Software for Generalized iLUCK-models

Gero Walter

Institut für Statistik
Ludwig-Maximilians-Universität München

September 6th, 2010

- Generalized Bayesian Inference
  - General Idea
  - Conjugate Priors
  - Example: Dirichlet-Multinomial Model / IDM
- Generalized iLUCK-models
  - Definition
  - Examples
- Implementation
  - The R Project
  - Imperative vs. Object Oriented
  - R-Package `luck`

Generalized Bayesian Inference – General Idea

Bayesian Inference on some parameter θ:

prior knowledge on θ + data x → updated knowledge on θ

Generalized Bayesian Inference – General Idea

Bayesian Inference on some parameter θ:

prior knowledge on θ	+	data x	→	updated knowledge on θ
prior distribution $p(\theta)$	+	likelihood $f(x \mid \theta)$	→	posterior distribution $p(\theta \mid x)$

Generalized Bayesian Inference – General Idea

Bayesian Inference on some parameter θ:

prior knowledge on θ	+	data x	→	updated knowledge on θ
prior distribution $p(\theta)$	+	likelihood $f(x \mid \theta)$	→	posterior distribution $p(\theta \mid x)$
set of priors	+	likelihood	→	**set of** posteriors

Generalized Bayesian Inference – General Idea

Bayesian Inference on some parameter θ:

prior knowledge on θ	+	data x	→	updated knowledge on θ
prior distribution $p(\theta)$	+	likelihood $f(x \mid \theta)$	→	posterior distribution $p(\theta \mid x)$
set of priors	+	likelihood	→	**set of** posteriors

Tractability: use **conjugate** priors

→ choose $p(\theta)$ such that $p(\theta \mid x)$ is from same parametric class

↳ update only parameters!

Conjugate Priors

General result on construction of conjugate priors:

$X \overset{iid}{\sim}$ 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

General result on construction of conjugate priors:
$X \stackrel{iid}{\sim}$ 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 Priors

General result on construction of conjugate priors:
$X \overset{iid}{\sim}$ 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 $y^{(1)} = \frac{n^{(0)}}{n^{(0)} + n} \cdot y^{(0)} + \frac{n}{n^{(0)} + n} \cdot \frac{1}{n}\tau(x)$ and $n^{(1)} = n^{(0)} + n$.

Conjugate Priors — Interpretation of $y^{(0)}$ and $n^{(0)}$

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

$y^{(0)}$: **“main prior parameter”**

$n^{(0)}$: **“prior strength”** or **“pseudocounts”**

- for samples from a $\mathrm{N}(\mu, 1)$, $p(\mu)$ is a $\mathrm{N}(y^{(0)}, \frac{1}{n^{(0)}})$
- for samples from a $\mathrm{Po}(\lambda)$, $p(\lambda)$ is a $\mathrm{Ga}(n^{(0)} y^{(0)}, n^{(0)})$
  $\rightarrow$ $\mathbb{E}[\lambda] = y^{(0)}$, $\mathbb{V}(\lambda) = \frac{y^{(0)}}{n^{(0)}}$
- for samples from a $\mathrm{M}(\boldsymbol{\theta})$, $p(\boldsymbol{\theta})$ is a $\mathrm{Dir}(n^{(0)}, \mathbf{y}^{(0)})$

Example: Dirichlet-Multinomial-Model

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

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

Example: Dirichlet-Multinomial-Model

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

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

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

Example: Dirichlet-Multinomial-Model

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

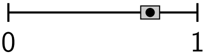
$$\mathbb{E}[\theta_j] = \frac{\alpha_j}{\sum \alpha_i} =: y_j^{(0)} \quad \mathbb{V}(\theta_j) = \frac{\mathbb{E}[\theta_j](1 - \mathbb{E}[\theta_j])}{\sum \alpha_i + 1} = \frac{y_j^{(0)}(1 - y_j^{(0)})}{n^{(0)} + 1}$$

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

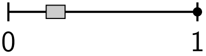
$$y_j^{(1)} = \frac{n^{(0)}}{n^{(0)} + n} \cdot y_j^{(0)} + \frac{n}{n^{(0)} + n} \cdot \frac{k_j}{n} \qquad n^{(1)} = n^{(0)} + n$$

Example: Imprecise Dirichlet Model (IDM)

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

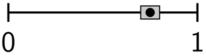


Case (ii): $y_j^{(0)} \in [0.2, 0.3]\,,$ $(n^{(0)} = 8)$ $\qquad k_j/n = 1$ $(n = 16)$

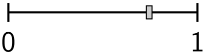


Example: Imprecise Dirichlet Model (IDM)

