Journal Club

Tue Herlau

Technical University of Denmark

herlau@stanford.edu

Wellcome Open Research

Wellcome Open Research 2017, 1:23 Last updated: 17 MAY 2017

Check for updates

METHOD ARTICLE

REVISED Fixing the stimulus-as-fixed-effect fallacy in task fMRI [version 2; referees: 1 approved, 2 approved with reservations]

Jacob Westfall[1*], Thomas E. Nichols [2], Tal Yarkoni[1*]

[1]Department of Psychology, University of Texas, Austin, USA
[2]Department of Statistics & WMG, University of Warwick, Coventry, UK

Wellcome Open Research

Wellcome Open Research 2017, 1:23 Last updated: 17 MAY 2017

Check for updates

METHOD ARTICLE

REVISED **Fixing the stimulus-as-fixed-effect fallacy in task fMRI [version 2; referees: 1 approved, 2 approved with reservations]**

Jacob Westfall[1*], Thomas E. Nichols [2], Tal Yarkoni[1*]

[1]Department of Psychology, University of Texas, Austin, USA
[2]Department of Statistics & WMG, University of Warwick, Coventry, UK

- How 'recycling' stimuli may lead to statistical problems

Wellcome Open Research

Wellcome Open Research 2017, 1:23 Last updated: 17 MAY 2017

Check for updates

METHOD ARTICLE

REVISED **Fixing the stimulus-as-fixed-effect fallacy in task fMRI [version 2; referees: 1 approved, 2 approved with reservations]**

Jacob Westfall[1*], Thomas E. Nichols[2], Tal Yarkoni[1*]

[1]Department of Psychology, University of Texas, Austin, USA
[2]Department of Statistics & WMG, University of Warwick, Coventry, UK

- How 'recycling' stimuli may lead to statistical problems
- That these problems are to be taken serious

Wellcome Open Research

Wellcome Open Research 2017, 1:23 Last updated: 17 MAY 2017

Check for updates

METHOD ARTICLE

REVISED **Fixing the stimulus-as-fixed-effect fallacy in task fMRI [version 2; referees: 1 approved, 2 approved with reservations]**

Jacob Westfall[1*], Thomas E. Nichols [2], Tal Yarkoni[1*]

[1]Department of Psychology, University of Texas, Austin, USA
[2]Department of Statistics & WMG, University of Warwick, Coventry, UK

- How 'recycling' stimuli may lead to statistical problems
- That these problems are to be taken serious
- A statistical model (RSM) that accounts for of these problems

- Do people in Mountain View or Palo Alto throw longest?

- Do people in Mountain View or Palo Alto throw longest?
- Method 1: Get 300 people from MV and 300 from PA. Let them throw balls and do hypothesis testing on the $n = 600$ throws ($H_0 : A = B$)

- Do people in Mountain View or Palo Alto throw longest?
- Method 1: Get 300 people from MV and 300 from PA. Let them throw balls and do hypothesis testing on the $n = 600$ throws ($H_0 : A = B$)
- Method 2: Get 3 people from MV (Jack, Julia, Pete) and 3 from PA (Mary, Sue, Bob). Let them throw balls 100 times each and do hypothesis testing on the $n = 600$ throws ($H_0 : A = B$)

Rejecting H_0 is evidence for different things:

Rejecting H_0 is evidence for different things:

- Method 1: People in MV and PA have different throwing range

Rejecting H_0 is evidence for different things:

- Method 1: People in MV and PA have different throwing range
- Method 2: (Jack, Julie, Pete) has different throwing range than (Marie, Sue, Bob).

Rejecting H_0 is evidence for different things:

- Method 1: People in MV and PA have different throwing range
- Method 2: (Jack, Julie, Pete) has different throwing range than (Marie, Sue, Bob).
- Normally we are interested in the first form of conclusion

Rejecting H_0 is evidence for different things:

- Method 1: People in MV and PA have different throwing range
- Method 2: (Jack, Julie, Pete) has different throwing range than (Marie, Sue, Bob).
- Normally we are interested in the first form of conclusion
- Second conclusion always true(!)

Rejecting H_0 is evidence for different things:

- Method 1: People in MV and PA have different throwing range
- Method 2: (Jack, Julie, Pete) has different throwing range than (Marie, Sue, Bob).
- Normally we are interested in the first form of conclusion
- Second conclusion always true(!)
- Important not to examine method 2 and conclude as method 1

The SPM GLM estimate the response of a voxel to stimuli.
More subjects = better estimates.

