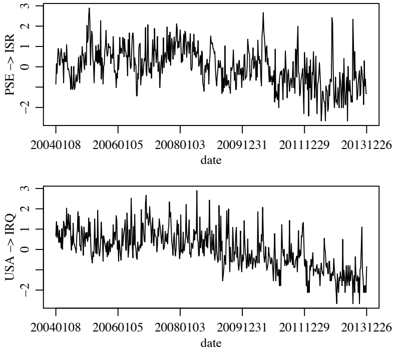
Adaptive Tensor Testing

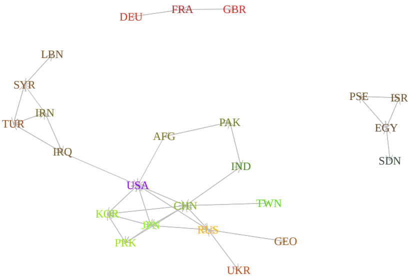
Peter Hoff
Duke University

Supported by the NSF (DMS-1505136)

Outline

ICEWS data





Data: `www.lockheedmartin.com/us/products/W-ICEWS/iData.html`

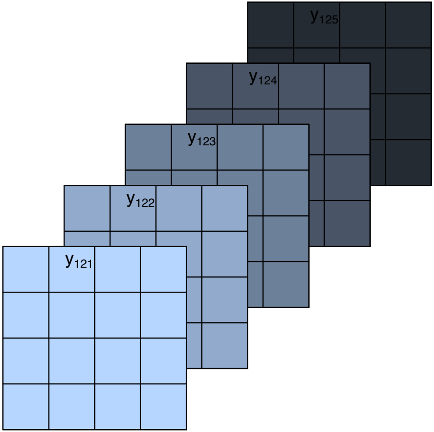
(Thanks to Mike Ward, Duke University, for data access and polysci consulting.)

Dynamic network data

A time series of sociomatrices:

$$\mathbf{Y} = \{\mathbf{Y}_t : t = 1, \ldots, T\}$$

$y_{i,j,t}$ = time-t relation from i to j.

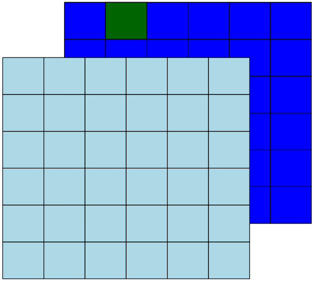


Temporal dependence

How does $y_{i,j,t}$ depend on $\mathbf{Y}_{t-1}$?

Maybe

- $y_{i,j,t-1}$?
- $y_{j,i,t-1}$?
- $y_{i,k,t-1}, y_{k,j,t-1}$?
- $y_{k,i,t-1}, y_{j,k,t-1}$?
- $y_{k,l,t-1}$?

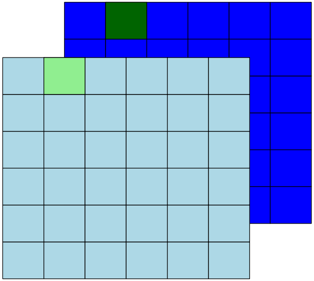


Temporal dependence

How does $y_{i,j,t}$ depend on $\mathbf{Y}_{t-1}$?

Maybe

- $y_{i,j,t-1}$?
- $y_{j,i,t-1}$?
- $y_{i,k,t-1}, y_{k,j,t-1}$?
- $y_{k,i,t-1}, y_{j,k,t-1}$?
- $y_{k,l,t-1}$?

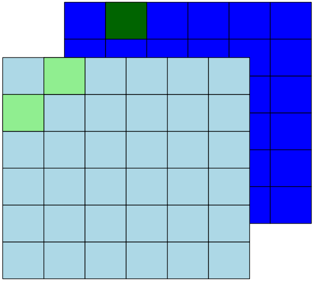


Temporal dependence

How does $y_{i,j,t}$ depend on $\mathbf{Y}_{t-1}$?

Maybe

- $y_{i,j,t-1}$?
- $y_{j,i,t-1}$?
- $y_{i,k,t-1}, y_{k,j,t-1}$?
- $y_{k,i,t-1}, y_{j,k,t-1}$?
- $y_{k,l,t-1}$?

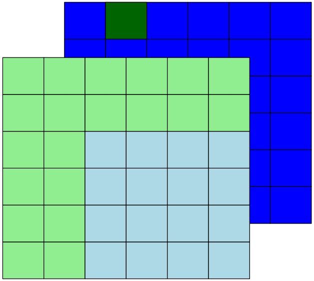


Temporal dependence

How does $y_{i,j,t}$ depend on $\mathbf{Y}_{t-1}$?

Maybe

- $y_{i,j,t-1}$?
- $y_{j,i,t-1}$?
- $y_{i,k,t-1}, y_{k,j,t-1}$?
- $y_{k,i,t-1}, y_{j,k,t-1}$?
- $y_{k,l,t-1}$?

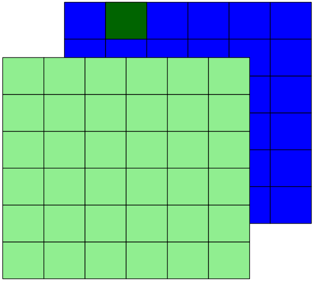


Temporal dependence

How does $y_{i,j,t}$ depend on $\mathbf{Y}_{t-1}$?

Maybe

- $y_{i,j,t-1}$?
- $y_{j,i,t-1}$?
- $y_{i,k,t-1}, y_{k,j,t-1}$?
- $y_{k,i,t-1}, y_{j,k,t-1}$?
- $y_{k,l,t-1}$?

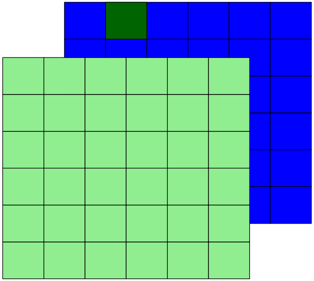


Temporal dependence

How does $y_{i,j,t}$ depend on $\mathbf{Y}_{t-1}$?

Maybe

- $y_{i,j,t-1}$?
- $y_{j,i,t-1}$?
- $y_{i,k,t-1}, y_{k,j,t-1}$?
- $y_{k,i,t-1}, y_{j,k,t-1}$?
- $y_{k,l,t-1}$?

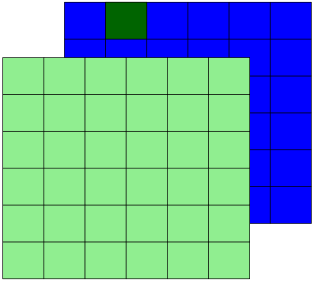


Temporal dependence

How does $y_{i,j,t}$ depend on $\mathbf{Y}_{t-1}$?

Maybe

- $y_{i,j,t-1}$?
- $y_{j,i,t-1}$?
- $y_{i,k,t-1}, y_{k,j,t-1}$?
- $y_{k,i,t-1}, y_{j,k,t-1}$?
- $y_{k,l,t-1}$?



First order VAR model

Let

- $\mathbf{y}_t = \text{vec}(\mathbf{Y}_t)$
- $\mathbf{x}_t = \text{vec}(\mathbf{X}_t)$

VAR: A first-order VAR model posits that

$$\mathbf{y}_t = \Theta \mathbf{x}_t + \mathbf{e}_t \ , \quad \text{E}[\mathbf{e}_t] = \mathbf{0} \ , \quad \text{E}[\mathbf{e}_t \mathbf{e}_s^T] = \begin{cases} \Sigma & \text{if } t = s \\ \mathbf{0} & \text{if } t \neq s, \end{cases}$$

where Θ and Σ are parameters to be estimated.

Is it possible to estimate an effect of each entry of $\mathbf{x}_t$ on each entry of $\mathbf{y}_t$?

First order VAR model

Let

- $\mathbf{y}_t = \text{vec}(\mathbf{Y}_t)$
- $\mathbf{x}_t = \text{vec}(\mathbf{X}_t)$

VAR: A first-order VAR model posits that

$$\mathbf{y}_t = \boldsymbol{\Theta}\mathbf{x}_t + \mathbf{e}_t \ , \quad \text{E}[\mathbf{e}_t] = \mathbf{0} \ , \quad \text{E}[\mathbf{e}_t\mathbf{e}_s^T] = \begin{cases} \boldsymbol{\Sigma} & \text{if } t = s \\ \mathbf{0} & \text{if } t \neq s, \end{cases}$$

where $\boldsymbol{\Theta}$ and $\boldsymbol{\Sigma}$ are parameters to be estimated.

Is it possible to estimate an effect of each entry of $\mathbf{x}_t$ on each entry of $\mathbf{y}_t$?

First order VAR model

Let

- $\mathbf{y}_t = \text{vec}(\mathbf{Y}_t)$
- $\mathbf{x}_t = \text{vec}(\mathbf{X}_t)$

VAR: A first-order VAR model posits that

$$\mathbf{y}_t = \mathbf{\Theta}\mathbf{x}_t + \mathbf{e}_t \ , \quad \text{E}[\mathbf{e}_t] = \mathbf{0} \ , \quad \text{E}[\mathbf{e}_t\mathbf{e}_s^T] = \begin{cases} \Sigma & \text{if } t = s \\ \mathbf{0} & \text{if } t \neq s, \end{cases}$$

where $\mathbf{\Theta}$ and Σ are parameters to be estimated.

Is it possible to estimate an effect of each entry of $\mathbf{x}_t$ on each entry of $\mathbf{y}_t$?

Multilinear models

Assume $\mathbf{\Theta}$ is low-dimensional:

$$\mathbf{Y}_t = \mathbf{A}\mathbf{X}_t\mathbf{B}^\top + \mathbf{E}_t$$
$$\mathbf{y}_t = (\mathbf{B} \otimes \mathbf{A})\mathbf{x}_t + \mathbf{e}_t$$

Extension to higher-dimensional tensors:

$$\mathbf{Y}_t = \mathbf{X}_t \times \{\mathbf{A}, \mathbf{B}, \mathbf{C}\} + \mathbf{E}_t$$
$$\mathbf{y}_t = (\mathbf{C} \otimes \mathbf{B} \otimes \mathbf{A})\mathbf{x}_t + \mathbf{e}_t$$

(Hoff 2015)

Multilinear models

Assume $\mathbf{\Theta}$ is low-dimensional:

$$\mathbf{Y}_t = \mathbf{A}\mathbf{X}_t\mathbf{B}^\top + \mathbf{E}_t$$
$$\mathbf{y}_t = (\mathbf{B} \otimes \mathbf{A})\mathbf{x}_t + \mathbf{e}_t$$

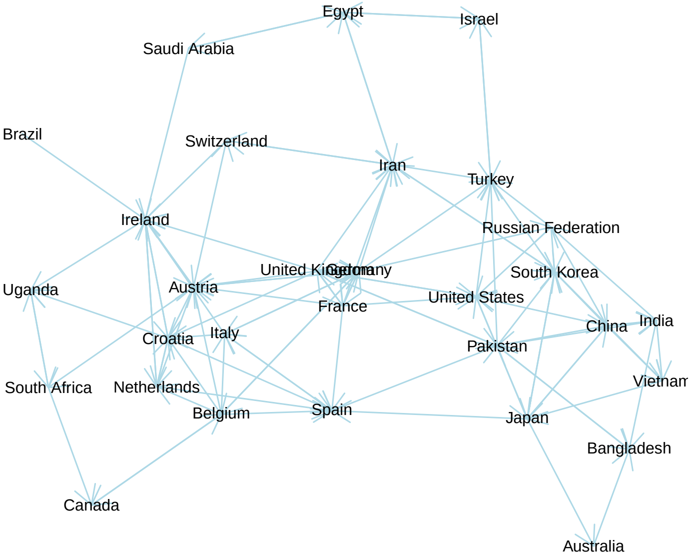
Extension to higher-dimensional tensors:

$$\mathbf{Y}_t = \mathbf{X}_t \times \{\mathbf{A}, \mathbf{B}, \mathbf{C}\} + \mathbf{E}_t$$
$$\mathbf{y}_t = (\mathbf{C} \otimes \mathbf{B} \otimes \mathbf{A})\mathbf{x}_t + \mathbf{e}_t$$

(Hoff 2015)

Social influence model for symmetric relations

$$\mathbf{y}_t = (\mathbf{A} \otimes \mathbf{A})\mathbf{x}_t + \mathbf{e}_t$$



Inference concerns

Restricted model: $\mathbf{y}_t = (\mathbf{A} \otimes \mathbf{A})\mathbf{x}_t + \mathbf{e}_t$

- parameter estimates stable;
- model is wrong;
- CIs wont have proper coverage, hard to interpret.