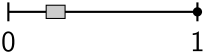
Case (i): $y_j^{(0)} \in [0.7,\, 0.8]$, $(n^{(0)} = 8)$ $\quad k_j/n = 0.75$ $(n = 16)$



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



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



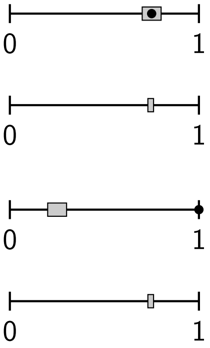
Example: Imprecise Dirichlet Model (IDM)

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

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

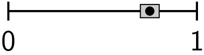
Case (ii): $y_j^{(0)} \in [0.2, 0.3]$, $(n^{(0)} = 8)$ $\quad k_j/n = 1$ $(n = 16)$

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

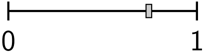


Example: Imprecise Dirichlet Model (IDM)

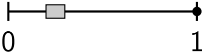
Case (i): $y_j^{(0)} \in [0.7, 0.8]\,,$ $(n^{(0)} = 8)$ $\quad k_j/n = 0.75$ $(n = 16)$



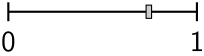
→ $y_j^{(1)} \in [0.73, 0.76]$ $(n^{(0)} = 24)$



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



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



$$\left[\ \mathbb{V}(\theta_j) \in [0.0178, 0.0233] \longrightarrow \mathbb{V}(\theta_j|\mathbf{k}) \in [0.0072, 0.0078]\ \right]$$

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

Generalized iLUCK-models

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

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

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
$y^{(0)} \in \mathcal{Y}^{(0)} \times n^{(0)} \in \mathcal{N}^{(0)}$

Generalized iLUCK-models

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

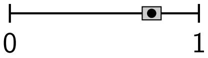
1. use conjugate priors from general construction method (prior parameters $y^{(0)}$, $n^{(0)}$)
2. construct sets of priors via sets of parameters $y^{(0)} \in \mathcal{Y}^{(0)} \times n^{(0)} \in \mathcal{N}^{(0)}$
3. set of posteriors $\hat{=}$ set of (element-wise) updated priors
   ➡ still easy to handle: described as set of $(y^{(1)}, n^{(1)})$'s

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

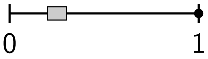
$$n^{(1)} = n^{(0)} + n$$

Generalized iLUCK-models — "Generalized IDM"

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



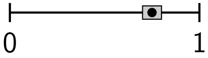
Case (ii): $y_j^{(0)} \in [0.2, 0.3]\,, \quad (n^{(0)} \in [1, 8])$ $\quad k_j/n = 1 \quad (n = 16)$



Generalized iLUCK-models — "Generalized IDM"

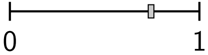
Case (i): $y_j^{(0)} \in [0.7, 0.8]\,,$ $\quad k_j/n = 0.75$

$(n^{(0)} \in [1, 8])$ $\quad (n = 16)$



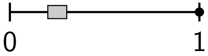
→ $y_j^{(1)} \in [0.73, 0.76]$

$(n^{(0)} \in [17, 24])$



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

$(n^{(0)} \in [1, 8])$ $\quad (n = 16)$



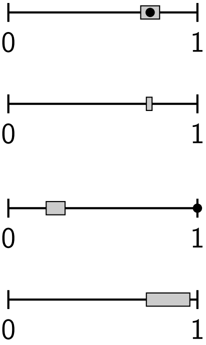
Generalized iLUCK-models — "Generalized IDM"

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

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

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

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



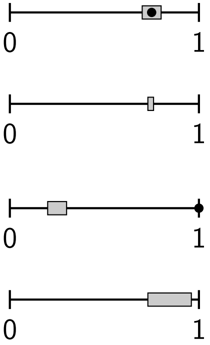
Generalized iLUCK-models — "Generalized IDM"