- Do voxel 117 respond differently to angry vs. fearful faces?

The SPM GLM estimate the response of a voxel to stimuli.
More subjects = better estimates.

- Do voxel 117 respond differently to angry vs. fearful faces?
- Method 1: Get 300 images of angry faces and 300 fearful faces. Show faces ONCE to available S subjects. Do statistics on $600S$ exposures

The SPM GLM estimate the response of a voxel to stimuli.
More subjects = better estimates.

- Do voxel 117 respond differently to angry vs. fearful faces?
- Method 1: Get 300 images of angry faces and 300 fearful faces. Show faces ONCE to available S subjects. Do statistics on $600S$ exposures
- Method 1: Get 3 images of angry faces and 3 of fearful faces. Show each face 100 times to available S subjects. Do statistics on $600S$ exposures

The SPM GLM estimate the response of a voxel to stimuli.
More subjects = better estimates.

- Do voxel 117 respond differently to angry vs. fearful faces?
- Method 1: Get 300 images of angry faces and 300 fearful faces. Show faces ONCE to available S subjects. Do statistics on $600S$ exposures
- Method 1: Get 3 images of angry faces and 3 of fearful faces. Show each face 100 times to available S subjects. Do statistics on $600S$ exposures
- Rejecting H_0 (same response) lead us to conclude:

The SPM GLM estimate the response of a voxel to stimuli.
More subjects = better estimates.

- Do voxel 117 respond differently to angry vs. fearful faces?
- Method 1: Get 300 images of angry faces and 300 fearful faces. Show faces ONCE to available S subjects. Do statistics on $600S$ exposures
- Method 1: Get 3 images of angry faces and 3 of fearful faces. Show each face 100 times to available S subjects. Do statistics on $600S$ exposures
- Rejecting H_0 (same response) lead us to conclude:
- Case 1: Voxel 117 show differentiated activity to angry / fearful faces

The SPM GLM estimate the response of a voxel to stimuli.
More subjects = better estimates.

- Do voxel 117 respond differently to angry vs. fearful faces?
- Method 1: Get 300 images of angry faces and 300 fearful faces. Show faces ONCE to available S subjects. Do statistics on $600S$ exposures
- Method 1: Get 3 images of angry faces and 3 of fearful faces. Show each face 100 times to available S subjects. Do statistics on $600S$ exposures
- Rejecting H_0 (same response) lead us to conclude:
- Case 1: Voxel 117 show differentiated activity to angry / fearful faces
- Case 2: Voxel 117 show differential activity to (`angry.png1`, `angry.png2`, `angry.png3`) vs. (`fear.png1`, `fear.png2`, `fear.png3`)

The SPM GLM estimate the response of a voxel to stimuli.
More subjects = better estimates.

- Do voxel 117 respond differently to angry vs. fearful faces?
- Method 1: Get 300 images of angry faces and 300 fearful faces. Show faces ONCE to available S subjects. Do statistics on $600S$ exposures
- Method 1: Get 3 images of angry faces and 3 of fearful faces. Show each face 100 times to available S subjects. Do statistics on $600S$ exposures
- Rejecting H_0 (same response) lead us to conclude:
- Case 1: Voxel 117 show differentiated activity to angry / fearful faces
- Case 2: Voxel 117 show differential activity to (`angry.png1`, `angry.png2`, `angry.png3`) vs. (`fear.png1`, `fear.png2`, `fear.png3`)
- As for throwing, under mild assumptions Case 2 will be trivially true, but it is not interesting

- Is this a real problem?

- Is this a real problem?
- A: Yes

> *We ultimately found that 63/100 of the experiments (...) published test statistics for these experiments [that] are likely inflated (...)*

- Is the effect significant?

