Software for Generalized Bayesian Inference

An object-oriented R implementation of generalized iLUCK-models

Gero Walter

Department of Statistics
Ludwig-Maximilians-University Munich

gero.walter@stat.uni-muenchen.de
www.stat.uni-muenchen.de/~walter

July 17th, 2009

Generalized Bayesian inference, prior-data conflict, Imprecise Dirichlet Model (IDM), R, Software

- iLUCK-models are a generalization of the IDM to arbitrary sample distributions that form a so-called exponential family.
- iLUCK-models offer a general, manageable and powerful calculous for Bayesian inference with sets of priors.
- However, they are insensitive to *prior-data conflict* and thus do not use the full expressive power of imprecise probability.
- Generalized iLUCK-models extend iLUCK-models such that *prior-data conflict* is accounted for.
- A basic framework for display and updating of generalized iLUCK-models is implemented in the statistical software environment R.
- The framework can be easily extended to give inferences for arbitrary sample distributions.

Background

Examples

Software Environment

Implementation

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, i.e.
choose $p(\theta)$ such that $p(\theta \mid x)$ is from the same parametric class
→ update only parameters!

LUCK-models: Single Conjugate Prior

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

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

(includes Binomial, Multinomial, Normal, Poisson,... distr.)

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

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

$y^{(0)}$: **"main prior parameter"**

- for samples from a $\mathrm{N}(\mu, 1)$, $p(\mu)$ is a $\mathrm{N}(y^{(0)}, \frac{1}{n^{(0)}})$
- for samples from a $\mathrm{M}(\boldsymbol{\theta})$, $p(\boldsymbol{\theta})$ is a $\mathrm{Dir}(n^{(0)}, y^{(0)})$ $(y_j^{(0)} = t_j \hat{=}$ prior probability for class j, $n^{(0)} = s)$

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

with $\tilde{\tau}(x) =: \frac{1}{n}\tau(x)$: $\quad\left[\ \tau(x) = \sum_{i=1}^{n} \tau(x_i)\ \right]$

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

sets of LUCK-models – iLUCK-models

iLUCK-model:

(inspired by Quaeghebeur and de Cooman, 2005)

vary $y^{(0)}$ in $\mathcal{Y}^{(0)}$ $\quad [\ \mathcal{Y}^{(0)}$ convex $]$,
i.e. allow for ambiguity on the main prior parameter

➡ prior credal set contains *all finite convex mixtures* of $p(\theta)$s with $y^{(0)} \in \mathcal{Y}^{(0)}$

➡ posterior credal set easy to calculate:
all finite convex mixtures of $p(\theta \mid x)$s with

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

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

⚠ unfavourable behavior in case of prior–data conflict! ⚠

Prior-Data Conflict

Situation in which *informative prior beliefs*
and *trusted data* (no outliers, etc.) are in conflict

Example: (Walley 1991)

$$\begin{array}{r|ccl} \text{Data :} & X & \sim & \mathrm{N}(\vartheta, 1) \\ \text{conjugate prior:} & \vartheta & \sim & \mathrm{N}(\mu, 1) \\ \hline \text{posterior:} & \vartheta \mid x & \sim & \mathrm{N}\left(\dfrac{\mu + x}{2}, \dfrac{1}{2}\right) \end{array}$$

- Case (i): $\mu = 5.5$, $x = 6.5$ $\Longrightarrow \vartheta \sim \mathrm{N}(6, \frac{1}{2})$
- Case (ii): $\mu = 3.5$, $x = 8.5$ $\Longrightarrow \vartheta \sim \mathrm{N}(6, \frac{1}{2})$

⚠ In Bayesian analysis all inference is based only on the posterior!

Prior-Data Conflict in iLUCK-models

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

→ Posterior imprecision does not depend on $\tau(x)$!

For *any* sample of size n, posterior imprecision is reduced by the same amount!

sets of LUCK-models – Generalized iLUCK-models

generalized iLUCK-model:

vary $y^{(0)}$ in $\mathcal{Y}^{(0)}$ **and** $n^{(0)}$ in $\mathcal{N}^{(0)}$, i.e. weigh prior information $\mathcal{Y}^{(0)}$ and sample information $\tilde{\tau}(x)$ more flexible in

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

- ➡ prior credal set contains *all finite convex mixtures* of $p(\theta)$s with $y^{(0)} \in \mathcal{Y}^{(0)}$ **and** $n^{(0)} \in \mathcal{N}^{(0)}$
- ➡ posterior credal set still quite easy to calculate: *all finite convex mixtures* of $p(\theta \mid x)$s with

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

Defines a general framework for two models proposed by Walley (1991) for Binomial and scaled Normal data.

