"Strong happiness" and other properties of certain imprecise probability models when treating samples sequentially

Gero Walter

Institut für Statistik
Ludwig-Maximilians-Universität München

May 12th, 2010

Prior-Data Conflict

Prior-Data Conflict $\hat{=}$ situation in which. . .

- ...*informative prior beliefs* and *trusted data*
  (sampling model correct, no outliers, etc.) are in conflict

- "... the prior [places] its mass primarily on distributions in
  the sampling model for which the observed data is surprising."
  (Evans & Moshonov, 2006)

Simple Example: Dirichlet-Multinomial-Model

Data:	$\mathbf{k}$	$\sim$	$\mathsf{M}(\boldsymbol{\theta})$	$(\sum k_j = n)$
conjugate prior:	$\boldsymbol{\theta}$	$\sim$	$\mathsf{Dir}(\boldsymbol{\alpha})$	$(\sum \theta_j = 1)$
posterior:	$\boldsymbol{\theta} \mid \mathbf{k}$	$\sim$	$\mathsf{Dir}(\boldsymbol{\alpha} + \mathbf{k})$	

$$\mathbb{E}[\theta_j] = \frac{\alpha_j}{\sum \alpha_i} \qquad\qquad \mathbb{V}(\theta_j) = \frac{\mathbb{E}[\theta_j](1 - \mathbb{E}[\theta_j])}{\sum \alpha_i + 1}$$

Simple Example: Dirichlet-Multinomial-Model

Data:	$\mathbf{k}$	$\sim$	$M(\boldsymbol{\theta})$	$(\sum k_j = n)$
conjugate prior:	$\boldsymbol{\theta}$	$\sim$	$Dir(\boldsymbol{\alpha})$	$(\sum \theta_j = 1)$
posterior:	$\boldsymbol{\theta} \mid \mathbf{k}$	$\sim$	$Dir(\boldsymbol{\alpha} + \mathbf{k})$	

$$\mathbb{E}[\theta_j] = \frac{\alpha_j}{\sum \alpha_i} =: y_j^{(0)} \quad \mathbb{V}(\theta_j) = \frac{\mathbb{E}[\theta_j](1 - \mathbb{E}[\theta_j])}{\sum \alpha_i + 1} = \frac{y_j^{(0)}(1 - y_j^{(0)})}{n^{(0)} + 1}$$

Data :	$\mathbf{k}$	$\sim$	$M(\boldsymbol{\theta})$	
conjugate prior:	$\boldsymbol{\theta}$	$\sim$	$Dir(n^{(0)}, \mathbf{y}^{(0)})$	$n^{(0)} = \sum \alpha_i$
posterior:	$\boldsymbol{\theta} \mid \mathbf{k}$	$\sim$	$Dir(n^{(1)}, \mathbf{y}^{(1)})$	

Simple Example: Dirichlet-Multinomial-Model

Data:	$\mathbf{k}$	$\sim$	$\mathsf{M}(\boldsymbol{\theta})$	$(\sum k_j = n)$
conjugate prior:	$\boldsymbol{\theta}$	$\sim$	$\mathsf{Dir}(\boldsymbol{\alpha})$	$(\sum \theta_j = 1)$
posterior:	$\boldsymbol{\theta} \mid \mathbf{k}$	$\sim$	$\mathsf{Dir}(\boldsymbol{\alpha} + \mathbf{k})$	

$$\mathbb{E}[\theta_j] = \frac{\alpha_j}{\sum \alpha_i} =: y_j^{(0)} \quad \mathbb{V}(\theta_j) = \frac{\mathbb{E}[\theta_j](1 - \mathbb{E}[\theta_j])}{\sum \alpha_i + 1} = \frac{y_j^{(0)}(1 - y_j^{(0)})}{n^{(0)} + 1}$$
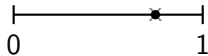
Data :	$\mathbf{k}$	$\sim$	$\mathsf{M}(\boldsymbol{\theta})$	
conjugate prior:	$\boldsymbol{\theta}$	$\sim$	$\mathsf{Dir}(n^{(0)}, \mathbf{y}^{(0)})$	$n^{(0)} = \sum \alpha_i$
posterior:	$\boldsymbol{\theta} \mid \mathbf{k}$	$\sim$	$\mathsf{Dir}(n^{(1)}, \mathbf{y}^{(1)})$	