Full model: $\mathbf{y}_t = \Theta \mathbf{x}_t + \mathbf{e}_t$

- parameter estimates unstable;
- model is less wrong ;
- CIs have proper coverage, but lack precision.

Objective: Obtain proper CIs that adapt to restricted model.

- maintain proper coverage;
- improve precision if model is correct.

Strategy: Use tests with more power in likely parts of parameter space, less power elsewhere.

Inference concerns

Restricted model: $\mathbf{y}_t = (\mathbf{A} \otimes \mathbf{A})\mathbf{x}_t + \mathbf{e}_t$

- parameter estimates stable;
- model is wrong;
- CIs wont have proper coverage, hard to interpret.

Full model: $\mathbf{y}_t = \mathbf{\Theta}\mathbf{x}_t + \mathbf{e}_t$

- parameter estimates unstable;
- model is less wrong ;
- CIs have proper coverage, but lack precision.

Objective: Obtain proper CIs that adapt to restricted model.

- maintain proper coverage;
- improve precision if model is correct.

Strategy: Use tests with more power in likely parts of parameter space, less power elsewhere.

Inference concerns

Restricted model: $\mathbf{y}_t = (\mathbf{A} \otimes \mathbf{A})\mathbf{x}_t + \mathbf{e}_t$

- parameter estimates stable;
- model is wrong;
- CIs wont have proper coverage, hard to interpret.

Full model: $\mathbf{y}_t = \mathbf{\Theta}\mathbf{x}_t + \mathbf{e}_t$

- parameter estimates unstable;
- model is less wrong ;
- CIs have proper coverage, but lack precision.

Objective: Obtain proper CIs that adapt to restricted model.

- maintain proper coverage;
- improve precision if model is correct.

Strategy: Use tests with more power in likely parts of parameter space, less power elsewhere.

Inference concerns

Restricted model: $\mathbf{y}_t = (\mathbf{A} \otimes \mathbf{A})\mathbf{x}_t + \mathbf{e}_t$

- parameter estimates stable;
- model is wrong;
- CIs wont have proper coverage, hard to interpret.

Full model: $\mathbf{y}_t = \mathbf{\Theta}\mathbf{x}_t + \mathbf{e}_t$

- parameter estimates unstable;
- model is less wrong ;
- CIs have proper coverage, but lack precision.

Objective: Obtain proper CIs that adapt to restricted model.

- maintain proper coverage;
- improve precision if model is correct.

Strategy: Use tests with more power in likely parts of parameter space, less power elsewhere.

Tensor data and models

FAB tests and CIs

AFAB tests and CIs

Multilevel data

Linear regression

Tensor parameters

Bayes-optimal confidence interval

Normal model:

- $y \sim N(\theta, \sigma^2)$, σ^2 known.
- $\pi(\theta)$ is prior information about θ.

Goal: Construct confidence interval C having

- Constant coverage: $\Pr(\theta \in C(y)|\theta) = 1 - \alpha \; \forall \theta$.
- Optimal precision: $\int \int |C(y)| \, p(dy|\theta)\pi(d\theta) < \int \int |\tilde{C}(y)| \, p(dy|\theta)\pi(d\theta)$.

The first criterion is "frequentist" - conditional on θ_j.

The second is "Bayesian" - on average across likely values of θ_j.

Bayes-optimal confidence interval

Normal model:

- $y \sim N(\theta, \sigma^2)$, σ^2 known.
- $\pi(\theta)$ is prior information about θ.

Goal: Construct confidence interval C having

- Constant coverage: $\Pr(\theta \in C(y)|\theta) = 1 - \alpha \ \forall \theta$.
- Optimal precision: $\int \int |C(y)| \, p(dy|\theta)\pi(d\theta) < \int \int |\tilde{C}(y)| \, p(dy|\theta)\pi(d\theta)$.

The first criterion is "frequentist" - conditional on θ_j.

The second is "Bayesian" - on average across likely values of θ_j.

Bayes-optimal confidence interval

Normal model:

- $y \sim N(\theta, \sigma^2)$, σ^2 known.
- $\pi(\theta)$ is prior information about θ.

Goal: Construct confidence interval C having

- Constant coverage: $\Pr(\theta \in C(y)|\theta) = 1 - \alpha \ \forall \theta$.
- Optimal precision: $\int \int |C(y)| \, p(dy|\theta)\pi(d\theta) < \int \int |\tilde{C}(y)| \, p(dy|\theta)\pi(d\theta)$.

The first criterion is "frequentist" - conditional on θ_j.

The second is "Bayesian" - on average across likely values of θ_j.

Bayes-optimal confidence interval

Normal model:

- $y \sim N(\theta, \sigma^2)$, σ^2 known.
- $\pi(\theta)$ is prior information about θ.

Goal: Construct confidence interval C having

- Constant coverage: $\Pr(\theta \in C(y)|\theta) = 1 - \alpha \ \forall \theta$.
- Optimal precision: $\int \int |C(y)| \, p(dy|\theta)\pi(d\theta) < \int \int |\tilde{C}(y)| \, p(dy|\theta)\pi(d\theta)$.

The first criterion is "frequentist" - conditional on θ_j.

The second is "Bayesian" - on average across likely values of θ_j.

Bayes-optimal confidence interval

Normal model:

- $y \sim N(\theta, \sigma^2)$, σ^2 known.
- $\pi(\theta)$ is prior information about θ.

Goal: Construct confidence interval C having

- Constant coverage: $\Pr(\theta \in C(y)|\theta) = 1 - \alpha \ \forall \theta$.
- Optimal precision: $\int \int |C(y)| \, p(dy|\theta)\pi(d\theta) < \int \int |\tilde{C}(y)| \, p(dy|\theta)\pi(d\theta)$.

The first criterion is "frequentist" - conditional on θ_j.

The second is "Bayesian" - on average across likely values of θ_j.

F.A.B : Frequentist, Assisted by Bayes.

All CIs

UMAU procedure:

$$C_{1/2}(y) = \{\theta : y + \sigma z_{\alpha/2} < \theta < y + \sigma z_{1-\alpha/2}\}$$

Any procedure:

$$C_s(y) = \{\theta : y + \sigma z_{\alpha(1-s)} < \theta < y + \sigma z_{1-\alpha s}\}$$

In fact, s may depend on θ: If $s : \mathbb{R} \to [0, 1]$ then

$A_s(\theta) = \{y : y + \sigma z_{\alpha(1-s(\theta))} < \theta < y + \sigma z_{1-\alpha s(\theta)}\}$ is a class of level-α tests;
$C_s(y) = \{\theta : y + \sigma z_{\alpha(1-s(\theta))} < \theta < y + \sigma z_{1-\alpha s(\theta)}\}$ is a $1 - \alpha$ CR.

$\Pr(\theta \in C_s(y)|\theta) = 1 - \alpha.$

All CIs

UMAU procedure:

$$C_{1/2}(y) = \{\theta : y + \sigma z_{\alpha/2} < \theta < y + \sigma z_{1-\alpha/2}\}$$

Any procedure:

$$C_s(y) = \{\theta : y + \sigma z_{\alpha(1-s)} < \theta < y + \sigma z_{1-\alpha s}\}$$

In fact, s may depend on θ: If $s : \mathbb{R} \to [0,1]$ then

$A_s(\theta) = \{y : y + \sigma z_{\alpha(1-s(\theta))} < \theta < y + \sigma z_{1-\alpha s(\theta)}\}$ is a class of level-α tests;
$C_s(y) = \{\theta : y + \sigma z_{\alpha(1-s(\theta))} < \theta < y + \sigma z_{1-\alpha s(\theta)}\}$ is a $1-\alpha$ CR.

$$\Pr(\theta \in C_s(y)|\theta) = 1 - \alpha.$$

All CIs

UMAU procedure:

$$C_{1/2}(y) = \{\theta : y + \sigma z_{\alpha/2} < \theta < y + \sigma z_{1-\alpha/2}\}$$

Any procedure:

$$C_s(y) = \{\theta : y + \sigma z_{\alpha(1-s)} < \theta < y + \sigma z_{1-\alpha s}\}$$

In fact, s may depend on θ: If $s : \mathbb{R} \to [0,1]$ then

$A_s(\theta) = \{y : y + \sigma z_{\alpha(1-s(\theta))} < \theta < y + \sigma z_{1-\alpha s(\theta)}\}$ is a class of level-α tests;

$C_s(y) = \{\theta : y + \sigma z_{\alpha(1-s(\theta))} < \theta < y + \sigma z_{1-\alpha s(\theta)}\}$ is a $1-\alpha$ CR.

$\Pr(\theta \in C_s(y)|\theta) = 1 - \alpha.$

All CIs

UMAU procedure:

$$C_{1/2}(y) = \{\theta : y + \sigma z_{\alpha/2} < \theta < y + \sigma z_{1-\alpha/2}\}$$

Any procedure:

$$C_s(y) = \{\theta : y + \sigma z_{\alpha(1-s)} < \theta < y + \sigma z_{1-\alpha s}\}$$

In fact, s may depend on θ: If $s : \mathbb{R} \to [0, 1]$ then

$A_s(\theta) = \{y : y + \sigma z_{\alpha(1-s(\theta))} < \theta < y + \sigma z_{1-\alpha s(\theta)}\}$ is a class of level-α tests;
$C_s(y) = \{\theta : y + \sigma z_{\alpha(1-s(\theta))} < \theta < y + \sigma z_{1-\alpha s(\theta)}\}$ is a $1 - \alpha$ CR.