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

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

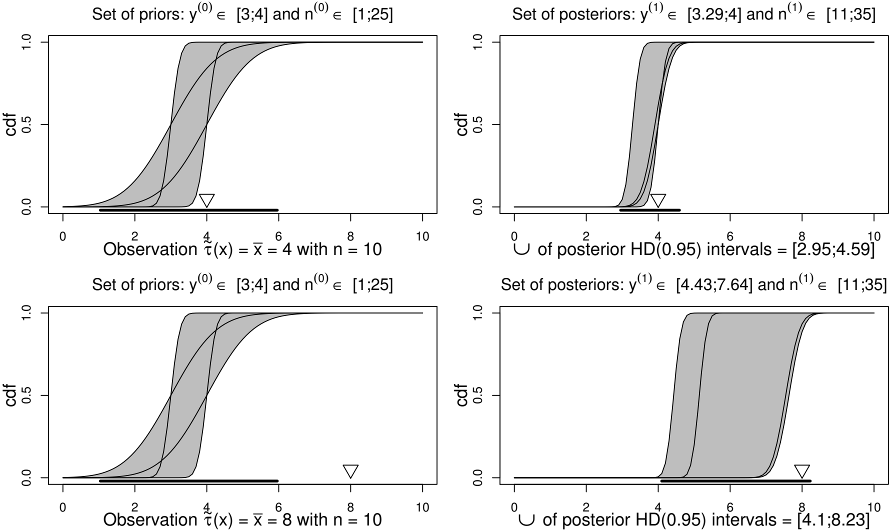
Case (ii): $y_j^{(0)} \in [0.2, 0.3]\,,$ $(n^{(0)} \in [1, 8])$ $\quad k_j/n = 1$ $(n = 16)$

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

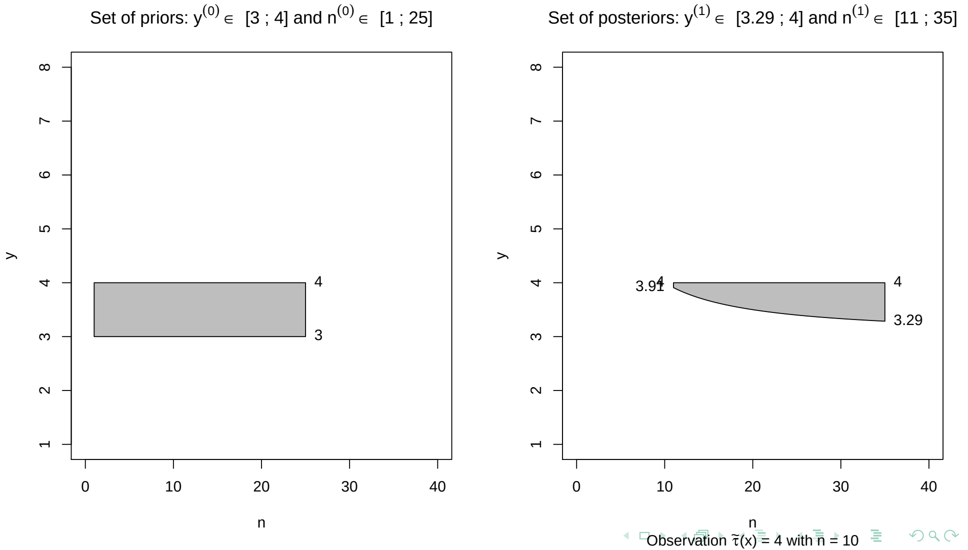


Generalized iLUCK-models lead to cautious inferences if, and only if, caution is needed.

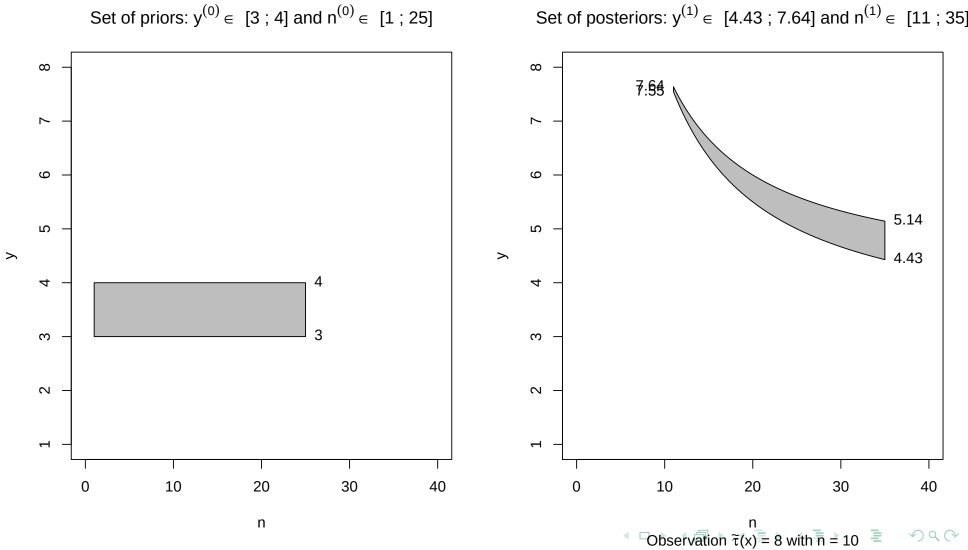
Generalized iLUCK-models: $X_i \sim \mathrm{N}(\mu, 1)$