$$y_j^{(1)} = \frac{n^{(0)}}{n^{(0)} + n} \cdot y_j^{(0)} + \frac{n}{n^{(0)} + n} \cdot \frac{k_j}{n} \qquad n^{(1)} = n^{(0)} + n$$
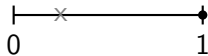
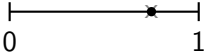# Simple Example: Dirichlet-Multinomial-Model

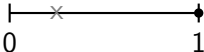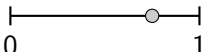Case (i): $\quad y_j^{(0)} = 0.75$, $\quad k_j/n = 0.75$

$(n^{(0)} = 8)$ $\quad (n = 16)$

Case (ii): $\quad y_j^{(0)} = 0.25$, $\quad k_j/n = 1$

$(n^{(0)} = 8)$ $\quad (n = 16)$

Simple Example: Dirichlet-Multinomial-Model

Case (i): $\quad y_j^{(0)} = 0.75$, $\quad k_j/n = 0.75$
$\quad\quad\quad (n^{(0)} = 8)$ $\quad\quad (n = 16)$

Case (ii): $\quad y_j^{(0)} = 0.25$, $\quad k_j/n = 1$
$\quad\quad\quad (n^{(0)} = 8)$ $\quad\quad (n = 16)$

$\longrightarrow \quad \mathbb{E}[\theta_j \mid \mathbf{k}] = y_j^{(1)} = 0.75$, $\quad \mathbb{V}(\theta_j \mid \mathbf{k}) = 3/400$

$$(\mathbb{V}(\theta_j) = 1/48)$$

> ⚠ Posterior inferences do not reflect uncertainty ⚠
> due to unexpected observations!

Conjugate Priors

Weighted average structure is underneath *all common* conjugate priors for exponential family sampling distributions!

$X \overset{iid}{\sim}$ linear, canonical exponential family, i.e.

$$p(x \mid \theta) \propto \exp\left\{ \langle \psi, \tau(x) \rangle - n\mathbf{b}(\psi) \right\} \qquad \left[\psi \text{ transformation of } \theta\right]$$

Conjugate Priors

Weighted average structure is underneath *all common* conjugate priors for exponential family sampling distributions!

$X \overset{iid}{\sim}$ linear, canonical exponential family, i.e.

$$p(x \mid \theta) \propto \exp \left\{ \langle \psi, \tau(x) \rangle - n\mathbf{b}(\psi) \right\} \qquad \left[\psi \text{ transformation of } \theta \right]$$

➡ conjugate prior:

$$p(\theta) \propto \exp \left\{ n^{(0)} \big[\langle \psi, y^{(0)} \rangle - \mathbf{b}(\psi) \big] \right\}$$

Conjugate Priors

Weighted average structure is underneath *all common* conjugate priors for exponential family sampling distributions!

$X \overset{iid}{\sim}$ linear, canonical exponential family, i.e.

$$p(x \mid \theta) \propto \exp\left\{\langle \psi, \tau(x) \rangle - n\mathbf{b}(\psi)\right\} \qquad \left[\psi \text{ transformation of } \theta\right]$$

➡ conjugate prior:

$$p(\theta) \propto \exp\left\{n^{(0)}\left[\langle \psi, y^{(0)} \rangle - \mathbf{b}(\psi)\right]\right\}$$

➡ (conjugate) posterior:

$$p(\theta \mid x) \propto \exp\left\{n^{(1)}\left[\langle \psi, y^{(1)} \rangle - \mathbf{b}(\psi)\right]\right\},$$

where $\quad y^{(1)} = \dfrac{n^{(0)}}{n^{(0)} + n} \cdot y^{(0)} + \dfrac{n}{n^{(0)} + n} \cdot \dfrac{1}{n}\tau(x) \quad$ and $\quad n^{(1)} = n^{(0)} + n$.