$\Pr(\theta \in C_s(y)|\theta) = 1 - \alpha.$

All CIs

UMAU procedure:

$$C_{1/2}(y) = \{\theta : y + \sigma z_{\alpha/2} < \theta < y + \sigma z_{1-\alpha/2}\}$$

Any procedure:

$$C_s(y) = \{\theta : y + \sigma z_{\alpha(1-s)} < \theta < y + \sigma z_{1-\alpha s}\}$$

In fact, s may depend on θ: If $s : \mathbb{R} \to [0, 1]$ then

$A_s(\theta) = \{y : y + \sigma z_{\alpha(1-s(\theta))} < \theta < y + \sigma z_{1-\alpha s(\theta)}\}$ is a class of level-α tests;
$C_s(y) = \{\theta : y + \sigma z_{\alpha(1-s(\theta))} < \theta < y + \sigma z_{1-\alpha s(\theta)}\}$ is a $1 - \alpha$ CR.

$$\Pr(\theta \in C_s(y)|\theta) = 1 - \alpha.$$

Properties of C_s

1. $\Pr(\theta \in C_s(y)|\theta) = 1 - \alpha$.
2. $C_s(y)$ is an interval if $s(\theta)$ is monotonic (not necessary).
3. In the monotonic case, the interval can be found by solving

$$\theta^L = y + \sigma z_{\alpha(1-s(\theta^L))}$$
$$\theta^U = y + \sigma z_{1-\alpha s(\theta^U)}$$

noting that $\theta^L < y + \sigma z_\alpha$ and $y + \sigma z_{1-\alpha} < \theta^U$.

Properties of C_s

1. $\Pr(\theta \in C_s(y)|\theta) = 1 - \alpha$.
2. $C_s(y)$ is an interval if $s(\theta)$ is monotonic (not necessary).
3. In the monotonic case, the interval can be found by solving

$$\theta^L = y + \sigma z_{\alpha(1-s(\theta^L))}$$
$$\theta^U = y + \sigma z_{1-\alpha s(\theta^U)}$$

noting that $\theta^L < y + \sigma z_\alpha$ and $y + \sigma z_{1-\alpha} < \theta^U$.

Properties of C_s

1. $\Pr(\theta \in C_s(y)|\theta) = 1 - \alpha$.
2. $C_s(y)$ is an interval if $s(\theta)$ is monotonic (not necessary).
3. In the monotonic case, the interval can be found by solving

$$\theta^L = y + \sigma z_{\alpha(1-s(\theta^L))}$$
$$\theta^U = y + \sigma z_{1-\alpha s(\theta^U)}$$

noting that $\theta^L < y + \sigma z_\alpha$ and $y + \sigma z_{1-\alpha} < \theta^U$.

Properties of C_s

1. $\Pr(\theta \in C_s(y)|\theta) = 1 - \alpha$.
2. $C_s(y)$ is an interval if $s(\theta)$ is monotonic (not necessary).
3. In the monotonic case, the interval can be found by solving

$$\theta^L = y + \sigma z_{\alpha(1-s(\theta^L))}$$
$$\theta^U = y + \sigma z_{1-\alpha s(\theta^U)}$$

noting that $\theta^L < y + \sigma z_\alpha$ and $y + \sigma z_{1-\alpha} < \theta^U$.

Properties of C_s

1. $\Pr(\theta \in C_s(y)|\theta) = 1 - \alpha$.
2. $C_s(y)$ is an interval if $s(\theta)$ is monotonic (not necessary).
3. In the monotonic case, the interval can be found by solving

$$\theta^L = y + \sigma z_{\alpha(1-s(\theta^L))}$$
$$\theta^U = y + \sigma z_{1-\alpha s(\theta^U)}$$

noting that $\theta^L < y + \sigma z_\alpha$ and $y + \sigma z_{1-\alpha} < \theta^U$.

Optimal s-function

If $\pi(\theta)$ is the $N(\mu, \tau^2)$ density, then

$$R(\pi, C_s) = \int \int |C_s(y)|\, p(dy|\theta)\, \pi(d\theta)$$

is minimized by

$$s(\theta) = g^{-1}(2\sigma(\theta - \mu)/\tau^2)$$
$$g(s) = \Phi^{-1}(\alpha s) - \Phi^{-1}(\alpha(1 - s))$$

This s-function yields Pratt's (1963) z-interval.

Optimal s-function

If $\pi(\theta)$ is the $N(\mu, \tau^2)$ density, then

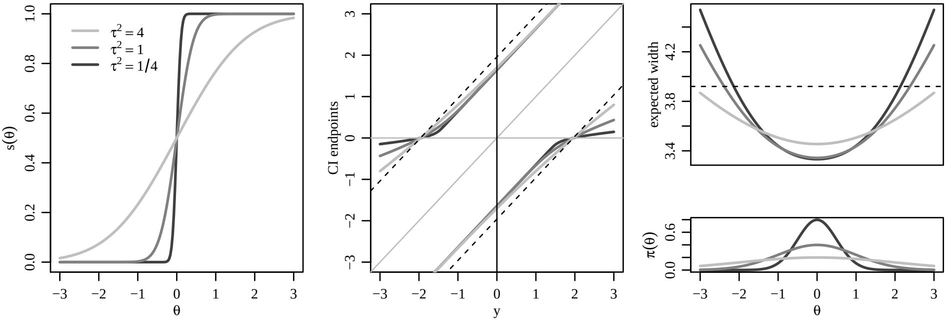
$$R(\pi, C_s) = \int\int |C_s(y)|\, p(dy|\theta)\, \pi(d\theta)$$

is minimized by

$$s(\theta) = g^{-1}(2\sigma(\theta - \mu)/\tau^2)$$
$$g(s) = \Phi^{-1}(\alpha s) - \Phi^{-1}(\alpha(1 - s))$$

This s-function yields Pratt's (1963) z-interval.

Bayes-optimal frequentist procedure



Limitations of FAB z-interval

This idea has not been widely adopted:

- Pratt's original derivation hard to generalize for unknown variance.
- Non-Bayesians uneasy with specifying a prior distribution.
- Bayesians less concerned with frequentist coverage.

We address some of these issues:

- Extend z-interval to t-interval using s-functions.
- Apply method in multiparameter problems, sharing information.
  - "prior" is determined by data;
  - exact frequentist coverage is maintained.

Limitations of FAB z-interval

This idea has not been widely adopted:

- Pratt's original derivation hard to generalize for unknown variance.
- Non-Bayesians uneasy with specifying a prior distribution.
- Bayesians less concerned with frequentist coverage.

We address some of these issues:

- Extend z-interval to t-interval using s-functions.
- Apply method in multiparameter problems, sharing information.
    - "prior" is determined by data;
    - exact frequentist coverage is maintained.

Tensor data and models

FAB tests and CIs

AFAB tests and CIs

Multilevel data

Linear regression

Tensor parameters

FAB t-intervals

Data:

$$\hat{\theta} \sim N(\theta, h^2\sigma^2) \ , \quad q\hat{\sigma}^2/\sigma^2 \sim \chi^2_q$$

h^2 known, $\hat{\theta}$ and $\hat{\sigma}^2$ independent.

UMAU CI:

$$C_{1/2}(\hat{\theta}, \hat{\sigma}) = \{\theta : \hat{\theta} + h\hat{\sigma}t_{\alpha/2} < \theta < \hat{\theta} + h\hat{\sigma}t_{1-\alpha/2}\}$$

All CIs:

$$C_s(\hat{\theta}, \hat{\sigma}) = \{\theta : \hat{\theta} + h\hat{\sigma}t_{\alpha(1-s(\theta))} < \theta < \hat{\theta} + h\hat{\sigma}t_{1-\alpha s(\theta)}\}$$

FAB t-intervals

Data:

$$\hat{\theta} \sim N(\theta, h^2\sigma^2) \; , \quad q\hat{\sigma}^2/\sigma^2 \sim \chi^2_q$$

h^2 known, $\hat{\theta}$ and $\hat{\sigma}^2$ independent.

UMAU CI:

$$C_{1/2}(\hat{\theta}, \hat{\sigma}) = \{\theta : \hat{\theta} + h\hat{\sigma} t_{\alpha/2} < \theta < \hat{\theta} + h\hat{\sigma} t_{1-\alpha/2}\}$$

All CIs:

$$C_s(\hat{\theta}, \hat{\sigma}) = \{\theta : \hat{\theta} + h\hat{\sigma} t_{\alpha(1-s(\theta))} < \theta < \hat{\theta} + h\hat{\sigma} t_{1-\alpha s(\theta)}\}$$

FAB t-intervals

Data:

$$\hat{\theta} \sim N(\theta, h^2\sigma^2) \;, \quad q\hat{\sigma}^2/\sigma^2 \sim \chi^2_q$$

h^2 known, $\hat{\theta}$ and $\hat{\sigma}^2$ independent.

UMAU CI:

$$C_{1/2}(\hat{\theta}, \hat{\sigma}) = \{\theta : \hat{\theta} + h\hat{\sigma} t_{\alpha/2} < \theta < \hat{\theta} + h\hat{\sigma} t_{1-\alpha/2}\}$$

All CIs:

$$C_s(\hat{\theta}, \hat{\sigma}) = \{\theta : \hat{\theta} + h\hat{\sigma} t_{\alpha(1-s(\theta))} < \theta < \hat{\theta} + h\hat{\sigma} t_{1-\alpha s(\theta)}\}$$

FAB t-intervals

$$C_s(\hat{\theta}, \hat{\sigma}) = \{\theta : \hat{\theta} + h\hat{\sigma} t_{\alpha(1-s(\theta))} < \theta < \hat{\theta} + h\hat{\sigma} t_{1-\alpha s(\theta)}\}$$

Thm: If $\hat{\theta} \sim N(\theta, h^2\sigma^2)$ and $q\hat{\sigma}^2/\sigma^2 \sim \chi^2_q$ are independent, then

1. $\Pr(\theta \in C_s(\hat{\theta}, \hat{\sigma}) | \theta, \sigma) = 1 - \alpha \; \forall (\theta, \sigma) \in \mathbb{R} \times \mathbb{R}^+$;
2. If $s(\theta)$ nondecreasing, $C_s(\hat{\theta}, \hat{\sigma})$ is an interval with probability 1;
3. If $s(\theta)$ not nondecreasing, $C_s(\hat{\theta}, \hat{\sigma})$ is an interval with probability < 1.

FAB t-intervals

$$C_s(\hat{\theta}, \hat{\sigma}) = \{\theta : \hat{\theta} + h\hat{\sigma} t_{\alpha(1-s(\theta))} < \theta < \hat{\theta} + h\hat{\sigma} t_{1-\alpha s(\theta)}\}$$

Thm: If $\hat{\theta} \sim N(\theta, h^2\sigma^2)$ and $q\hat{\sigma}^2/\sigma^2 \sim \chi^2_q$ are independent, then

1. $\Pr(\theta \in C_s(\hat{\theta}, \hat{\sigma}) | \theta, \sigma) = 1 - \alpha \, \forall \, (\theta, \sigma) \in \mathbb{R} \times \mathbb{R}^+$;
2. If $s(\theta)$ nondecreasing, $C_s(\hat{\theta}, \hat{\sigma})$ is an interval with probability 1;
3. If $s(\theta)$ not nondecreasing, $C_s(\hat{\theta}, \hat{\sigma})$ is an interval with probability < 1.