Parameter Sets



Parameter Sets



The R project for Statistical Computing

- not just a (statistical) software package,
  rather a full-grown programming language
- open source implementation of the
  (award-winning) S language
- extremely widespread in universitary research
  (reference implementation of new methods are often in R)
- extensions providing additional functionality
  can be made readily available as “packages”
- can be linked with LaTeX (package Sweave)
- can be used as imperative or as object oriented language

Imperative vs. Object Oriented Programming

imperative: do this, then that

- ↪ functions (on arguments) that produce some output

object oriented: create 'objects' (that mirror real-world concepts), do things with them

- ↪ blueprints for objects are called 'classes', defining slots (properties) and methods (what you can do with it)

objects created according to a blueprint are called an 'instance'

Imperative vs. Object Oriented Programming

imperative: do this, then that

→ functions (on arguments) that produce some output

object oriented: create 'objects' (that mirror real-world concepts), do things with them

→ blueprints for objects are called 'classes', defining slots (properties) and methods (what you can do with it)

objects created according to a blueprint are called an 'instance'

example: brewery administrating their customers' orders

class: `Order` (slots: `name`, `bottles`, `variety`, . . .
methods: `writeInvoice`, . . .)

instances: order no. 41 (`name`=M. T., `bottles`=6, `variety`=Archangel)
order no. 42 (`name`=G. W., `bottles`=12, `variety`=Vice)
⋮

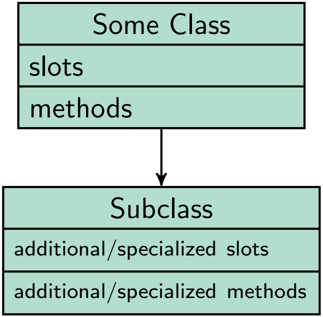
Object Oriented Programming — Class hierarchies