- Is the effect significant?
- Artificial data analyzed with SPM (false positive rate of test)

		alpha = .05	alpha = .01	alpha = .005	alpha = .001
16	16	0.344	0.190	0.148	0.082
	32	0.274	0.150	0.106	0.046
	64	0.208	0.104	0.080	0.030
32	16	0.478	0.316	0.270	0.200
	32	0.438	0.322	0.262	0.170
	64	0.290	0.186	0.148	0.080
64	16	0.642	0.548	0.514	0.410
	32	0.568	0.442	0.400	0.310
	64	0.452	0.298	0.254	0.184

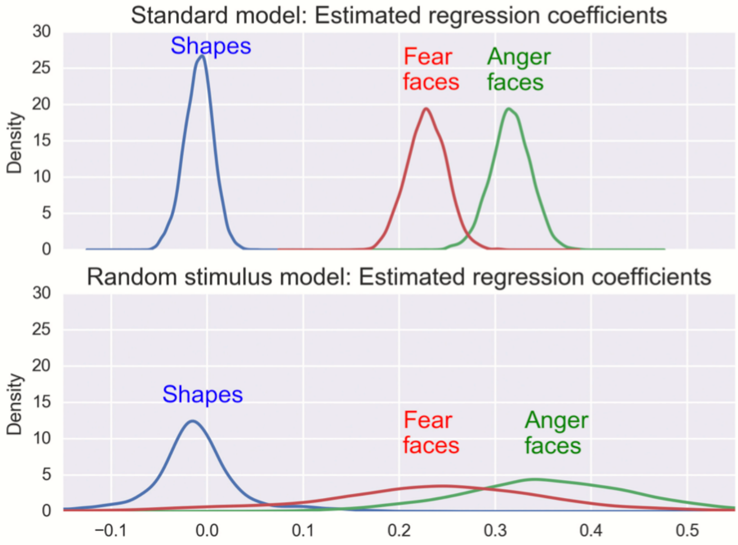
FPR

Bad news bears!

Number of different stimuli

Subjects

- Is the effect significant?
- HPC amygdala response to shapes, anger and fear faces



- Don't recycle stimuli

- Don't recycle stimuli
- Frame generalizations narrowly (i.e. "the set of stimuli considered from image database X") rather than the entire class the stimuli represents (i.e. "sad and angry faces").

- Don't recycle stimuli
- Frame generalizations narrowly (i.e. "the set of stimuli considered from image database X") rather than the entire class the stimuli represents (i.e. "sad and angry faces").
- Extend GLM to get around this problem

- Recall $Y_{it} = \mu_{it} + \epsilon_{it}$, $\epsilon_{it} \sim \mathrm{AR}(2)$

$$\text{Basic GLM:} \quad \mu_{it} = \beta_0 + \sum_{k=1}^{K} (\beta_k + p_{ki}) X_{kit}$$

Y_{it} is observed BOLD for subject i at time t, k is experimental condition (i.e. angry, fearful) and X_{kit} is the (canonical) response for stimuli k obtained as

$$X_{kit} = \sum_{j \text{ image in condition } k} x_{ijt}$$

- Recall $Y_{it} = \mu_{it} + \epsilon_{it}$, $\epsilon_{it} \sim \mathrm{AR}(2)$

$$\text{Basic GLM:} \quad \mu_{it} = \beta_0 + \sum_{k=1}^{K} (\beta_k + p_{ki}) X_{kit}$$

  Y_{it} is observed BOLD for subject i at time t, k is experimental condition (i.e. angry, fearful) and X_{kit} is the (canonical) response for stimuli k obtained as

$$X_{kit} = \sum_{j \text{ image in condition } k} x_{ijt}$$

- Why do we need the subject-specific p_{ki} values?

- Recall $Y_{it} = \mu_{it} + \epsilon_{it}$, $\epsilon_{it} \sim \mathrm{AR}(2)$

$$\text{Basic GLM:} \quad \mu_{it} = \beta_0 + \sum_{k=1}^{K} (\beta_k + p_{ki}) X_{kit}$$

  Y_{it} is observed BOLD for subject i at time t, k is experimental condition (i.e. angry, fearful) and X_{kit} is the (canonical) response for stimuli k obtained as

$$X_{kit} = \sum_{j \text{ image in condition } k} x_{ijt}$$

- Why do we need the subject-specific p_{ki} values?
- Ensures outliers ($p_{k,37} \gg 0$) not affect estimate of β_k thereby inducing false positives

- Recall $Y_{it} = \mu_{it} + \epsilon_{it}$, $\epsilon_{it} \sim \mathrm{AR}(2)$

$$\text{Basic GLM:} \quad \mu_{it} = \beta_0 + \sum_{k=1}^{K} (\beta_k + p_{ki}) X_{kit}$$

  Y_{it} is observed BOLD for subject i at time t, k is experimental condition (i.e. angry, fearful) and X_{kit} is the (canonical) response for stimuli k obtained as

$$X_{kit} = \sum_{j \text{ image in condition } k} x_{ijt}$$

- Why do we need the subject-specific p_{ki} values?
- Ensures outliers ($p_{k,37} \gg 0$) not affect estimate of β_k thereby inducing false positives
- Bayesian perspective: Model subject-specific effects if they are (plausibly) present

- If we assume stimuli j (as presented to subject i) has a stimuli-specific effect (i.e. an extra-angry face, extra-hard math problem, etc.)