Prior-Data Conflict
Generalized iLUCK-models
"Strong Happiness"

Simple Example
Parameter Sets
Direct vs. Sequential Updating

Generalized iLUCK-models

Model for Bayesian inference with sets of priors
(Walter & Augustin, 2009)

1. use conjugate priors from general construction method
   (prior parameters $y^{(0)}$, $n^{(0)}$)

Prior-Data Conflict
Generalized iLUCK-models
"Strong Happiness"

Simple Example
Parameter Sets
Direct vs. Sequential Updating

Generalized iLUCK-models

Model for Bayesian inference with sets of priors
(Walter & Augustin, 2009)

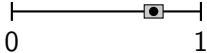1. use conjugate priors from general construction method
   (prior parameters $y^{(0)}$, $n^{(0)}$)

2. construct sets of priors via sets of parameters
   $y^{(0)} \in \mathcal{Y}^{(0)} \times n^{(0)} \in \mathcal{N}^{(0)}$

Prior-Data Conflict
Generalized iLUCK-models
"Strong Happiness"

Simple Example
Parameter Sets
Direct vs. Sequential Updating

Generalized iLUCK-models

Model for Bayesian inference with sets of priors
(Walter & Augustin, 2009)

1. use conjugate priors from general construction method
   (prior parameters $y^{(0)}$, $n^{(0)}$)

2. construct sets of priors via sets of parameters
   $y^{(0)} \in \mathcal{Y}^{(0)} \times n^{(0)} \in \mathcal{N}^{(0)}$

3. set of posteriors $\hat{=}$ set of (element-wise) updated priors
   ➡ still easy to handle: described as set of $(y^{(1)}, n^{(1)})$'s

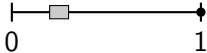$$y^{(1)} = \frac{n^{(0)}}{n^{(0)} + n} \cdot y^{(0)} + \frac{n}{n^{(0)} + n} \cdot \frac{1}{n}\tau(x)$$

$$n^{(1)} = n^{(0)} + n$$

Prior-Data Conflict
Generalized iLUCK-models
"Strong Happiness"

Simple Example
Parameter Sets
Direct vs. Sequential Updating

Generalized iLUCK-models: Simple Example

Case (i): $\quad y_j^{(0)} \in [0.7, 0.8]\,, \quad k_j/n = 0.75$
$\quad\quad\quad (n^{(0)} \in [1, 8]) \quad\quad\quad\quad\quad (n = 16)$

Case (ii): $\quad y_j^{(0)} \in [0.2, 0.3]\,, \quad k_j/n = 1$
$\quad\quad\quad (n^{(0)} \in [1, 8]) \quad\quad\quad\quad\quad (n = 16)$

Prior-Data Conflict
Generalized iLUCK-models
"Strong Happiness"

Simple Example
Parameter Sets
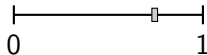Direct vs. Sequential Updating

Generalized iLUCK-models: Simple Example

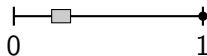Case (i): $y_j^{(0)} \in [0.7, 0.8]$, $k_j/n = 0.75$
              $(n^{(0)} \in [1, 8])$ $(n = 16)$


⟶ $y_j^{(1)} \in [0.73, 0.76]$
       $(n^{(0)} \in [17, 24])$

Case (ii): $y_j^{(0)} \in [0.2, 0.3]$, $k_j/n = 1$
              $(n^{(0)} \in [1, 8])$ $(n = 16)$

Prior-Data Conflict
Generalized iLUCK-models
"Strong Happiness"

Simple Example
Parameter Sets
Direct vs. Sequential Updating

Generalized iLUCK-models: Simple Example

Case (i):
$$y_j^{(0)} \in [0.7, 0.8], \quad k_j/n = 0.75$$
$$(n^{(0)} \in [1,8]) \qquad (n = 16)$$