FAB t-intervals

$$C_s(\hat{\theta}, \hat{\sigma}) = \{\theta : \hat{\theta} + h\hat{\sigma} t_{\alpha(1-s(\theta))} < \theta < \hat{\theta} + h\hat{\sigma} t_{1-\alpha s(\theta)}\}$$

Thm: If $\hat{\theta} \sim N(\theta, h^2\sigma^2)$ and $q\hat{\sigma}^2/\sigma^2 \sim \chi^2_q$ are independent, then

1. $\Pr(\theta \in C_s(\hat{\theta}, \hat{\sigma})|\theta, \sigma) = 1 - \alpha \, \forall \, (\theta, \sigma) \in \mathbb{R} \times \mathbb{R}^+$;
2. If $s(\theta)$ nondecreasing, $C_s(\hat{\theta}, \hat{\sigma})$ is an interval with probability 1;
3. If $s(\theta)$ not nondecreasing, $C_s(\hat{\theta}, \hat{\sigma})$ is an interval with probability < 1.

FAB t-intervals

$$C_s(\hat{\theta},\hat{\sigma}) = \{\theta : \hat{\theta} + h\hat{\sigma} t_{\alpha(1-s(\theta))} < \theta < \hat{\theta} + h\hat{\sigma} t_{1-\alpha s(\theta)}\}$$

Thm: If $\hat{\theta} \sim N(\theta, h^2\sigma^2)$ and $q\hat{\sigma}^2/\sigma^2 \sim \chi^2_q$ are independent, then

1. $\Pr(\theta \in C_s(\hat{\theta},\hat{\sigma})|\theta,\sigma) = 1 - \alpha \,\forall\, (\theta,\sigma) \in \mathbb{R} \times \mathbb{R}^+$;
2. If $s(\theta)$ nondecreasing, $C_s(\hat{\theta},\hat{\sigma})$ is an interval with probability 1;
3. If $s(\theta)$ not nondecreasing, $C_s(\hat{\theta},\hat{\sigma})$ is an interval with probability < 1.

Optimal and suboptimal spending functions

Prior information:
$\pi(\theta, \sigma^2)$, $\mathsf{E}[(\theta, \sigma^2)] = (\mu, \sigma_0^2)$, $\mathsf{V}[\theta] = \tau^2$.

Principled procedure:
Find Bayes optimal s-function under this prior (numerically).

Simpler procedure:
Use Bayes optimal s-function from z-interval, with $\theta \sim N(\mu, \tau^2)$, $\sigma^2 = \sigma_0^2$.

Optimal and suboptimal spending functions

Prior information:
$\pi(\theta, \sigma^2)$, $\mathsf{E}[(\theta, \sigma^2)] = (\mu, \sigma_0^2)$, $\mathsf{V}[\theta] = \tau^2$.

Principled procedure:
Find Bayes optimal *s*-function under this prior (numerically).

Simpler procedure:
Use Bayes optimal *s*-function from *z*-interval, with $\theta \sim N(\mu, \tau^2)$, $\sigma^2 = \sigma_0^2$.

Optimal and suboptimal spending functions

Prior information:
$\pi(\theta, \sigma^2)$, $\mathsf{E}[(\theta, \sigma^2)] = (\mu, \sigma_0^2)$, $\mathsf{V}[\theta] = \tau^2$.

Principled procedure:
Find Bayes optimal s-function under this prior (numerically).

Simpler procedure:
Use Bayes optimal s-function from z-interval, with $\theta \sim N(\mu, \tau^2)$, $\sigma^2 = \sigma_0^2$.

AFAB t-intervals

Lacking a prior, can we obtain $s(\theta)$ from data?

Cor: If

- $\hat{\theta} \sim N(\theta, h^2\sigma^2)$,
- $q\hat{\sigma}^2/\sigma^2 \sim \chi^2_q$,
- $\hat{\theta}$, $\hat{\sigma}^2$ and $\tilde{s}$ are independent, then

$$\Pr(\theta \in C_{\tilde{s}}(\hat{\theta}, \hat{\sigma})|\theta, \sigma) = 1 - \alpha.$$

AFAB procedure:

- $\tilde{s}(\theta) = g^{-1}(2\tilde{\sigma}(\theta - \tilde{\mu})/\tilde{\tau}^2)$;
- $\tilde{\gamma} = (\tilde{\theta}, \tilde{\tau}^2, \tilde{\sigma}^2)$ estimated from data;
- Constant coverage assured if $\tilde{\gamma}$ independent of $\hat{\theta}, \hat{\sigma}^2$.

AFAB t-intervals

Lacking a prior, can we obtain $s(\theta)$ from data?

Cor: If

- $\hat{\theta} \sim N(\theta, h^2\sigma^2)$,
- $q\hat{\sigma}^2/\sigma^2 \sim \chi^2_q$,
- $\hat{\theta}$, $\hat{\sigma}^2$ and $\tilde{s}$ are independent, then

$$\Pr(\theta \in C_{\tilde{s}}(\hat{\theta}, \hat{\sigma})|\theta, \sigma) = 1 - \alpha.$$

AFAB procedure:

- $\tilde{s}(\theta) = g^{-1}(2\tilde{\sigma}(\theta - \tilde{\mu})/\tilde{\tau}^2)$;
- $\tilde{\gamma} = (\tilde{\theta}, \tilde{\tau}^2, \tilde{\sigma}^2)$ estimated from data;
- Constant coverage assured if $\tilde{\gamma}$ independent of $\hat{\theta}, \hat{\sigma}^2$.

AFAB t-intervals

Lacking a prior, can we obtain $s(\theta)$ from data?

Cor: If

- $\hat{\theta} \sim N(\theta, h^2\sigma^2)$,
- $q\hat{\sigma}^2/\sigma^2 \sim \chi^2_q$,
- $\hat{\theta}$, $\hat{\sigma}^2$ and $\tilde{s}$ are independent, then

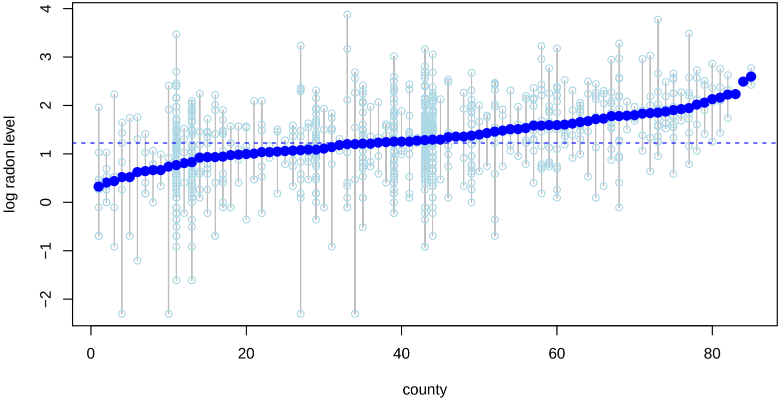
$$\Pr(\theta \in C_{\tilde{s}}(\hat{\theta}, \hat{\sigma})|\theta, \sigma) = 1 - \alpha.$$

AFAB procedure:

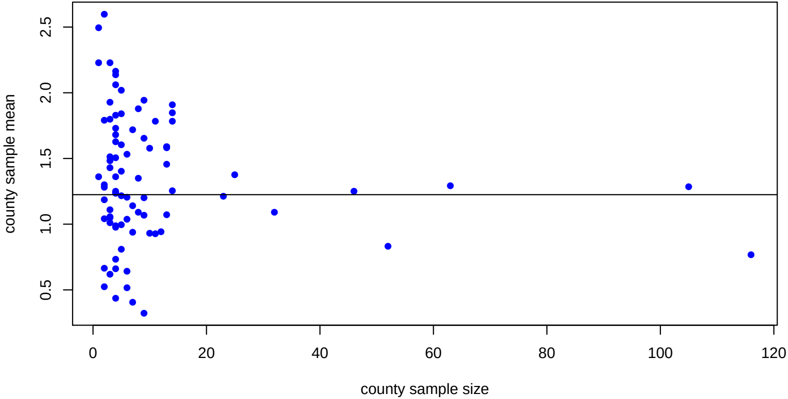
- $\tilde{s}(\theta) = g^{-1}(2\tilde{\sigma}(\theta - \tilde{\mu})/\tilde{\tau}^2)$;
- $\tilde{\gamma} = (\tilde{\theta}, \tilde{\tau}^2, \tilde{\sigma}^2)$ estimated from data;
- Constant coverage assured if $\tilde{\gamma}$ independent of $\hat{\theta}, \hat{\sigma}^2$.

Radon data

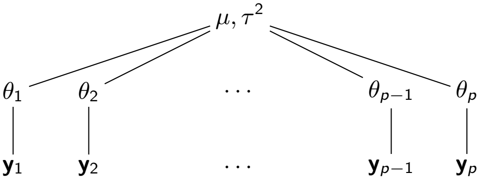
Log radon levels of 85 Minnesota counties.



Varying variability

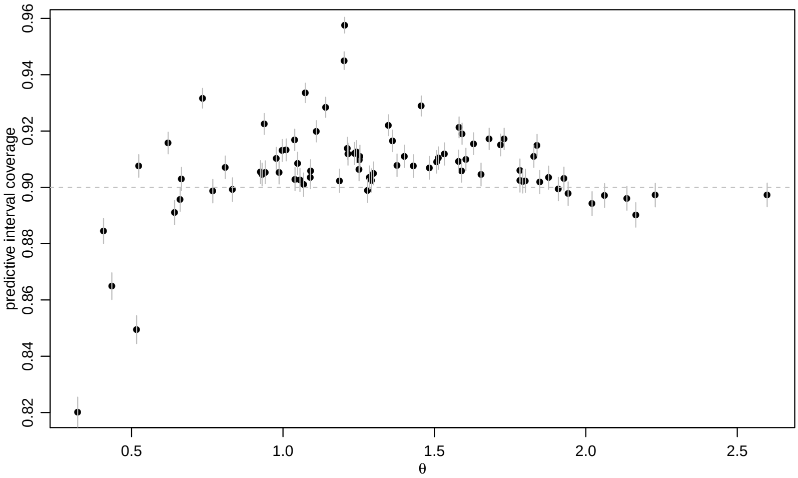


Hierarchical normal model:



$$y_{1,j}, \ldots, y_{n_j,j} \sim \text{i.i.d. } N(\theta_j, \sigma^2)$$
$$\theta_1, \ldots, \theta_p \sim \text{i.i.d. } N(\mu, \tau^2)$$

Nonconstant credible interval coverage



$$\Pr(\theta_j \in C(\hat{\theta}_j)) \approx 1 - \alpha$$

$$\Pr(\theta_j \in C(\hat{\theta}_j) | \theta_j) \quad \text{depends on } \theta_j.$$

Empirical Bayes prior

$$y_{1,j}, \dots, y_{n_j,j} \sim \text{i.i.d. } N(\theta_j, \sigma^2)$$
$$\theta_1, \dots, \theta_p \sim \text{i.i.d. } N(\mu, \tau^2)$$

FAB CIP: $C^j_{\tilde{s}}(\mathbf{y}) = \{\theta : \bar{y}_j + n_j^{-1/2}\hat{\sigma} t_{\alpha(1-\tilde{s}(\theta))} < \theta < \bar{y}_j + n_j^{-1/2}\hat{\sigma} t_{1-\alpha\tilde{s}(\theta)}\}$

Has $1-\alpha$ constant coverage if

- $q\hat{\sigma}^2/\sigma^2 \sim \chi^2_q$, independent of $\bar{y}_j$;
- $\tilde{s}$-function independent of $(\bar{y}_j, \hat{\sigma}^2)$.