- If we assume stimuli j (as presented to subject i) has a stimuli-specific effect (i.e. an extra-angry face, extra-hard math problem, etc.)
- $$\text{Proposed GLM:} \quad \mu_{it} = \beta_0 + \sum_{k=1}^{K} (\beta_k + p_{ki}) X_{kit} + \sum_{j=1}^{J} s_j x_{ijt}$$

- If we assume stimuli j (as presented to subject i) has a stimuli-specific effect (i.e. an extra-angry face, extra-hard math problem, etc.)
- $$\text{Proposed GLM:} \quad \mu_{it} = \beta_0 + \sum_{k=1}^{K} (\beta_k + p_{ki}) X_{kit} + \sum_{j=1}^{J} s_j x_{ijt}$$
- If this feature *is* present in data, and our model *does not* capture it, the model will more often find spurious results (false positives, bad estimates, etc.)

- Full model specification: $Y_{it} \sim \mathcal{N}(\mu_{it}, \sigma_Y)$ where

$$\mu_{it} = \alpha_1 Y_{i,t-1} + \alpha_2 Y_{i,t-2} + \beta_0 + \sum_{k=1}^{K} (\beta_k + p_{ki}) X_{kit} + \sum_{j=1}^{J} s_j x_{ijt}$$

$$\beta_k \sim \mathcal{N}(0, 10^6), \quad p_{ki} \sim \mathcal{N}(0, \sigma_k), \quad s_j \sim \mathcal{N}(0, \sigma_s)$$

$$\sigma_k, \sigma_s, \sigma_Y \sim \mathrm{Cauchy}^+(10), \quad \alpha_1, \alpha_2 \sim \mathrm{Cauchy}(0, 1)$$

- Full model specification: $Y_{it} \sim \mathcal{N}(\mu_{it}, \sigma_Y)$ where

$$\mu_{it} = \alpha_1 Y_{i,t-1} + \alpha_2 Y_{i,t-2} + \beta_0 + \sum_{k=1}^{K} (\beta_k + p_{ki}) X_{kit} + \sum_{j=1}^{J} s_j x_{ijt}$$

$$\beta_k \sim \mathcal{N}(0, 10^6), \quad p_{ki} \sim \mathcal{N}(0, \sigma_k), \quad s_j \sim \mathcal{N}(0, \sigma_s)$$

$$\sigma_k, \sigma_s, \sigma_Y \sim \mathrm{Cauchy}^+(10), \quad \alpha_1, \alpha_2 \sim \mathrm{Cauchy}(0, 1)$$

- *"This model cannot be fit using standard mixed modeling statistical packages, such as lme4 in R or SAS PROC MIXED (...) For the models in this paper, we used the PyMC3 Python package, which is built on the Theano deep learning package and implements the state-of-the-art No U-Turn MCMC Sampler"*

- If we generate $z \sim \mathcal{G}(\alpha = \frac{1}{2}, \beta = \frac{\sigma^2}{2})$, $x \sim \mathcal{N}(0, \sigma^{-2} = z)$ then:

$$x \sim \mathrm{Cauchy}(0, \sigma)$$

- If we generate $z \sim \mathcal{G}(\alpha = \frac{1}{2}, \beta = \frac{\sigma^2}{2})$, $x \sim \mathcal{N}(0, \sigma^{-2} = z)$ then:

$$x \sim \mathrm{Cauchy}(0, \sigma)$$

- Therefore, we can replace all the Cauchy distributions with $\mathrm{Cauchy}(x; 0, \sigma) = \int \mathcal{N}(x; 0, \sigma^{-2} = z)\mathcal{G}(z; \alpha = \frac{1}{2}, \beta = \frac{\sigma^2}{2})dz$