Example: samples from a N$(\mu, 1)$

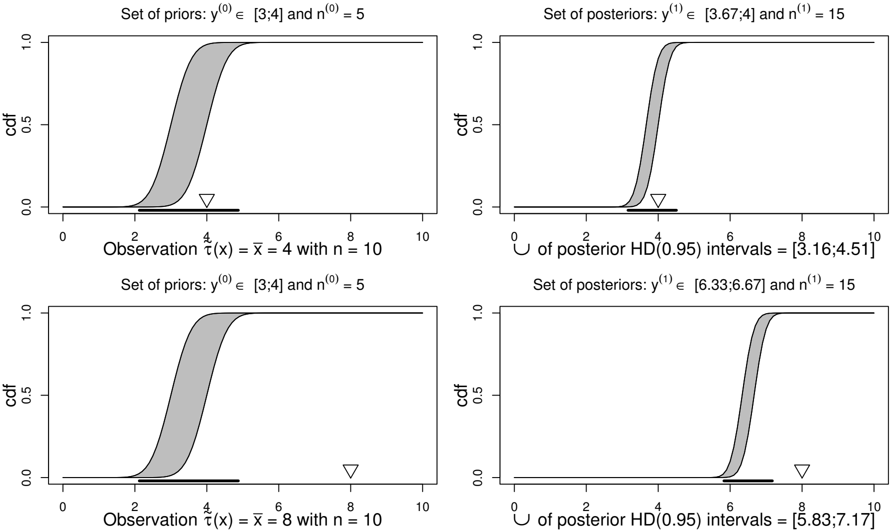
likelihood $$p(x \mid \theta) \propto \exp\left\{ -\frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^2 \right\}$$

$$\propto \exp\left\{ \underbrace{\mu}_{=\psi} \underbrace{\sum_{i=1}^{n} x_i}_{\tau(x)} - \underbrace{\frac{\mu^2}{2}}_{=\mathbf{b}(\psi)} \right\}$$

conjugate prior $$p(\theta) \propto \exp\left\{ n^{(0)} \left[\langle \mu, y^{(0)} \rangle - \frac{\mu^2}{2} \right] \right\}$$

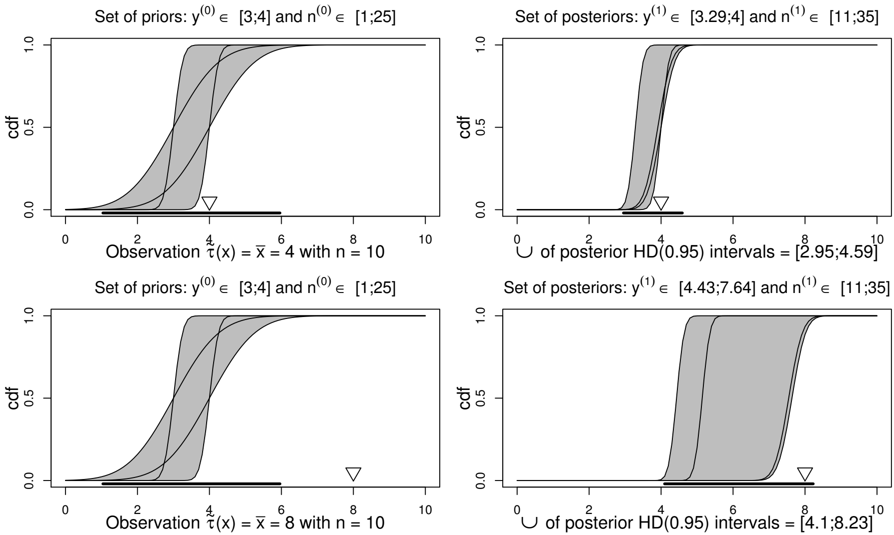
$$\propto \exp\left\{ -\frac{n^{(0)}}{2}\left(\mu - y^{(0)}\right)^2 \right\} \propto \mathsf{N}\left(y^{(0)}, \frac{1}{n^{(0)}}\right)$$

iLUCK-model



Set of priors: y(0) ∈ [3;4] and n(0) = 5
cdf
1.0
0.5
0.0
0
2
4
6
8
10
Observation τ̃(x) = x̄ = 4 with n = 10
Set of posteriors: y(1) ∈ [3.67;4] and n(1) = 15
cdf
∪ of posterior HD(0.95) intervals = [3.16;4.51]
Set of priors: y(0) ∈ [3;4] and n(0) = 5
cdf
Observation τ̃(x) = x̄ = 8 with n = 10
Set of posteriors: y(1) ∈ [6.33;6.67] and n(1) = 15
cdf
∪ of posterior HD(0.95) intervals = [5.83;7.17]