Adaptive inference: Obtain $\tilde{s}$ from groups other than j

- Group sample variances provide independent estimates of σ^2: $\hat{\sigma}^2$, $\tilde{\sigma}^2$;
- Group sample means provide $\tilde{\mu}$, $\tilde{\tau}^2$;
- $(\tilde{\mu}, \tilde{\tau}^2, \tilde{\sigma}^2)$ are independent of $(\bar{y}_j, \hat{\sigma}^2)$.
- $\tilde{s}(\theta) = g^{-1}(2\tilde{\sigma}(\theta - \tilde{\mu})/\tilde{\tau}^2)$ approximately Bayes-optimal.

Empirical Bayes prior

$$y_{1,j}, \dots, y_{n_j,j} \sim \text{i.i.d. } N(\theta_j, \sigma^2)$$
$$\theta_1, \dots, \theta_p \sim \text{i.i.d. } N(\mu, \tau^2)$$

FAB CIP: $C^j_{\tilde{s}}(\mathbf{y}) = \{\theta : \bar{y}_j + n_j^{-1/2}\hat{\sigma} t_{\alpha(1-\tilde{s}(\theta))} < \theta < \bar{y}_j + n_j^{-1/2}\hat{\sigma} t_{1-\alpha\tilde{s}(\theta)}\}$

Has $1-\alpha$ constant coverage if

- $q\hat{\sigma}^2/\sigma^2 \sim \chi^2_q$, independent of $\bar{y}_j$;
- $\tilde{s}$-function independent of $(\bar{y}_j, \hat{\sigma}^2)$.

Adaptive inference: Obtain $\tilde{s}$ from groups other than j

- Group sample variances provide independent estimates of σ^2: $\hat{\sigma}^2$, $\tilde{\sigma}^2$;
- Group sample means provide $\tilde{\mu}$, $\tilde{\tau}^2$;
- $(\tilde{\mu}, \tilde{\tau}^2, \tilde{\sigma}^2)$ are independent of $(\bar{y}_j, \hat{\sigma}^2)$.
- $\tilde{s}(\theta) = g^{-1}(2\tilde{\sigma}(\theta - \tilde{\mu})/\tilde{\tau}^2)$ approximately Bayes-optimal.

Empirical Bayes prior

$$y_{1,j}, \ldots, y_{n_j,j} \sim \text{i.i.d. } N(\theta_j, \sigma^2)$$
$$\theta_1, \ldots, \theta_p \sim \text{i.i.d. } N(\mu, \tau^2)$$

FAB CIP: $C^j_{\tilde{s}}(\mathbf{y}) = \{\theta : \bar{y}_j + n_j^{-1/2}\hat{\sigma} t_{\alpha(1-\tilde{s}(\theta))} < \theta < \bar{y}_j + n_j^{-1/2}\hat{\sigma} t_{1-\alpha\tilde{s}(\theta)}\}$

Has $1 - \alpha$ constant coverage if

- $q\hat{\sigma}^2/\sigma^2 \sim \chi^2_q$, independent of $\bar{y}_j$;
- $\tilde{s}$-function independent of $(\bar{y}_j, \hat{\sigma}^2)$.

Adaptive inference: Obtain $\tilde{s}$ from groups other than j

- Group sample variances provide independent estimates of σ^2: $\hat{\sigma}^2$, $\tilde{\sigma}^2$;
- Group sample means provide $\tilde{\mu}$, $\tilde{\tau}^2$;
- $(\tilde{\mu}, \tilde{\tau}^2, \tilde{\sigma}^2)$ are independent of $(\bar{y}_j, \hat{\sigma}^2)$.
- $\tilde{s}(\theta) = g^{-1}(2\tilde{\sigma}(\theta - \tilde{\mu})/\tilde{\tau}^2)$ approximately Bayes-optimal.

Empirical Bayes prior

$$y_{1,j}, \ldots, y_{n_j,j} \sim \text{i.i.d. } N(\theta_j, \sigma^2)$$
$$\theta_1, \ldots, \theta_p \sim \text{i.i.d. } N(\mu, \tau^2)$$

FAB CIP: $C^j_{\tilde{s}}(\mathbf{y}) = \{\theta : \bar{y}_j + n_j^{-1/2}\hat{\sigma} t_{\alpha(1-\tilde{s}(\theta))} < \theta < \bar{y}_j + n_j^{-1/2}\hat{\sigma} t_{1-\alpha\tilde{s}(\theta)}\}$

Has $1-\alpha$ constant coverage if

- $q\hat{\sigma}^2/\sigma^2 \sim \chi^2_q$, independent of $\bar{y}_j$;
- $\tilde{s}$-function independent of $(\bar{y}_j, \hat{\sigma}^2)$.

Adaptive inference: Obtain $\tilde{s}$ from groups other than j

- Group sample variances provide independent estimates of σ^2: $\hat{\sigma}^2$, $\tilde{\sigma}^2$;
- Group sample means provide $\tilde{\mu}$, $\tilde{\tau}^2$;
- $(\tilde{\mu}, \tilde{\tau}^2, \tilde{\sigma}^2)$ are independent of $(\bar{y}_j, \hat{\sigma}^2)$.
- $\tilde{s}(\theta) = g^{-1}(2\tilde{\sigma}(\theta - \tilde{\mu})/\tilde{\tau}^2)$ approximately Bayes-optimal.

Empirical Bayes prior

$$y_{1,j}, \dots, y_{n_j,j} \sim \text{i.i.d. } N(\theta_j, \sigma^2)$$
$$\theta_1, \dots, \theta_p \sim \text{i.i.d. } N(\mu, \tau^2)$$

FAB CIP: $C^j_{\tilde{s}}(\mathbf{y}) = \{\theta : \bar{y}_j + n_j^{-1/2}\hat{\sigma} t_{\alpha(1-\tilde{s}(\theta))} < \theta < \bar{y}_j + n_j^{-1/2}\hat{\sigma} t_{1-\alpha\tilde{s}(\theta)}\}$

Has $1-\alpha$ constant coverage if

- $q\hat{\sigma}^2/\sigma^2 \sim \chi^2_q$, independent of $\bar{y}_j$;
- $\tilde{s}$-function independent of $(\bar{y}_j, \hat{\sigma}^2)$.

Adaptive inference: Obtain $\tilde{s}$ from groups other than j

- Group sample variances provide independent estimates of σ^2: $\hat{\sigma}^2$, $\tilde{\sigma}^2$;
- Group sample means provide $\tilde{\mu}$, $\tilde{\tau}^2$;
- $(\tilde{\mu}, \tilde{\tau}^2, \tilde{\sigma}^2)$ are independent of $(\bar{y}_j, \hat{\sigma}^2)$.
- $\tilde{s}(\theta) = g^{-1}(2\tilde{\sigma}(\theta - \tilde{\mu})/\tilde{\tau}^2)$ approximately Bayes-optimal.

Empirical Bayes prior

$$y_{1,j}, \ldots, y_{n_j,j} \sim \text{i.i.d. } N(\theta_j, \sigma^2)$$
$$\theta_1, \ldots, \theta_p \sim \text{i.i.d. } N(\mu, \tau^2)$$

FAB CIP: $C^j_{\tilde{s}}(\mathbf{y}) = \{\theta : \bar{y}_j + n_j^{-1/2}\hat{\sigma} t_{\alpha(1-\tilde{s}(\theta))} < \theta < \bar{y}_j + n_j^{-1/2}\hat{\sigma} t_{1-\alpha\tilde{s}(\theta)}\}$

Has $1-\alpha$ constant coverage if

- $q\hat{\sigma}^2/\sigma^2 \sim \chi^2_q$, independent of $\bar{y}_j$;
- $\tilde{s}$-function independent of $(\bar{y}_j, \hat{\sigma}^2)$.

Adaptive inference: Obtain $\tilde{s}$ from groups other than j

- Group sample variances provide independent estimates of σ^2: $\hat{\sigma}^2$, $\tilde{\sigma}^2$;
- Group sample means provide $\tilde{\mu}$, $\tilde{\tau}^2$;
- $(\tilde{\mu}, \tilde{\tau}^2, \tilde{\sigma}^2)$ are independent of $(\bar{y}_j, \hat{\sigma}^2)$.
- $\tilde{s}(\theta) = g^{-1}(2\tilde{\sigma}(\theta - \tilde{\mu})/\tilde{\tau}^2)$ approximately Bayes-optimal.

Empirical Bayes prior

$$y_{1,j}, \ldots, y_{n_j,j} \sim \text{i.i.d. } N(\theta_j, \sigma^2)$$
$$\theta_1, \ldots, \theta_p \sim \text{i.i.d. } N(\mu, \tau^2)$$

FAB CIP: $C^j_{\tilde{s}}(\mathbf{y}) = \{\theta : \bar{y}_j + n_j^{-1/2}\hat{\sigma} t_{\alpha(1-\tilde{s}(\theta))} < \theta < \bar{y}_j + n_j^{-1/2}\hat{\sigma} t_{1-\alpha\tilde{s}(\theta)}\}$

Has $1-\alpha$ constant coverage if

- $q\hat{\sigma}^2/\sigma^2 \sim \chi^2_q$, independent of $\bar{y}_j$;
- $\tilde{s}$-function independent of $(\bar{y}_j, \hat{\sigma}^2)$.

Adaptive inference: Obtain $\tilde{s}$ from groups other than j

- Group sample variances provide independent estimates of σ^2: $\hat{\sigma}^2$, $\tilde{\sigma}^2$;
- Group sample means provide $\tilde{\mu}$, $\tilde{\tau}^2$;
- $(\tilde{\mu}, \tilde{\tau}^2, \tilde{\sigma}^2)$ are independent of $(\bar{y}_j, \hat{\sigma}^2)$.
- $\tilde{s}(\theta) = g^{-1}(2\tilde{\sigma}(\theta - \tilde{\mu})/\tilde{\tau}^2)$ approximately Bayes-optimal.