➡ $$y_j^{(1)} \in [0.73, 0.76]$$
$$(n^{(0)} \in [17, 24])$$

Case (ii):
$$y_j^{(0)} \in [0.2, 0.3], \quad k_j/n = 1$$
$$(n^{(0)} \in [1,8]) \qquad (n = 16)$$

➡ $$y_j^{(1)} \in [0.73, 0.96]$$
$$(n^{(0)} \in [17, 24])$$

Prior-Data Conflict
Generalized iLUCK-models
"Strong Happiness"

Simple Example
Parameter Sets
Direct vs. Sequential Updating

Generalized iLUCK-models: Simple Example

Case (i): $y_j^{(0)} \in [0.7, 0.8]$, $k_j/n = 0.75$
              $(n^{(0)} \in [1, 8])$ $(n = 16)$

→ $y_j^{(1)} \in [0.73, 0.76]$
  $(n^{(0)} \in [17, 24])$

Case (ii): $y_j^{(0)} \in [0.2, 0.3]$, $k_j/n = 1$
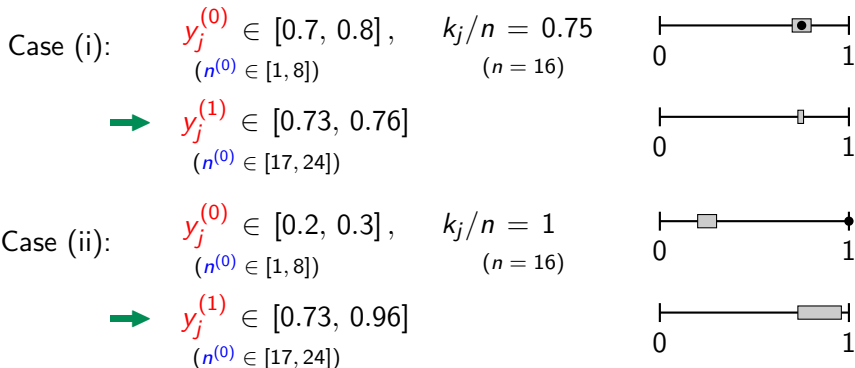 $(n^{(0)} \in [1, 8])$ $(n = 16)$

→ $y_j^{(1)} \in [0.73, 0.96]$
  $(n^{(0)} \in [17, 24])$
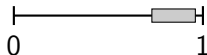
Generalized iLUCK-models lead to cautious inferences
if, and only if, caution is needed.

Prior-Data Conflict
Generalized iLUCK-models
"Strong Happiness"

Simple Example
Parameter Sets
Direct vs. Sequential Updating
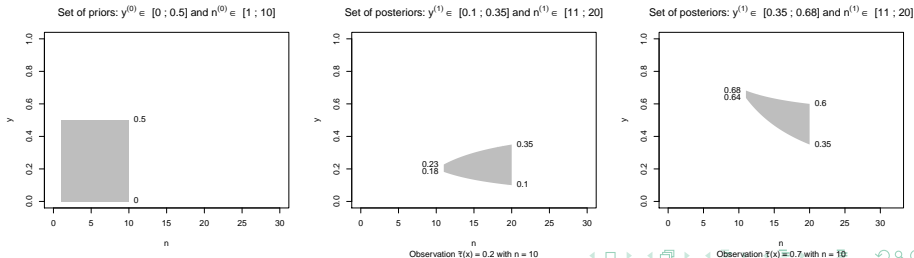
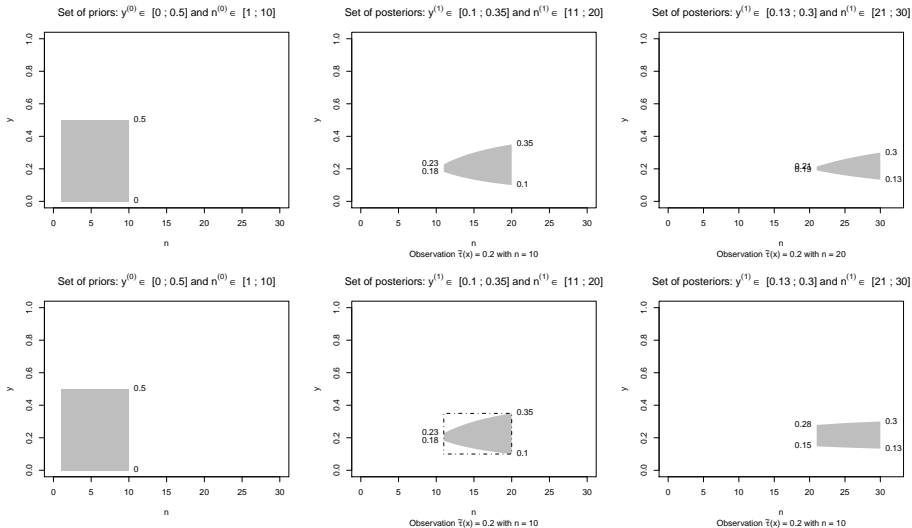# Generalized iLUCK-models: Parameter Sets

