (1, (1, 3)), (1, (1, 2), (2, 3))\}$$

Path monomials

Associate

$$i \in V(\mathcal{G}) \longrightarrow \theta'_{0i} = \theta_{0i} + \varepsilon_i, \qquad (j,k) \in E(\mathcal{G}) \longrightarrow \theta_{jk},$$

and, for $P \in \vec{P}_i$, define

$$\boldsymbol{\theta}_P = \prod_{i \in P} \theta_{0i} \prod_{(j,k) \in P} \theta_{jk},$$

as the path monomial.

Proposition

For a linear SEM over a DAG $\mathcal{G}$

$$\mathbb{E}(Y_i \mid \boldsymbol{\theta}, d) = \sum_{P \in \vec{P}_i} \boldsymbol{\theta}_P$$

Multilinear Factorizations

Let

- $\boldsymbol{\theta}_{\vec{P}_i} = \prod_{P \in \vec{P}_i} \boldsymbol{\theta}_P$, $\boldsymbol{\theta}_{Tot} = \prod_{i \in [m]} \boldsymbol{\theta}_{\vec{P}_i}$;
- $\boldsymbol{a}_i = (a_{ij})_{j \in [\# \vec{P}_i]}$, $\boldsymbol{a} = (\boldsymbol{a}_i^{\mathsf{T}})$, $\boldsymbol{r} = (r_i)_{i \in [m]}$;
- $\boldsymbol{r} \simeq \boldsymbol{a}$ if both $|\boldsymbol{a}| = |\boldsymbol{r}|$ and $|\boldsymbol{a}_i| = r_i$

Theorem

Suppose

- *a linear SEM;*
- *a panel separable utility;*
- *u_i is polynomial of degree n_i;*

then

$$\bar{u}(d \mid \theta) = \sum_{\boldsymbol{0} <_{lex} \boldsymbol{r} \leq_{lex} \boldsymbol{n}} c_{\boldsymbol{r}} \sum_{\boldsymbol{a} \simeq \boldsymbol{r}} \binom{|\boldsymbol{r}|}{\boldsymbol{a}} \boldsymbol{\theta}_{Tot}^{\boldsymbol{r}},$$

where $\boldsymbol{n} = (n_i)_{i \in [m]}$, $|\boldsymbol{r}| = \sum_{i \in [m]} r_i$, $c_{\boldsymbol{r}} = k_J \prod_{j \in J} \rho_{j r_j}$ and $J = \{j \in [m] : r_j \neq 0\}$

A graphical interpretation

Note that

$$u(\boldsymbol{y}, d) = \sum_{\mathbf{0} <_{lex} \boldsymbol{r} \leq_{lex} \boldsymbol{n}} c_{\boldsymbol{r}} \boldsymbol{y}^{\boldsymbol{r}}$$

Let

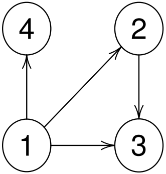
- $\vec{P}_i^j$ be the set of unordered j-tuples of paths ending in i;
- for $P \in \vec{P}_i^j$, n_{P_i} the number of distinct permutations of the elements of P;
- for $\boldsymbol{r} \in \mathbb{Z}_{\geq 0}^m$, $\vec{P}_{\boldsymbol{r}} = \times_{r_i \neq 0} \vec{P}_i^{r_i}$ and $n_P = \sum_{r_i \neq 0} n_{P_i}$,

then

$$\sum_{\boldsymbol{a} \simeq \boldsymbol{r}} \binom{|\boldsymbol{r}|}{\boldsymbol{a}} \boldsymbol{\theta}_{Tot}^{\boldsymbol{r}} = \sum_{P \in \vec{P}_{\boldsymbol{r}}} n_P \prod_{p \in P} \boldsymbol{\theta}_p$$

An example

$$\mathbb{E}(Y_2^2 Y_4^2 \mid \boldsymbol{\theta}) = \theta_{02}'^2\theta_{04}'^2 + 2\theta_{12}\theta_{02}'\theta_{04}'^2 + \theta_{12}^2\theta_{04}'^2 + 2\theta_{02}'^2\theta_{14}\theta_{04}' + 4\theta_{12}\theta_{02}'\theta_{14}\theta_{04} + 2\theta_{12}^2\theta_{14}\theta_{04}' + \theta_{02}'^2\theta_{14}^2 + 2\theta_{12}\theta_{02}'\theta_{14}^2 + \theta_{12}^2\theta_{14}^2.$$



((2), (2), (4), (4))
((1, (1, 2)), (2), (4), (4))
((1, (1, 2)), (1, (1, 2)), (4), (4))
((2), (2), (1, (1, 4)), (4))
((1, (1, 2)), (2), (1, (1, 4)), (4))
((1, (1, 2)), (1, (1, 2)), (1, (1, 4)), (4))
((2), (2), (1, (1, 4)), (1, (1, 4)))
((1, (1, 2)), (2), (1, (1, 4)), (1, (1, 4)))
((1, (1, 2)), (1, (1, 2)), (1, (1, 4)), (1, (1, 4)))

Discussion

New application in Bayesian decision making;

Algebra helped us identifying minimal sets of separation conditions for fast multi-expert analyses;

Extensions:

- Bayesian dynamic forecasting;
- Tensor propagation;

Thanks for your attention

Discussion

New application in Bayesian decision making;

Algebra helped us identifying minimal sets of separation conditions for fast multi-expert analyses;

Extensions:

- Bayesian dynamic forecasting;
- Tensor propagation;

Thanks for your attention