- If we generate $z \sim \mathcal{G}(\alpha = \frac{1}{2}, \beta = \frac{\sigma^2}{2})$, $x \sim \mathcal{N}(0, \sigma^{-2} = z)$ then:

$$x \sim \mathrm{Cauchy}(0, \sigma)$$

- Therefore, we can replace all the Cauchy distributions with $\mathrm{Cauchy}(x; 0, \sigma) = \int \mathcal{N}(x; 0, \sigma^{-2} = z)\mathcal{G}(z; \alpha = \frac{1}{2}, \beta = \frac{\sigma^2}{2})dz$
- For our model

$$\mu_{it} = \alpha_1 Y_{i,t-1} + \alpha_2 Y_{i,t-2} + \beta_0 + \sum_{k=1}^{K}(\beta_k + p_{ki})X_{kit} + \sum_{j=1}^{J} s_j x_{ijt}$$

we collect all parameters $\boldsymbol{\theta} = \begin{bmatrix} \alpha_1 & \alpha_2 & \beta_0 & \beta_k & \cdots \end{bmatrix}^T$

- If we generate $z \sim \mathcal{G}(\alpha = \frac{1}{2}, \beta = \frac{\sigma^2}{2})$, $x \sim \mathcal{N}(0, \sigma^{-2} = z)$ then:

$$x \sim \mathrm{Cauchy}(0, \sigma)$$

- Therefore, we can replace all the Cauchy distributions with $\mathrm{Cauchy}(x; 0, \sigma) = \int \mathcal{N}(x; 0, \sigma^{-2} = z)\mathcal{G}(z; \alpha = \frac{1}{2}, \beta = \frac{\sigma^2}{2})dz$
- For our model

$$\mu_{it} = \alpha_1 Y_{i,t-1} + \alpha_2 Y_{i,t-2} + \beta_0 + \sum_{k=1}^{K}(\beta_k + p_{ki})X_{kit} + \sum_{j=1}^{J} s_j x_{ijt}$$

  we collect all parameters $\boldsymbol{\theta} = \begin{bmatrix} \alpha_1 & \alpha_2 & \beta_0 & \beta_k & \cdots \end{bmatrix}^T$
- (i) the prior distribution of $\boldsymbol{\theta}$ is a multivariate normal

- If we generate $z \sim \mathcal{G}(\alpha = \frac{1}{2}, \beta = \frac{\sigma^2}{2})$, $x \sim \mathcal{N}(0, \sigma^{-2} = z)$ then:

$$x \sim \mathrm{Cauchy}(0, \sigma)$$

- Therefore, we can replace all the Cauchy distributions with $\mathrm{Cauchy}(x; 0, \sigma) = \int \mathcal{N}(x; 0, \sigma^{-2} = z)\mathcal{G}(z; \alpha = \frac{1}{2}, \beta = \frac{\sigma^2}{2})dz$
- For our model

$$\mu_{it} = \alpha_1 Y_{i,t-1} + \alpha_2 Y_{i,t-2} + \beta_0 + \sum_{k=1}^{K} (\beta_k + p_{ki}) X_{kit} + \sum_{j=1}^{J} s_j x_{ijt}$$

we collect all parameters $\boldsymbol{\theta} = \begin{bmatrix} \alpha_1 & \alpha_2 & \beta_0 & \beta_k & \cdots \end{bmatrix}^T$

- (i) the prior distribution of $\boldsymbol{\theta}$ is a multivariate normal
- (ii) $\boldsymbol{\theta}$ enters linearly in the likelihood

- If we generate $z \sim \mathcal{G}(\alpha = \frac{1}{2}, \beta = \frac{\sigma^2}{2})$, $x \sim \mathcal{N}(0, \sigma^{-2} = z)$ then:

$$x \sim \mathrm{Cauchy}(0, \sigma)$$

- Therefore, we can replace all the Cauchy distributions with $\mathrm{Cauchy}(x; 0, \sigma) = \int \mathcal{N}(x; 0, \sigma^{-2} = z)\mathcal{G}(z; \alpha = \frac{1}{2}, \beta = \frac{\sigma^2}{2})dz$
- For our model

$$\mu_{it} = \alpha_1 Y_{i,t-1} + \alpha_2 Y_{i,t-2} + \beta_0 + \sum_{k=1}^{K}(\beta_k + p_{ki})X_{kit} + \sum_{j=1}^{J} s_j x_{ijt}$$

we collect all parameters $\boldsymbol{\theta} = \begin{bmatrix} \alpha_1 & \alpha_2 & \beta_0 & \beta_k & \cdots \end{bmatrix}^T$