Empirical Bayes prior

$$y_{1,j}, \dots, y_{n_j,j} \sim \text{i.i.d. } N(\theta_j, \sigma^2)$$
$$\theta_1, \dots, \theta_p \sim \text{i.i.d. } N(\mu, \tau^2)$$

FAB CIP: $C^j_{\tilde{s}}(\mathbf{y}) = \{\theta : \bar{y}_j + n_j^{-1/2}\hat{\sigma} t_{\alpha(1-\tilde{s}(\theta))} < \theta < \bar{y}_j + n_j^{-1/2}\hat{\sigma} t_{1-\alpha\tilde{s}(\theta)}\}$

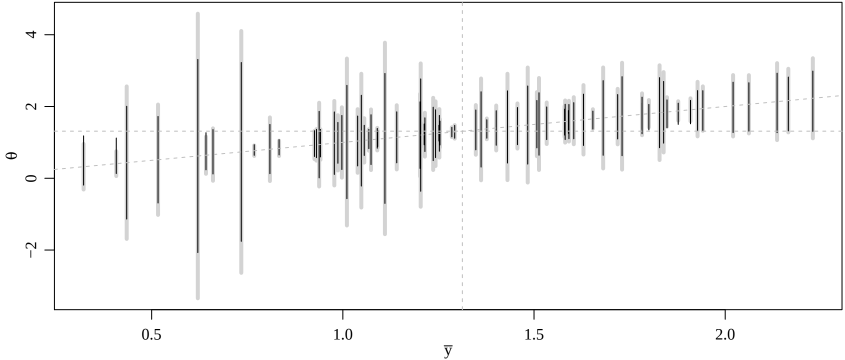
Has $1-\alpha$ constant coverage if

- $q\hat{\sigma}^2/\sigma^2 \sim \chi^2_q$, independent of $\bar{y}_j$;
- $\tilde{s}$-function independent of $(\bar{y}_j, \hat{\sigma}^2)$.

Adaptive inference: Obtain $\tilde{s}$ from groups other than j

- Group sample variances provide independent estimates of σ^2: $\hat{\sigma}^2$, $\tilde{\sigma}^2$;
- Group sample means provide $\tilde{\mu}$, $\tilde{\tau}^2$;
- $(\tilde{\mu}, \tilde{\tau}^2, \tilde{\sigma}^2)$ are independent of $(\bar{y}_j, \hat{\sigma}^2)$.
- $\tilde{s}(\theta) = g^{-1}(2\tilde{\sigma}(\theta - \tilde{\mu})/\tilde{\tau}^2)$ approximately Bayes-optimal.

Radon intervals



- FAB narrower than UMAU for 77 out of 82 counties.
- Average width ratio is 0.77.

Asymptotic optimality

Under the hierarchical normal model, the oracle CIP for θ_j is

$$C_s^j(\mathbf{y}) = \{\theta : \bar{y}_j + n_j^{-1/2}\sigma z_{\alpha(1-s(\theta))} < \theta < \bar{y}_j + n_j^{-1/2}\sigma z_{1-\alpha s(\theta)}\}$$

Our adaptive CIP for θ_j is

$$C_{\tilde{s}}^j(\mathbf{y}) = \{\theta : \bar{y}_j + n_j^{-1/2}\hat{\sigma} t_{\alpha(1-\tilde{s}(\theta))} < \theta < \bar{y}_j + n_j^{-1/2} s t_{1-\alpha\tilde{s}(\theta)}\}$$

As $p \to \infty$, almost surely we have

$$(\tilde{\mu}, \tilde{\tau}^2, \tilde{\sigma}^2, \hat{\sigma}^2) \to (\mu, \tau^2, \sigma^2, \sigma^2)$$
$$\tilde{s}(\theta) \to s(\theta)$$
$$t_\gamma \to z_\gamma$$
$$C_{\tilde{s}}^j(\mathbf{y}) \to C_s^j(\mathbf{y})$$

- coverage rate is exact, constant, non-asymptotic;
- optimality is asymptotic, assumes correctness of θ-model;
- conjecture: $C_{\tilde{s}}^j(\mathbf{y})$ dominates UMAU asymptotically, regardless of θ-model.

Asymptotic optimality

Under the hierarchical normal model, the oracle CIP for θ_j is

$$C_s^j(\mathbf{y}) = \{\theta : \bar{y}_j + n_j^{-1/2}\sigma z_{\alpha(1-s(\theta))} < \theta < \bar{y}_j + n_j^{-1/2}\sigma z_{1-\alpha s(\theta)}\}$$

Our adaptive CIP for θ_j is

$$C_{\tilde{s}}^j(\mathbf{y}) = \{\theta : \bar{y}_j + n_j^{-1/2}\hat{\sigma} t_{\alpha(1-\tilde{s}(\theta))} < \theta < \bar{y}_j + n_j^{-1/2} s t_{1-\alpha\tilde{s}(\theta)}\}$$

As $p \to \infty$, almost surely we have

$$(\tilde{\mu}, \tilde{\tau}^2, \tilde{\sigma}^2, \hat{\sigma}^2) \to (\mu, \tau^2, \sigma^2, \sigma^2)$$
$$\tilde{s}(\theta) \to s(\theta)$$
$$t_\gamma \to z_\gamma$$
$$C_{\tilde{s}}^j(\mathbf{y}) \to C_s^j(\mathbf{y})$$

- coverage rate is exact, constant, non-asymptotic;
- optimality is asymptotic, assumes correctness of θ-model;
- conjecture: $C_{\tilde{s}}^j(\mathbf{y})$ dominates UMAU asymptotically, regardless of θ-model.

Asymptotic optimality

Under the hierarchical normal model, the oracle CIP for θ_j is

$$C_s^j(\mathbf{y}) = \{\theta : \bar{y}_j + n_j^{-1/2}\sigma z_{\alpha(1-s(\theta))} < \theta < \bar{y}_j + n_j^{-1/2}\sigma z_{1-\alpha s(\theta)}\}$$

Our adaptive CIP for θ_j is

$$C_{\tilde{s}}^j(\mathbf{y}) = \{\theta : \bar{y}_j + n_j^{-1/2}\hat{\sigma} t_{\alpha(1-\tilde{s}(\theta))} < \theta < \bar{y}_j + n_j^{-1/2} s t_{1-\alpha\tilde{s}(\theta)}\}$$

As $p \to \infty$, almost surely we have

$$(\tilde{\mu}, \tilde{\tau}^2, \tilde{\sigma}^2, \hat{\sigma}^2) \to (\mu, \tau^2, \sigma^2, \sigma^2)$$
$$\tilde{s}(\theta) \to s(\theta)$$
$$t_\gamma \to z_\gamma$$
$$C_{\tilde{s}}^j(\mathbf{y}) \to C_s^j(\mathbf{y})$$

- coverage rate is exact, constant, non-asymptotic;
- optimality is asymptotic, assumes correctness of θ-model;
- conjecture: $C_{\tilde{s}}^j(\mathbf{y})$ dominates UMAU asymptotically, regardless of θ-model.

Asymptotic optimality

Under the hierarchical normal model, the oracle CIP for θ_j is

$$C^j_s(\mathbf{y}) = \{\theta : \bar{y}_j + n_j^{-1/2}\sigma z_{\alpha(1-s(\theta))} < \theta < \bar{y}_j + n_j^{-1/2}\sigma z_{1-\alpha s(\theta)}\}$$

Our adaptive CIP for θ_j is

$$C^j_{\tilde{s}}(\mathbf{y}) = \{\theta : \bar{y}_j + n_j^{-1/2}\hat{\sigma} t_{\alpha(1-\tilde{s}(\theta))} < \theta < \bar{y}_j + n_j^{-1/2} s t_{1-\alpha\tilde{s}(\theta)}\}$$

As $p \to \infty$, almost surely we have

$$(\tilde{\mu}, \tilde{\tau}^2, \tilde{\sigma}^2, \hat{\sigma}^2) \to (\mu, \tau^2, \sigma^2, \sigma^2)$$
$$\tilde{s}(\theta) \to s(\theta)$$
$$t_\gamma \to z_\gamma$$
$$C^j_{\tilde{s}}(\mathbf{y}) \to C^j_s(\mathbf{y})$$

- coverage rate is exact, constant, non-asymptotic;
- optimality is asymptotic, assumes correctness of θ-model;
- conjecture: $C^j_{\tilde{s}}(\mathbf{y})$ dominates UMAU asymptotically, regardless of θ-model.

FAB intervals for regression coefficients

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y} \sim N_p(\boldsymbol{\beta}, \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1})$$
$$\hat{\sigma}^2 = ||\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}||^2/(n-p) \sim \frac{\sigma^2}{n-p}\chi^2_{n-p}$$

$$C_u(\hat{\beta}_j, \hat{\sigma}) = \{\beta_j : \hat{\beta}_j + h_j\hat{\sigma}t_{\alpha/2} < \beta_j < \hat{\beta}_j + h_j\hat{\sigma}t_{1-\alpha/2}\}.$$

$$C_{\tilde{s}}(\hat{\beta}_j, \hat{\sigma}) = \{\beta_j : \hat{\beta}_j + h_j\hat{\sigma}t_{\alpha(1-\tilde{s}(\beta_j))} < \beta_j < \hat{\beta}_j + h_j\hat{\sigma}t_{1-\alpha\tilde{s}(\beta_j)}\},$$

- $\hat{\beta}_j$, $\hat{\sigma}$, $\tilde{s}$ independent;
- $(n-p)\hat{\sigma}^2/\sigma^2 \sim \chi^2_{n-p}$;
- $\tilde{s}$ an estimate of Bayes-optimal s-function under the prior

$$\beta_1, \ldots, \beta_p \sim \text{i.i.d. } N(\mu, \tau^2).$$

FAB intervals for regression coefficients

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top \mathbf{y} \sim N_p(\boldsymbol{\beta}, \sigma^2(\mathbf{X}^\top \mathbf{X})^{-1})$$
$$\hat{\sigma}^2 = ||\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}||^2/(n-p) \sim \tfrac{\sigma^2}{n-p}\chi^2_{n-p}$$

$$C_u(\hat{\beta}_j, \hat{\sigma}) = \{\beta_j : \hat{\beta}_j + h_j\hat{\sigma}t_{\alpha/2} < \beta_j < \hat{\beta}_j + h_j\hat{\sigma}t_{1-\alpha/2}\}.$$

$$C_{\tilde{s}}(\hat{\beta}_j, \hat{\sigma}) = \{\beta_j : \hat{\beta}_j + h_j\hat{\sigma}t_{\alpha(1-\tilde{s}(\beta_j))} < \beta_j < \hat{\beta}_j + h_j\hat{\sigma}t_{1-\alpha\tilde{s}(\beta_j)}\},$$

- $\hat{\beta}_j$, $\hat{\sigma}$, $\tilde{s}$ independent;
- $(n-p)\hat{\sigma}^2/\sigma^2 \sim \chi^2_{n-p}$;
- $\tilde{s}$ an estimate of Bayes-optimal s-function under the prior

$$\beta_1, \ldots, \beta_p \sim \text{i.i.d. } N(\mu, \tau^2).$$

FAB intervals for regression coefficients

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top \mathbf{y} \sim N_p(\boldsymbol{\beta}, \sigma^2(\mathbf{X}^\top \mathbf{X})^{-1})$$

$$\hat{\sigma}^2 = ||\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}||^2/(n-p) \sim \tfrac{\sigma^2}{n-p}\chi^2_{n-p}$$

$$C_u(\hat{\beta}_j, \hat{\sigma}) = \{\beta_j : \hat{\beta}_j + h_j \hat{\sigma} t_{\alpha/2} < \beta_j < \hat{\beta}_j + h_j \hat{\sigma} t_{1-\alpha/2}\}.$$

$$C_{\tilde{s}}(\hat{\beta}_j, \hat{\sigma}) = \{\beta_j : \hat{\beta}_j + h_j \hat{\sigma} t_{\alpha(1-\tilde{s}(\beta_j))} < \beta_j < \hat{\beta}_j + h_j \hat{\sigma} t_{1-\alpha\tilde{s}(\beta_j)}\},$$