Prior parameter set $y^{(0)} \in \mathcal{Y}^{(0)} \times n^{(0)} \in \mathcal{N}^{(0)}$ is rectangular, but posterior parameter set

$$\left\{ (y^{(1)}, n^{(1)}) \,\middle|\, y^{(1)} = \frac{n^{(0)} y^{(0)} + \tau(x)}{n^{(0)} + n}, \; n^{(1)} = n^{(0)} + n, \right.$$

$$\left. y^{(0)} \in \mathcal{Y}^{(0)}, \; n^{(0)} \in \mathcal{N}^{(0)} \right\}$$

is not rectangular anymore!

Prior-Data Conflict
Generalized iLUCK-models
"Strong Happiness"
Simple Example
Parameter Sets
Direct vs. Sequential Updating

Generalized iLUCK-models: Parameter Sets

Prior parameter set $y^{(0)} \in \mathcal{Y}^{(0)} \times n^{(0)} \in \mathcal{N}^{(0)}$ is rectangular, but posterior parameter set

$$\left\{ (y^{(1)}, n^{(1)}) \,\middle|\, y^{(1)} = \frac{n^{(0)} y^{(0)} + \tau(x)}{n^{(0)} + n}, \; n^{(1)} = n^{(0)} + n, \right.$$
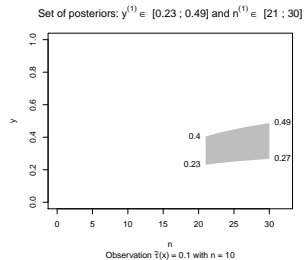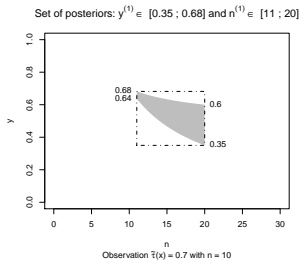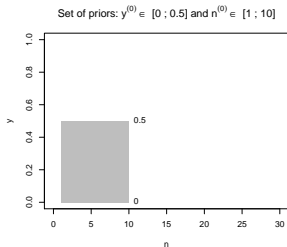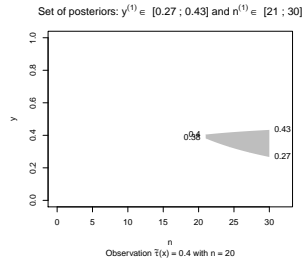$$\left. y^{(0)} \in \mathcal{Y}^{(0)}, \, n^{(0)} \in \mathcal{N}^{(0)} \right\}$$

is not rectangular anymore!