- (i) the prior distribution of $\boldsymbol{\theta}$ is a multivariate normal
- (ii) $\boldsymbol{\theta}$ enters linearly in the likelihood
- (iii) conditional on $\boldsymbol{\theta}$, all other parameters become independent

- We can easily derive the marginal of $\boldsymbol{\theta}$. Let's for simplicity just focus on the marginal of s_j. It can be found by noting $\mu_{it} = \{\text{other terms}\} + \sum_{j=1}^{J} s_j x_{ijt}$ or:

$$\boldsymbol{\mu} = \{\text{other terms}\} + \sum_{i=1}^{N} \boldsymbol{e}_i \boldsymbol{s}^T \boldsymbol{x}_{i,\cdot,\cdot}$$

- We can easily derive the marginal of $\boldsymbol{\theta}$. Let's for simplicity just focus on the marginal of s_j. It can be found by noting $\mu_{it} = \{\text{other terms}\} + \sum_{j=1}^{J} s_j x_{ijt}$ or:

$$\boldsymbol{\mu} = \{\text{other terms}\} + \sum_{i=1}^{N} \boldsymbol{e}_i \boldsymbol{s}^T \boldsymbol{x}_{i,\cdot,\cdot}$$

- This means that

$$\begin{aligned} \mathrm{vec}(\boldsymbol{\mu}) &= \mathrm{vec}(\sum_{i=1}^{N} \boldsymbol{e}_i \boldsymbol{s}^T \boldsymbol{x}_{i,\cdot,\cdot}) = \sum_{i=1}^{N} \mathrm{vec}(\boldsymbol{e}_i \boldsymbol{s}^T \boldsymbol{x}_{i,\cdot,\cdot}) \\ &= \sum_{i} (\boldsymbol{x}_{i,\cdot,\cdot} \otimes \boldsymbol{e}_i) \mathrm{vec}(\boldsymbol{s}^T) = \left[\sum_{i} (\boldsymbol{x}_{i,\cdot,\cdot} \otimes \boldsymbol{e}_i) \right] \boldsymbol{s} \end{aligned}$$

- We can easily derive the marginal of $\boldsymbol{\theta}$. Let's for simplicity just focus on the marginal of s_j. It can be found by noting $\mu_{it} = \{\text{other terms}\} + \sum_{j=1}^{J} s_j x_{ijt}$ or:

$$\boldsymbol{\mu} = \{\text{other terms}\} + \sum_{i=1}^{N} \boldsymbol{e}_i \boldsymbol{s}^T \boldsymbol{x}_{i,\cdot,\cdot}$$

- This means that

$$\begin{aligned}\mathrm{vec}(\boldsymbol{\mu}) &= \mathrm{vec}(\sum_{i=1}^{N} \boldsymbol{e}_i \boldsymbol{s}^T \boldsymbol{x}_{i,\cdot,\cdot}) = \sum_{i=1}^{N} \mathrm{vec}(\boldsymbol{e}_i \boldsymbol{s}^T \boldsymbol{x}_{i,\cdot,\cdot}) \\ &= \sum_i (\boldsymbol{x}_{i,\cdot,\cdot} \otimes \boldsymbol{e}_i) \mathrm{vec}(\boldsymbol{s}^T) = \left[\sum_i (\boldsymbol{x}_{i,\cdot,\cdot} \otimes \boldsymbol{e}_i)\right] \boldsymbol{s}\end{aligned}$$

- Therefore

$$\mathrm{vec}(\boldsymbol{\mu}) = \hat{\boldsymbol{X}}\boldsymbol{\theta} \Rightarrow \mathrm{vec}(\boldsymbol{Y} - \boldsymbol{\mu}) \sim \mathcal{N}(\mathrm{vec}(\boldsymbol{Y}) - \hat{\boldsymbol{X}}\boldsymbol{\theta}, \sigma_Y)$$

From which the posterior of $\boldsymbol{\theta}$ as a multivariate normal trivially follows

- Marginals can easily derived. Take as en example σ_Y and note likelihood has this form

$$p(\cdots, \sigma_Y, \cdots) = \{\text{Other stuff}\} \times \prod_{i,t} \mathcal{N}(Y_{it}|\mu_{it}, \sigma_Y)\mathcal{N}(\sigma_Y; 0, \tau)$$