- $\hat{\beta}_j$, $\hat{\sigma}$, $\tilde{s}$ independent;
- $(n-p)\hat{\sigma}^2/\sigma^2 \sim \chi^2_{n-p}$;
- $\tilde{s}$ an estimate of Bayes-optimal s-function under the prior

$$\beta_1, \ldots, \beta_p \sim \text{i.i.d. } N(\mu, \tau^2).$$

Independent pieces

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \sigma\boldsymbol{\epsilon}$$

- $\mathbf{P}_X = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$
- $\mathbf{P}_0 = \mathbf{I} - \mathbf{P}_X$
- $\mathbf{P}_1 = \mathbf{a}\mathbf{a}^\top/\mathbf{a}^\top\mathbf{a}$, where $\hat{\beta}_j = \mathbf{a}^\top\mathbf{y}$
- $\mathbf{P}_2 = \mathbf{P}_X(\mathbf{I} - \mathbf{P}_1)$.

$$\begin{aligned} \mathbf{y} = \mathbf{I}\mathbf{y} &= (\mathbf{P}_0 + \mathbf{P}_X)\mathbf{y} \\ &= (\mathbf{P}_0 + \mathbf{P}_1 + \mathbf{P}_2)\mathbf{y} \\ &= \mathbf{P}_0\mathbf{y} + \mathbf{P}_1\mathbf{y} + \mathbf{P}_2\mathbf{y} \equiv \mathbf{y}_0 + \mathbf{y}_1 + \mathbf{y}_2. \end{aligned}$$

Get independent estimates of

- σ^2 from $\mathbf{y}_0$,
- β_j from $\mathbf{y}_1$,
- τ^2, σ^2 from $\mathbf{y}_2$.

Independent pieces

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \sigma\boldsymbol{\epsilon}$$

- $\mathbf{P}_X = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$
- $\mathbf{P}_0 = \mathbf{I} - \mathbf{P}_X$
- $\mathbf{P}_1 = \mathbf{a}\mathbf{a}^\top/\mathbf{a}^\top\mathbf{a}$, where $\hat{\beta}_j = \mathbf{a}^\top\mathbf{y}$
- $\mathbf{P}_2 = \mathbf{P}_X(\mathbf{I} - \mathbf{P}_1)$.

$$\begin{aligned}
\mathbf{y} = \mathbf{I}\mathbf{y} &= (\mathbf{P}_0 + \mathbf{P}_X)\mathbf{y} \\
&= (\mathbf{P}_0 + \mathbf{P}_1 + \mathbf{P}_2)\mathbf{y} \\
&= \mathbf{P}_0\mathbf{y} + \mathbf{P}_1\mathbf{y} + \mathbf{P}_2\mathbf{y} \equiv \mathbf{y}_0 + \mathbf{y}_1 + \mathbf{y}_2.
\end{aligned}$$

Get independent estimates of

- σ^2 from $\mathbf{y}_0$,
- β_j from $\mathbf{y}_1$,
- τ^2, σ^2 from $\mathbf{y}_2$.

Independent pieces

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \sigma\boldsymbol{\epsilon}$$

- $\mathbf{P}_X = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$
- $\mathbf{P}_0 = \mathbf{I} - \mathbf{P}_X$
- $\mathbf{P}_1 = \mathbf{a}\mathbf{a}^\top/\mathbf{a}^\top\mathbf{a}$, where $\hat{\beta}_j = \mathbf{a}^\top\mathbf{y}$
- $\mathbf{P}_2 = \mathbf{P}_X(\mathbf{I} - \mathbf{P}_1)$.

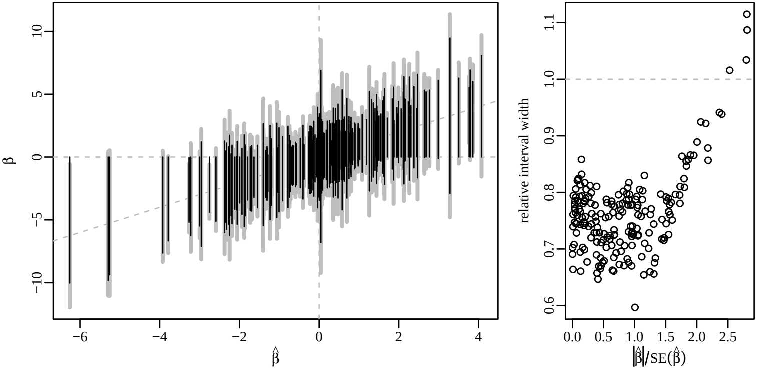
$$\begin{aligned}\mathbf{y} = \mathbf{I}\mathbf{y} &= (\mathbf{P}_0 + \mathbf{P}_X)\mathbf{y} \\ &= (\mathbf{P}_0 + \mathbf{P}_1 + \mathbf{P}_2)\mathbf{y} \\ &= \mathbf{P}_0\mathbf{y} + \mathbf{P}_1\mathbf{y} + \mathbf{P}_2\mathbf{y} \equiv \mathbf{y}_0 + \mathbf{y}_1 + \mathbf{y}_2.\end{aligned}$$

Get independent estimates of

- σ^2 from $\mathbf{y}_0$,
- β_j from $\mathbf{y}_1$,
- τ^2, σ^2 from $\mathbf{y}_2$.

Motif binding

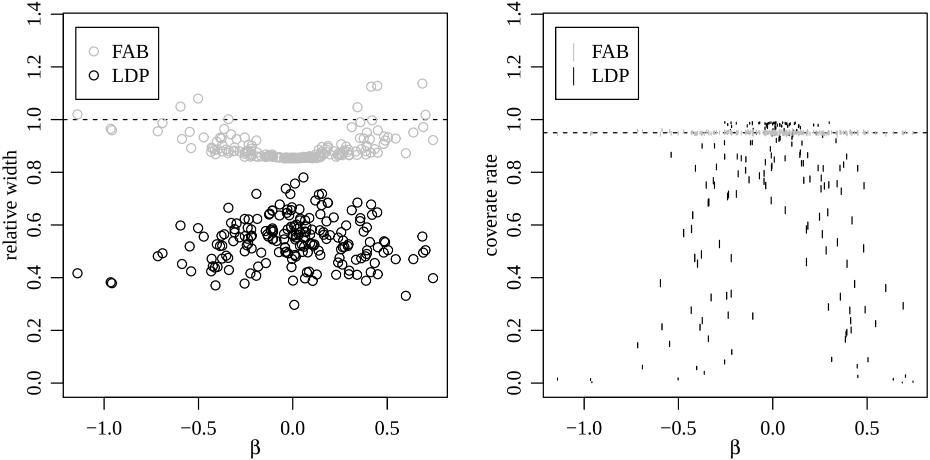
Conlon(2003) related protein binding intensity of $n = 287$ DNA segments to abundance of $p = 195$ genetic motifs.



- FAB intervals shorter than UMAU for 191 of 195 coefficients;
- relative widths ranging between 0.60 to 1.11, 0.76 on average.

Another adaptive procedure

Zhang and Zhang (2014) develop an adaptive CIP for sparse situations.



Coverage of LDP achieves nominal rate

- asymptotically;
- if certain sparsity conditions are met.

Asymptotic optimality

Intuition: If n and p are large then

- $(\hat{\sigma}, \tilde{\sigma}, \tilde{\tau}) \approx (\sigma, \sigma, \tau)$,
- $\tilde{s} \approx s$
- $z_{1-\alpha/2} \approx t_{1-\alpha/2}$,

and so we expect $C_{\tilde{s}} \approx C_s$.

Theorem: If as $n \to \infty$

- $p/n \to c \in (0, 1)$;
- $\sigma^2/n \to \sigma_0^2$;
- $\mathbf{X}$ grows in a reasonable way.

then

- $(\hat{\sigma}/n, \tilde{\sigma}/n, \tilde{\tau}) \xrightarrow{p} (\sigma_0, \sigma_0, \tau^2)$
- $\mathrm{E}[|C_{\tilde{s}}|] \to \mathrm{E}[|C_s|]$.

Asymptotic optimality

Intuition: If n and p are large then

- $(\hat{\sigma}, \tilde{\sigma}, \tilde{\tau}) \approx (\sigma, \sigma, \tau)$,
- $\tilde{s} \approx s$
- $z_{1-\alpha/2} \approx t_{1-\alpha/2}$,

and so we expect $C_{\tilde{s}} \approx C_s$.

Theorem: If as $n \to \infty$

- $p/n \to c \in (0, 1)$;
- $\sigma^2/n \to \sigma_0^2$;
- $\mathbf{X}$ grows in a reasonable way.

then

- $(\hat{\sigma}/n, \tilde{\sigma}/n, \tilde{\tau}) \xrightarrow{p} (\sigma_0, \sigma_0, \tau^2)$
- $\mathrm{E}[|C_{\tilde{s}}|] \to \mathrm{E}[|C_s|]$.

Tensor data analysis

Use tensor shape to assist with estimation and inference.

Tensor models: Assume $\mathbf{\Theta}$ is low dimensional

- $\text{vec}(\mathbf{\Theta}) = \mathbf{C} \otimes \mathbf{B} \otimes \mathbf{A}$;
- $\mathbf{\Theta}$ is low matrix or tensor rank;
- $\mathbf{\Theta}$ has other structure (e.g. spatial).

Goal: CIs, tests for elements of $\mathbf{\Theta}$ that

- are valid if tensor model is incorrect;
- adapt to tensor model if correct.

Method: Use structured tensor model as a working model for FAB.

Tensor data analysis

Use tensor shape to assist with estimation and inference.

Tensor models: Assume $\mathbf{\Theta}$ is low dimensional

- $\text{vec}(\mathbf{\Theta}) = \mathbf{C} \otimes \mathbf{B} \otimes \mathbf{A}$;
- $\mathbf{\Theta}$ is low matrix or tensor rank;
- $\mathbf{\Theta}$ has other structure (e.g. spatial).

Goal: CIs, tests for elements of $\mathbf{\Theta}$ that

- are valid if tensor model is incorrect;
- adapt to tensor model if correct.

Method: Use structured tensor model as a working model for FAB.

Tensor data analysis

Use tensor shape to assist with estimation and inference.

Tensor models: Assume $\mathbf{\Theta}$ is low dimensional

- $\text{vec}(\mathbf{\Theta}) = \mathbf{C} \otimes \mathbf{B} \otimes \mathbf{A}$;
- $\mathbf{\Theta}$ is low matrix or tensor rank;
- $\mathbf{\Theta}$ has other structure (e.g. spatial).