generalized iLUCK-model

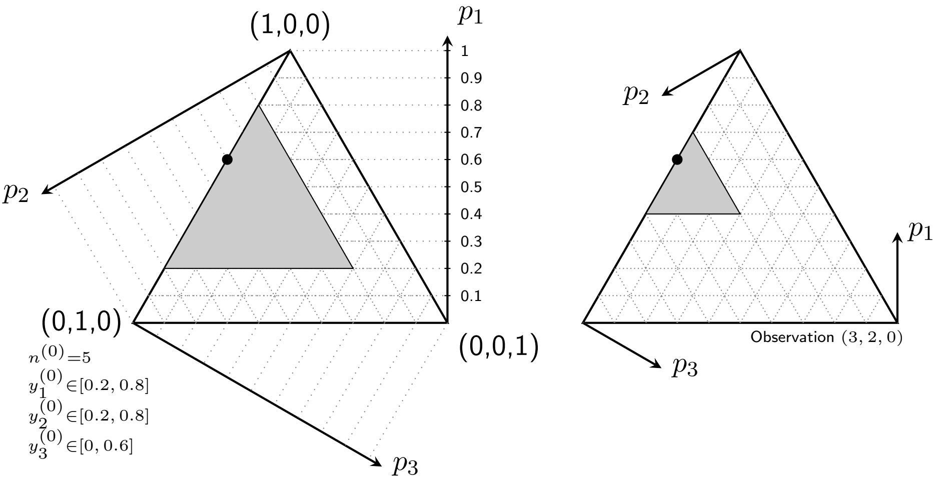


Example: samples from a M(θ)

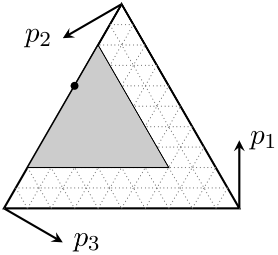
→ $\theta \sim \text{Dir}(n^{(0)}, \boldsymbol{y}^{(0)})$ ↔ **Imprecise Dirichlet Model (IDM)**

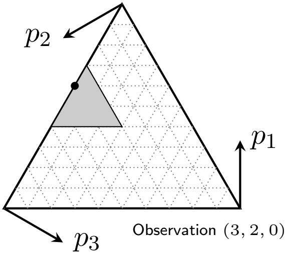
($y_j^{(0)} = t_j \hat{=}$ prior probability for class j, $n^{(0)} = s$)

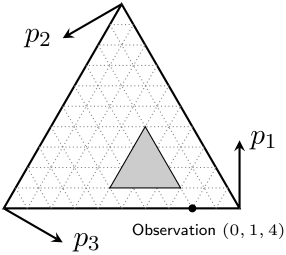
→ Walley (*JRSS*, 1991), Bernard (*IJAR Special Issue*, 2009)



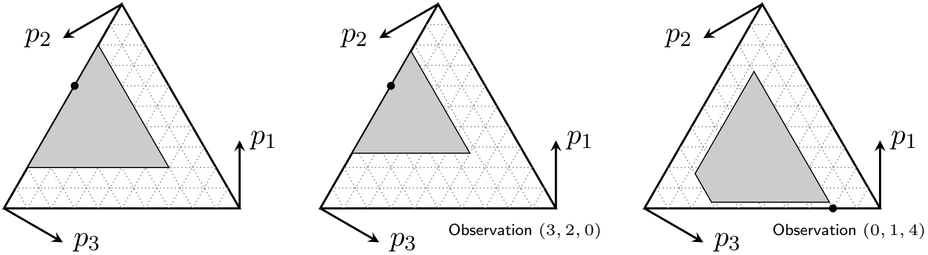
iLUCK-model ⟷ IDM







generalized iLUCK-model ⟷ generalized IDM



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)