- Marginals can easily derived. Take as en example σ_Y and note likelihood has this form

$$p(\cdots, \sigma_Y, \cdots) = \{\text{Other stuff}\} \times \prod_{i,t} \mathcal{N}(Y_{it}|\mu_{it}, \sigma_Y)\mathcal{N}(\sigma_Y; 0, \tau)$$

- Writing out this expression we see σ_Y enters as:

$$p(\cdots, \sigma_Y, \cdots) \propto \sigma_Y^q e^{-a\sigma_Y^2 - \frac{b}{\sigma_Y^2}}$$

(Generalized Inverse Gaussian (GiG) distribution)

- Marginals can easily derived. Take as en example σ_Y and note likelihood has this form

$$p(\cdots, \sigma_Y, \cdots) = \{\text{Other stuff}\} \times \prod_{i,t} \mathcal{N}(Y_{it} | \mu_{it}, \sigma_Y) \mathcal{N}(\sigma_Y; 0, \tau)$$

- Writing out this expression we see σ_Y enters as:

$$p(\cdots, \sigma_Y, \cdots) \propto \sigma_Y^q e^{-a\sigma_Y^2 - \frac{b}{\sigma_Y^2}}$$

  (Generalized Inverse Gaussian (GiG) distribution)
- MCMC sampling procedure:

- Marginals can easily derived. Take as en example σ_Y and note likelihood has this form

$$p(\cdots, \sigma_Y, \cdots) = \{\text{Other stuff}\} \times \prod_{i,t} \mathcal{N}(Y_{it}|\mu_{it}, \sigma_Y)\mathcal{N}(\sigma_Y; 0, \tau)$$

- Writing out this expression we see σ_Y enters as:

$$p(\cdots, \sigma_Y, \cdots) \propto \sigma_Y^q e^{-a\sigma_Y^2 - \frac{b}{\sigma_Y^2}}$$

(Generalized Inverse Gaussian (GiG) distribution)

- MCMC sampling procedure:
  - Sample θ from a single, multivariate normal distribution

- Marginals can easily derived. Take as en example σ_Y and note likelihood has this form

$$p(\cdots, \sigma_Y, \cdots) = \{\text{Other stuff}\} \times \prod_{i,t} \mathcal{N}(Y_{it}|\mu_{it}, \sigma_Y)\mathcal{N}(\sigma_Y; 0, \tau)$$

- Writing out this expression we see σ_Y enters as:

$$p(\cdots, \sigma_Y, \cdots) \propto \sigma_Y^q e^{-a\sigma_Y^2 - \frac{b}{\sigma_Y^2}}$$

(Generalized Inverse Gaussian (GiG) distribution)

- MCMC sampling procedure:
  - Sample θ from a single, multivariate normal distribution
  - Sample variances $(\sigma_Y, \sigma_s, \sigma_k)$ from GiG distributions

- Marginals can easily derived. Take as en example σ_Y and note likelihood has this form

$$p(\cdots, \sigma_Y, \cdots) = \{\text{Other stuff}\} \times \prod_{i,t} \mathcal{N}(Y_{it}|\mu_{it}, \sigma_Y)\mathcal{N}(\sigma_Y; 0, \tau)$$

- Writing out this expression we see σ_Y enters as:

$$p(\cdots, \sigma_Y, \cdots) \propto \sigma_Y^q e^{-a\sigma_Y^2 - \frac{b}{\sigma_Y^2}}$$

  (Generalized Inverse Gaussian (GiG) distribution)
- MCMC sampling procedure:
  - Sample θ from a single, multivariate normal distribution
  - Sample variances $(\sigma_Y, \sigma_s, \sigma_k)$ from GiG distributions
  - Sample the latent parameters $(z_Y, z_s, z_k, \dots)$ from their marginals (Gamma)

Conclusion

Conclusion

- When designing experiments, don't recycle stimuli

Conclusion

- When designing experiments, don't recycle stimuli
- If stimuli from the same class *cannot* be assumed to induce similar response, use a model which captures this feature to avoid many false positives

Conclusion

- When designing experiments, don't recycle stimuli
- If stimuli from the same class *cannot* be assumed to induce similar response, use a model which captures this feature to avoid many false positives
- Get code from `citron.stanford.edu` at: `/mnt/mapricot/musk2/home/herlau/shared/bayes_rsm/`