Set of priors: $y^{(0)} \in [0\,;0.5]$ and $n^{(0)} \in [1\,;10]$

Set of posteriors: $y^{(1)} \in [0.1\,;0.35]$ and $n^{(1)} \in [11\,;20]$

Observation $\bar{\tau}(x) = 0.2$ with $n = 10$

Set of posteriors: $y^{(1)} \in [0.35\,;0.68]$ and $n^{(1)} \in [11\,;20]$

Observation $\bar{\tau}(x) = 0.7$ with $n = 10$

Prior-Data Conflict
Generalized iLUCK-models
"Strong Happiness"
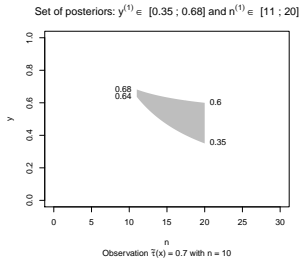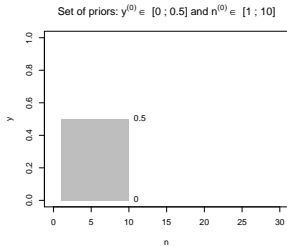
Simple Example
Parameter Sets
Direct vs. Sequential Updating

Generalized iLUCK-models: Direct vs. Sequential

Generalized iLUCK-models: Direct vs. Sequential

Prior-Data Conflict
Generalized iLUCK-models
"Strong Happiness"

Simple Example
Parameter Sets
Direct vs. Sequential Updating

Generalized iLUCK-models: Direct vs. Sequential

with "rectangularization"...

- ▶ different intervals for y in case of prior-data conflict

Prior-Data Conflict
Generalized iLUCK-models
"Strong Happiness"

Simple Example
Parameter Sets
Direct vs. Sequential Updating

Generalized iLUCK-models: Direct vs. Sequential

with "rectangularization"...

- different intervals for y in case of prior-data conflict
- sufficient statistic $\tau(x)$ not sufficient anymore?!?

Prior-Data Conflict
Generalized iLUCK-models
"Strong Happiness"

Simple Example
Parameter Sets
Direct vs. Sequential Updating

Generalized iLUCK-models: Direct vs. Sequential

with "rectangularization"...

- ▶ different intervals for y in case of prior-data conflict
- ▶ sufficient statistic $\tau(x)$ not sufficient anymore?!?
- ▶ more information (order) ➡ more imprecise posterior!

Prior-Data Conflict
Generalized iLUCK-models
"Strong Happiness"

Simple Example
Parameter Sets
Direct vs. Sequential Updating

Generalized iLUCK-models: Direct vs. Sequential

with "rectangularization". . .

- ▶ different intervals for y in case of prior-data conflict
- ▶ sufficient statistic $\tau(x)$ not sufficient anymore?!?
- ▶ more information (order) ➡ more imprecise posterior!

In general: $\mathcal{Y}_{\text{dir}}^{(2)} \subseteq \mathcal{Y}_{\text{seq}}^{(2)}$

if no pdc in both steps: $\qquad\qquad \mathcal{Y}_{\text{dir}}^{(2)} = \mathcal{Y}_{\text{seq}}^{(2)}$

if pdc in both steps in same direction: $\qquad \mathcal{Y}_{\text{dir}}^{(2)} = \mathcal{Y}_{\text{seq}}^{(2)}$

if neither of the former: $\qquad\qquad\quad \mathcal{Y}_{\text{dir}}^{(2)} \subset \mathcal{Y}_{\text{seq}}^{(2)}$

especially: if prior-data conflict 'on the way', then $\mathcal{Y}_{\text{dir}}^{(2)} \subset \mathcal{Y}_{\text{seq}}^{(2)}$!
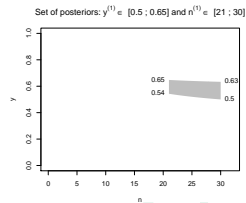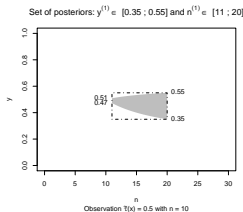
Prior-Data Conflict
Generalized iLUCK-models
"Strong Happiness"
Simple Example
Parameter Sets
Direct vs. Sequential Updating

LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

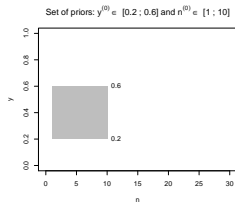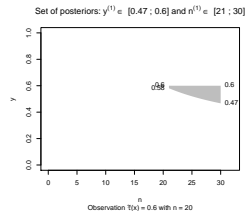## Generalized iLUCK-models: Direct vs. Sequential

If $\mathcal{Y}_{\text{dir}}^{(2)} \neq \mathcal{Y}_{\text{seq}}^{(2)}$, there was prior-data conflict 'on the way'.