Goal: CIs, tests for elements of $\mathbf{\Theta}$ that

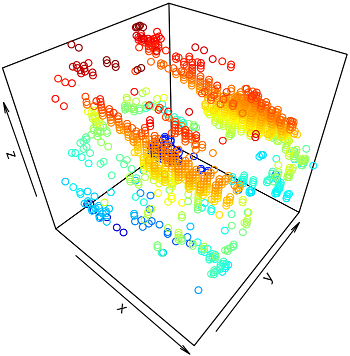
- are valid if tensor model is incorrect;
- adapt to tensor model if correct.

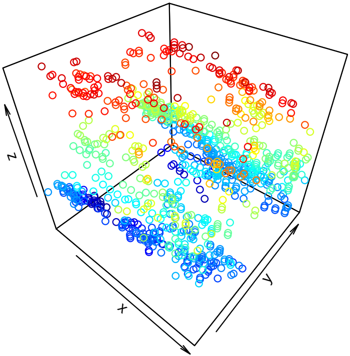
Method: Use structured tensor model as a working model for FAB.

Spatially adaptive FAB

DTI data (Deutsch et al. (2005), Efron (2010)):

- 3d brain images of 6 dyslexic and 6 non-dyslexic children;
- $p = 15,443$ voxel measurements from a $73 \times 55 \times 20$ grid;
- z-score from a two-sample t-test at each voxel.





Spatially adaptive FAB

Statistical model:

$$\mathbf{Y} = \mathbf{\Theta} + \mathbf{Z}$$

Working model for adaptation:

$$\theta_i | \mathbf{\Theta}_{-i} \sim N(\gamma \times \sum_{i' \sim i} \theta_i, \tau^2)$$

FAB CI/test: Is a value θ in the CI for θ_i? Do a level α test of

$$H : y_i \sim N(\theta, 1)$$
$$K : y_i \sim N(\hat{\theta}_i, 1)$$

- $\hat{\theta}_i$ estimated from $\mathbf{Y}_{-i}$;
- CI has $1 - \alpha$ coverage (assuming $\mathbf{Z}$ is i.i.d. $N(0, 1)$).

Spatially adaptive FAB

Statistical model:

$$\mathbf{Y} = \mathbf{\Theta} + \mathbf{Z}$$

Working model for adaptation:

$$\theta_i | \mathbf{\Theta}_{-i} \sim N(\gamma \times \sum_{i' \sim i} \theta_i, \tau^2)$$

FAB CI/test: Is a value θ in the CI for θ_i? Do a level α test of

$$H : y_i \sim N(\theta, 1)$$
$$K : y_i \sim N(\hat{\theta}_i, 1)$$

- $\hat{\theta}_i$ estimated from $\mathbf{Y}_{-i}$;
- CI has $1 - \alpha$ coverage (assuming $\mathbf{Z}$ is i.i.d. $N(0, 1)$).

Spatially adaptive FAB

Statistical model:

$$\mathbf{Y} = \mathbf{\Theta} + \mathbf{Z}$$

Working model for adaptation:

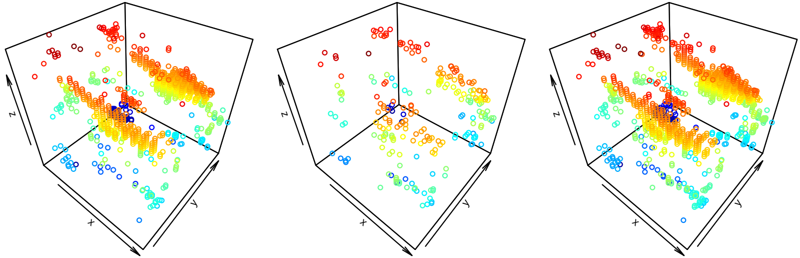
$$\theta_i | \mathbf{\Theta}_{-i} \sim N(\gamma \times \sum_{i' \sim i} \theta_i, \tau^2)$$

FAB CI/test: Is a value θ in the CI for θ_i? Do a level α test of

$$H : y_i \sim N(\theta, 1)$$
$$K : y_i \sim N(\hat{\theta}_i, 1)$$

- $\hat{\theta}_i$ estimated from $\mathbf{Y}_{-i}$;
- CI has $1 - \alpha$ coverage (assuming $\mathbf{Z}$ is i.i.d. $N(0, 1)$).

Spatially adaptive FAB



- 229 additional voxels significant ($\alpha = .05$), all with positive y_i values;
- 3 voxels no longer significant.

Adapting to multilinear tensor regression

Data: $\mathbf{y}_t = \text{vec}(\mathbf{Y}_t) \in \mathbb{R}^{p_1 p_2}$, $\mathbf{x}_t = \text{vec}(\mathbf{X}_t) \in \mathbb{R}^{q_1 q_2}$

Full model: $\boldsymbol{\Theta} \in \mathbb{R}^{p_1 p_2 \times q_1 q_2}$

$$\mathbf{y}_t = \boldsymbol{\Theta}\mathbf{x}_t + \mathbf{e}_t$$

$$y_{i,j,t} = \sum_{i'j'} \theta_{i,j,i',j'} x_{i',j',t} + e_{i,j,t}$$

Working model: $\theta_{i,j,i',j'} = a_{i,i'} a_{j,j'}$

$$\mathbf{y}_t = (\mathbf{A} \otimes \mathbf{A})\mathbf{x}_t + \mathbf{e}_t$$

$$y_{i,j,t} = \sum_{i'j'} a_{i,i'} a_{j,j'} x_{i',j',t} + e_{i,j,t}$$

Adapting to multilinear tensor regression

Data: $\mathbf{y}_t = \text{vec}(\mathbf{Y}_t) \in \mathbb{R}^{p_1 p_2}$, $\mathbf{x}_t = \text{vec}(\mathbf{X}_t) \in \mathbb{R}^{q_1 q_2}$

Full model: $\boldsymbol{\Theta} \in \mathbb{R}^{p_1 p_2 \times q_1 q_2}$

$$\mathbf{y}_t = \boldsymbol{\Theta}\mathbf{x}_t + \mathbf{e}_t$$
$$y_{i,j,t} = \sum_{i'j'} \theta_{i,j,i',j'} x_{i',j',t} + e_{i,j,t}$$

Working model: $\theta_{i,j,i',j'} = a_{i,i'} a_{j,j'}$

$$\mathbf{y}_t = (\mathbf{A} \otimes \mathbf{A})\mathbf{x}_t + \mathbf{e}_t$$
$$y_{i,j,t} = \sum_{i'j'} a_{i,i'} a_{j,j'} x_{i',j',t} + e_{i,j,t}$$

Adapting to multilinear tensor regression

Data: $\mathbf{y}_t = \text{vec}(\mathbf{Y}_t) \in \mathbb{R}^{p_1 p_2}$, $\mathbf{x}_t = \text{vec}(\mathbf{X}_t) \in \mathbb{R}^{q_1 q_2}$

Full model: $\mathbf{\Theta} \in \mathbb{R}^{p_1 p_2 \times q_1 q_2}$

$$\mathbf{y}_t = \mathbf{\Theta}\mathbf{x}_t + \mathbf{e}_t$$

$$y_{i,j,t} = \sum_{i'j'} \theta_{i,j,i',j'} x_{i',j',t} + e_{i,j,t}$$

Working model: $\theta_{i,j,i',j'} = a_{i,i'} a_{j,j'}$

$$\mathbf{y}_t = (\mathbf{A} \otimes \mathbf{A})\mathbf{x}_t + \mathbf{e}_t$$

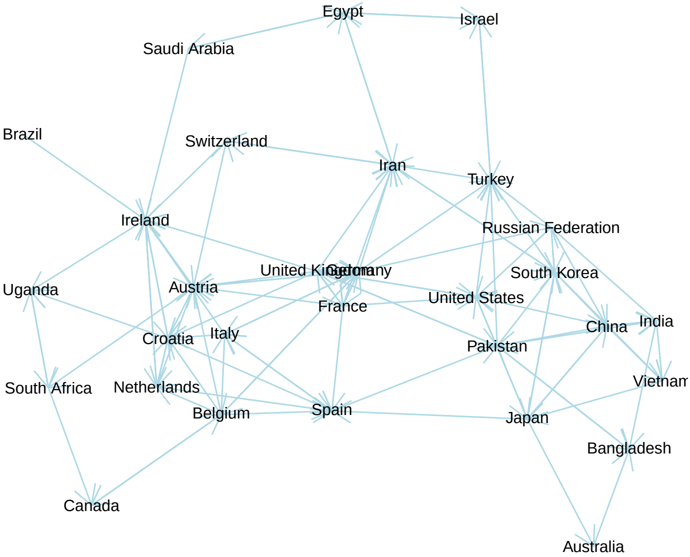
$$y_{i,j,t} = \sum_{i'j'} a_{i,i'} a_{j,j'} x_{i',j',t} + e_{i,j,t}$$

FAB approach

F.A.B : For a given dyad (i, j),

1. obtain OLS estimate $\hat{\boldsymbol{\theta}}_{ij}$ and its SE with $\{y_{i,j,t}\}$;
2. obtain $\hat{\mathbf{a}}_i$, $\hat{\mathbf{a}}_j$ of working model without $\{y_{i,j,t}\}$;
3. evaluate elements of $\boldsymbol{\theta}_{ij}$ using tests with power near $\boldsymbol{\theta}_{i,j} = \mathbf{a}_i \otimes \mathbf{a}_j$.

Adapt to this



Results

Some results: Testing at $\alpha = .05/\binom{p}{2}$,

- FAB identifies 13% more dyads as having significant influence;
- FAB-identified effects make more intuitive sense.

Some caveats:

- have not adequately addressed dependence among dyads at this stage;
- not sure how to extend precisely to sparse/non-Gaussian settings.

Results

Some results: Testing at $\alpha = .05/\binom{p}{2}$,

- FAB identifies 13% more dyads as having significant influence;
- FAB-identified effects make more intuitive sense.

Some caveats:

- have not adequately addressed dependence among dyads at this stage;
- not sure how to extend precisely to sparse/non-Gaussian settings.

Summary

- Standard approaches perform uniformly over whole parameter space;
- Vector, matrix or tensor parameters likely to be in small region of space.

- Adaptively estimate proximity of parameter to a given subregion;
- FAB approach can be used to guarantee valid adaptive tests/CIs.

Summary

- Standard approaches perform uniformly over whole parameter space;
- Vector, matrix or tensor parameters likely to be in small region of space.

- Adaptively estimate proximity of parameter to a given subregion;
- FAB approach can be used to guarantee valid adaptive tests/CIs.

References

Deutsch, G. K., R. F. Dougherty, R. Bammer, W. T. Siok, J. Gabrieli, and B. Wandell (2005). Correlations between white matter microstructure and reading performance in children. *Cortex 41*(3), 354–363.

Efron, B. (2010). *Large-scale inference*, Volume 1 of *Institute of Mathematical Statistics (IMS) Monographs*. Cambridge University Press, Cambridge. Empirical Bayes methods for estimation, testing, and prediction.

Zhang, C.-H. and S. S. Zhang (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *J. R. Stat. Soc. Ser. B. Stat. Methodol. 76*(1), 217–242.