general
concept

Some Class
slots
methods

Object Oriented Programming — Class hierarchies

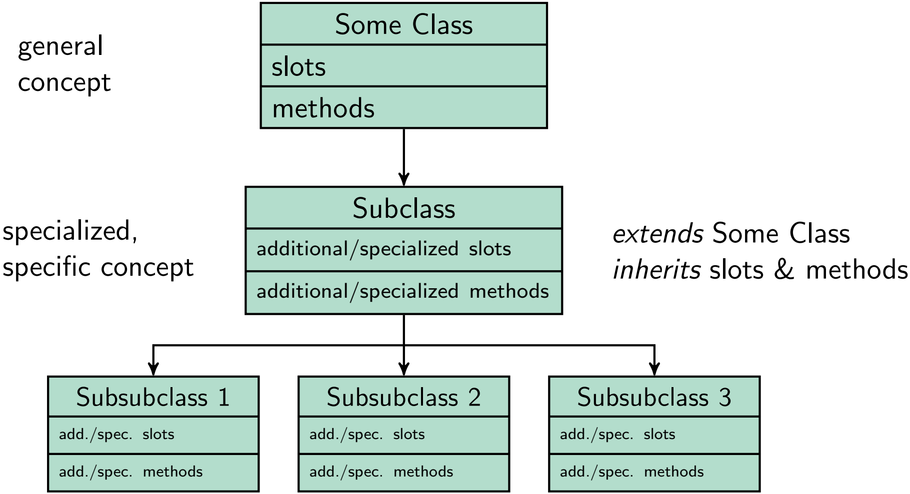
general concept



specialized, specific concept

extends Some Class
inherits slots & methods

Object Oriented Programming — Class hierarchies



R-package `luck` — General Concept

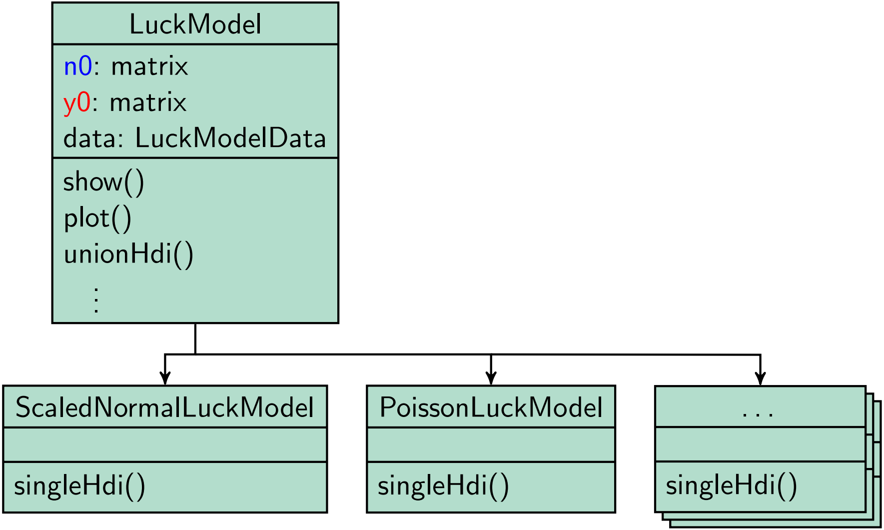
Mirror the hierarchy of general formulation (ψ, $\tau(x)$) to specific for a certain sample distribution ($X_i \sim \mathrm{N}(\mu, 1)$: $\psi = \mu$, $\tau(x) = \sum x_i$).

"Downward compatibility": $y^{(0)}$ and/or $n^{(0)}$ may be $\in \mathbb{R}$.

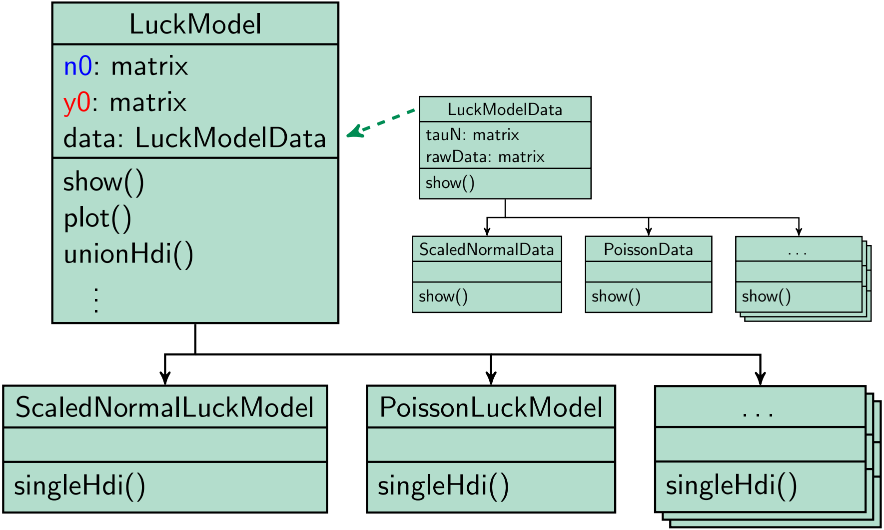
Provide basic utilities for working with generalized iLUCK-models:

- easy creation of prior distribution objects and data (compatibility check)
- "translation" into standard parametrizations
- visualize prior and posterior parameter sets
- easy general-purpose optimization over prior and posterior parameter sets
- allow easy implementation of subclasses for specific sample distributions by other users
- ...?

R-package `luck` — Class Structure



R-package luck — Class Structure



```
> ex1 <- LuckModel(n0=c(1,25), y0=c(3,4))
> ex1
generalized iLUCK model with prior parameter set:
  lower n0 = 1 upper n0 = 25
  lower y0 = 3 upper y0 = 4
 giving a main parameter prior imprecision of 1
> data1 <- LuckModelData(tau=40, n=10)
> data1
data object with sample statistic tau(x) = 40 and sample size n = 10
> ex2 <- ScaledNormalLuckModel(n0=c(1,25), y0=c(3,4), data=rnorm(mean=4
sd=1, n=10))
> ex2
generalized iLUCK model for inference from scaled normal data
with prior parameter set:
  lower n0 = 1 upper n0 = 25
  lower y0 = 3 upper y0 = 4
 giving a main parameter prior imprecision of 1
corresponding to a set of normal priors
 with means in [ 3 ; 4 ] and variances in [ 0.04 ; 1 ]
and ScaledNormalData object containing data of sample size 10
 with mean 4.170152 and variance 0.6234904 .
```

References

- E. Quaeghebeur and G. de Cooman. Imprecise probability models for inference in exponential families. In F.G. Cozman, R. Nau, and T. Seidenfeld, eds., *ISIPTA '05*, p. 287–296, 2005.
- R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, 2010. See also: `http://www.R-project.org`, `http://www.cran.R-project.org`
- G. Walter and T. Augustin. Imprecision and prior-data conflict in generalized Bayesian inference. *Journal of Statistical Theory and Practice*, 3 (Special Issue on Imprecision), p. 255–271, 2009.