(but: no need to use $\mathcal{Y}_{\text{seq}}^{(2)}$ for inference if you think, e.g,
that the sequence of data is exchangeable nevertheless.)

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Prior-Data Conflict
Generalized iLUCK-models
"Strong Happiness"

Simple Example
Parameter Sets
Direct vs. Sequential Updating
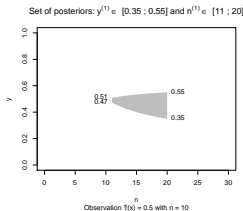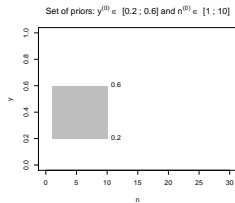
Generalized iLUCK-models: Direct vs. Sequential

If $\mathcal{Y}_{\text{dir}}^{(2)} \neq \mathcal{Y}_{\text{seq}}^{(2)}$, there was prior-data conflict 'on the way'.

(but: no need to use $\mathcal{Y}_{\text{seq}}^{(2)}$ for inference if you think, e.g,
that the sequence of data is exchangeable nevertheless.)

Prior-Data Conflict
Generalized iLUCK-models
"Strong Happiness"

Simple Example
Parameter Sets
Direct vs. Sequential Updating

Generalized iLUCK-models: Interval Lengths for 3 Steps

Data situation: Bernoulli sampling (observe 0 or 1)
Start with $\mathcal{Y}^{(0)} = [0, 1]$

Compare sequences:
- A $(1, 0, 1)$
- B $(1, 1, 0)$
- C $(0, 1, 1)$
- D $([2/3, n = 3])$ (direct)

Prior-Data Conflict
Generalized iLUCK-models
"Strong Happiness"

Simple Example
Parameter Sets
Direct vs. Sequential Updating

Generalized iLUCK-models: Interval Lengths for 3 Steps

Data situation: Bernoulli sampling (observe 0 or 1)
Start with $\mathcal{Y}^{(0)} = [0, 1]$

Compare sequences:

A $(1, 0, 1)$

B $(1, 1, 0)$

C $(0, 1, 1)$

D $([2/3, n = 3])$ (direct)

A* $=$ C* $([1/2, n = 2], 1)$

B* $([1, n = 2], 0)$

Prior-Data Conflict
Generalized iLUCK-models
"Strong Happiness"

Simple Example
Parameter Sets
Direct vs. Sequential Updating

Generalized iLUCK-models: Interval Lengths for 3 Steps

Data situation: Bernoulli sampling (observe 0 or 1)
Start with $\mathcal{Y}^{(0)} = [0, 1]$
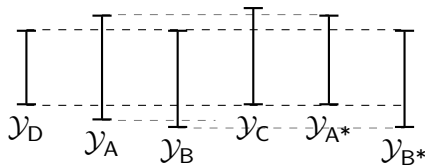
Compare sequences:

A $(1, 0, 1)$

B $(1, 1, 0)$

C $(0, 1, 1)$

D $([2/3, n = 3])$ (direct)

A* = C* $([1/2, n = 2], 1)$

B* $([1, n = 2], 0)$

Interval length for y: "$|\mathcal{Y}|$" =: l
$$l_A > l_B, \ l_A > l_C$$

Prior-Data Conflict
Generalized iLUCK-models
"Strong Happiness"