object-oriented: create 'objects', do things with them

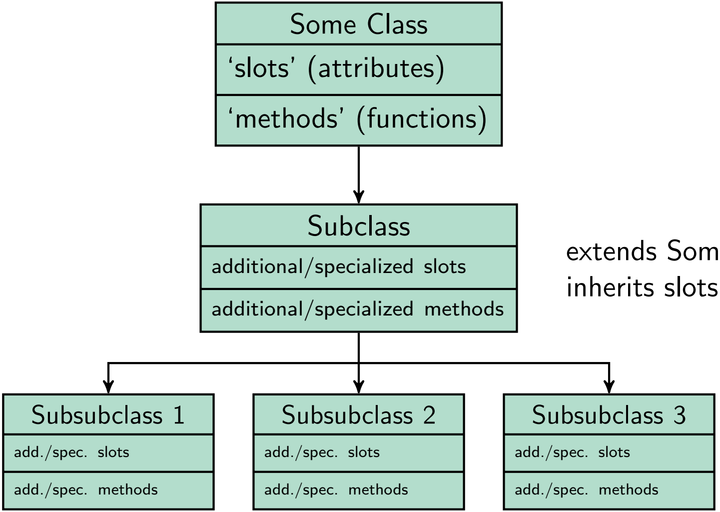
↪ blueprints for objects called 'classes'

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

example:
banking company administrating their customers' accounts

class:	`BankAccount`
instances:	bank account for customer A
	bank account for customer B
	⋮

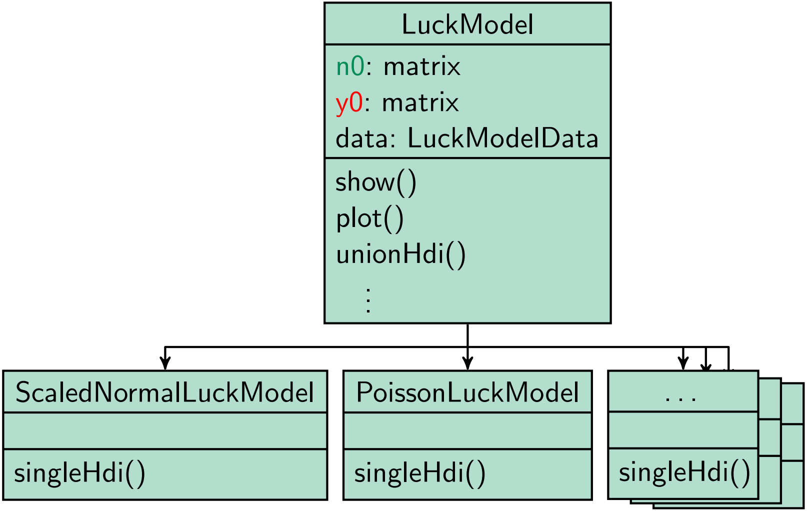
Object-oriented Programming: Class hierarchies

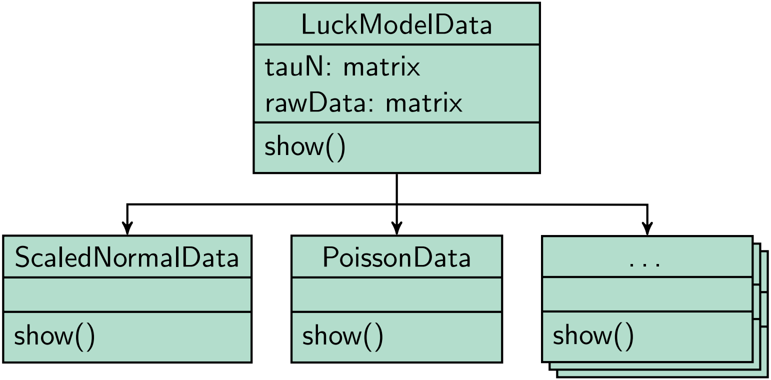


extends Some Class
inherits slots & methods

Implementation – Class Structure

Implemented class structure maps the hierarchy of the model:




LuckModelData
tauN: matrix
rawData: matrix
show()
ScaledNormalData
show()
PoissonData
show()
. . .
show()

Code Example

```
> ex1 <- LuckModel(n0=c(1,10), y0= c(0,5))
> ex1
generalized iLUCK model with prior parameter set:
  lower n0 = 1 upper n0 = 10
  lower y0 = 0 upper y0 = 5
 giving a main parameter prior imprecision of 5
> data1 <- LuckModelData(tau=11, n=2)
> data1
data object with sample statistic tau(x) = 11 and sample size n = 2
> ex2 <- ScaledNormalLuckModel(n0=c(1,2), 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 = 2
  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 [ 0.5 ; 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, 2008. 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.

G. Walter, T. Augustin, and A. Peters. Linear regression analysis under sets of conjugate priors. In G. de Cooman, J. Vejnarová, and M. Zaffalon, eds., *ISIPTA '07*, p. 445–455, 2007.