Simple Example
Parameter Sets
Direct vs. Sequential Updating

Generalized iLUCK-models: Interval Lengths for 3 Steps

Data situation: Bernoulli sampling (observe 0 or 1)
Start with $\mathcal{Y}^{(0)} = [0, 1]$

Compare sequences:
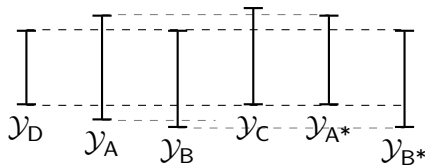
A $(1, 0, 1)$

B $(1, 1, 0)$

C $(0, 1, 1)$

D $([2/3, n = 3])$ (direct)

$A^* = C^*$ $([1/2, n = 2], 1)$

B* $([1, n = 2], 0)$

Interval length for y: "$|\mathcal{Y}|$" $=: l$
$$l_A > l_B, \; l_A > l_C$$

$A^{**} = B^{**}$ $(1, [1/2, n = 2]) \mathrel{\hat{=}} D$ (if $\overline{n}^{(0)} \geq 1$)

C^{**} $(0, [1, n = 2]) \mathrel{\hat{=}} C$

LMU
LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

"Strong Happiness"

Data situation: Bernoulli sampling (observe 0 or 1).

Idea: Choose sample size n_1 such that $I^{(1)}$ low enough under certain threshold I such that – even if we got another sample n_2 in conflict to $\mathcal{Y}^{(1)}$ – $I^{(2)}$ is still below I!

With a sample of such a size n_1 we attain "strong happiness", because whatever we would see as a following sample, we would never get more imprecise than I!

(Is only possible because degree of prior-data conflict is bounded due to Bernoulli sampling!).

"Strong Happiness"

$$I^{(2)} = \frac{\overline{n}^{(0)} I^{(0)}}{\overline{n}^{(0)} + n_1 + n_2} + \frac{\overline{n}^{(0)} - \underline{n}^{(0)}}{\overline{n}^{(0)} + n_1 + n_2} \left(\frac{n_1}{\underline{n}^{(0)} + n_1} \Delta\left(\frac{c_1}{n_1}, \mathcal{Y}^{(0)} \right) \right.$$

$$\left. + \frac{n_2}{\underline{n}^{(0)} + n_1 + n_2} \Delta\left(\frac{c_2}{n_2}, \mathcal{Y}^{(1)} \right) \right)$$

$$\overset{!}{\leq} I \quad \forall \, (k_2, n_2)$$

"Strong Happiness"

$$I^{(2)} = \frac{\overline{n}^{(0)} I^{(0)}}{\overline{n}^{(0)} + n_1 + n_2} + \frac{\overline{n}^{(0)} - \underline{n}^{(0)}}{\overline{n}^{(0)} + n_1 + n_2} \left(\frac{n_1}{\underline{n}^{(0)} + n_1} \Delta\left(\frac{c_1}{n_1}, \mathcal{Y}^{(0)} \right) \right.$$
$$\left. + \frac{n_2}{\underline{n}^{(0)} + n_1 + n_2} \Delta\left(\frac{c_2}{n_2}, \mathcal{Y}^{(1)} \right) \right)$$
$$\overset{!}{\leq} I \quad \forall \, (k_2, \, n_2)$$

If data from second sample in conflict with $\mathcal{Y}^{(1)}$ (and conflict strong enough), then imprecision $I^{(2)}$ should first increase and then decrease in n_2.

➡️ find n_2 that maximizes $I^{(2)}$ for maximal possible conflict with current $\mathcal{Y}^{(1)}$, plug into formula for $I^{(2)}$ and give n_1 as a function of $I^{(0)}$ (or $\mathcal{Y}^{(0)}$) and I (and $\mathcal{N}^{(0)}$).

Plans, Further Ideas

- ▶ Try to attain Strong Happiness by sorting out formula
- ▶ Write programs for Strong Happiness
- ▶ Investigate interval lengths in general situation (not only Bernoulli sampling)
- ▶ Focus considerations on certain event of interest instead of $\mathcal{Y}$?
- ▶ . . .