UPPSALA UNIVERSITET

Violent Crime

Bayesian Logistic Regression anything else than Lasso Regression?

Bodil Svennblad, PhD, UCR

2013-05-06

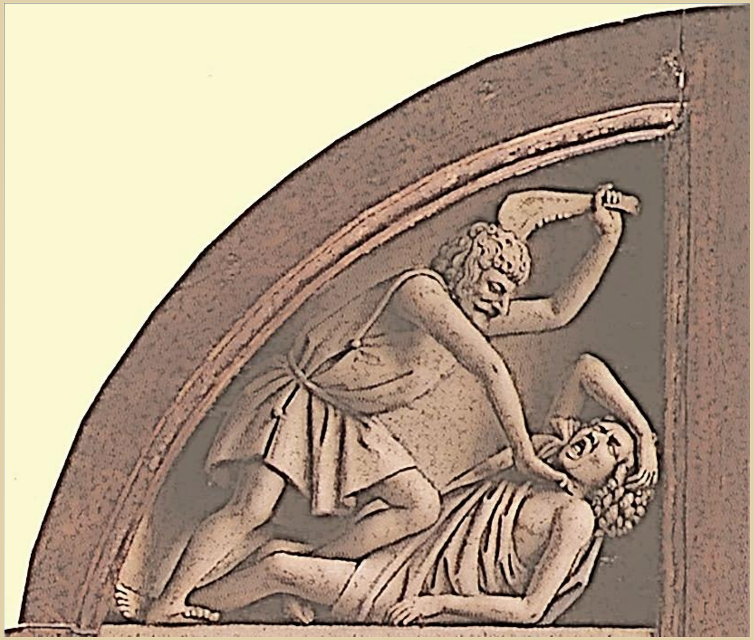
Rolf Gedeborg, MDr

Ingemar Thiblin, MDr

AKADEMISKA SJUKHUSET

UCR
UPPSALA CLINICAL RESEARCH CENTER

Violent Crime



- **Violence increases**?
- **Prediction model for mortality of victims**
  - As tool for legal process in court
  - In clinical practice

UPPSALA UNIVERSITET

Violent Crime, data

Swedish National Patient Registry ↔ Swedish Cause of Death Registry

- **Injury hospitalizations** from **1998 to 2004** with main ICD10 code S00-T80 (Injury, poisoning and certain other consequences of external causes) excluding adverse effects and poisoning
- Excluding readmissions
- ICD-10 cause of injury categorized using injury matrix (CDC)

~ 15.000 victims of homicide/ assault

AKADEMISKA SJUKHUSET

UCR UPPSALA CLINICAL RESEARCH CENTER

UPPSALA UNIVERSITET

Individual data

- Gender
- Age
- ICD10
- Date of admission
- DeathDate

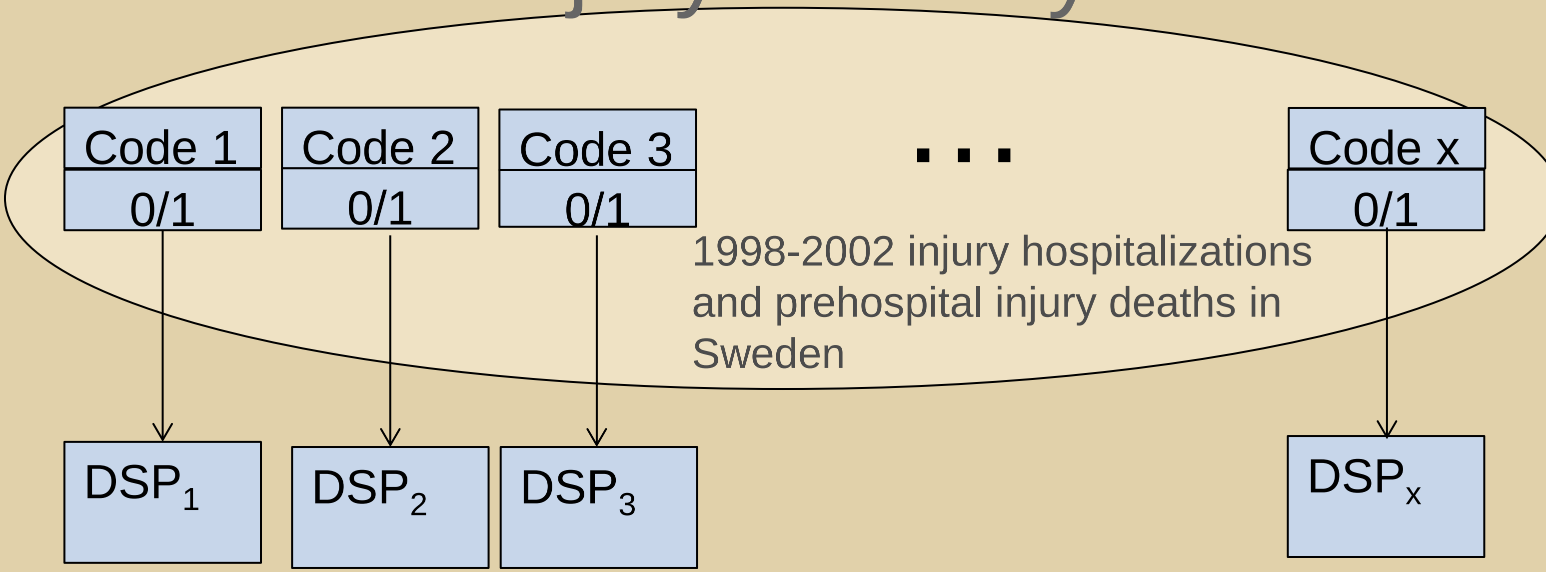
ICD10

- ICISS
- Cause
- Injury type
- Injury severity

AKADEMISKA SJUKHUSET

UCR UPPSALA CLINICAL RESEARCH CENTER

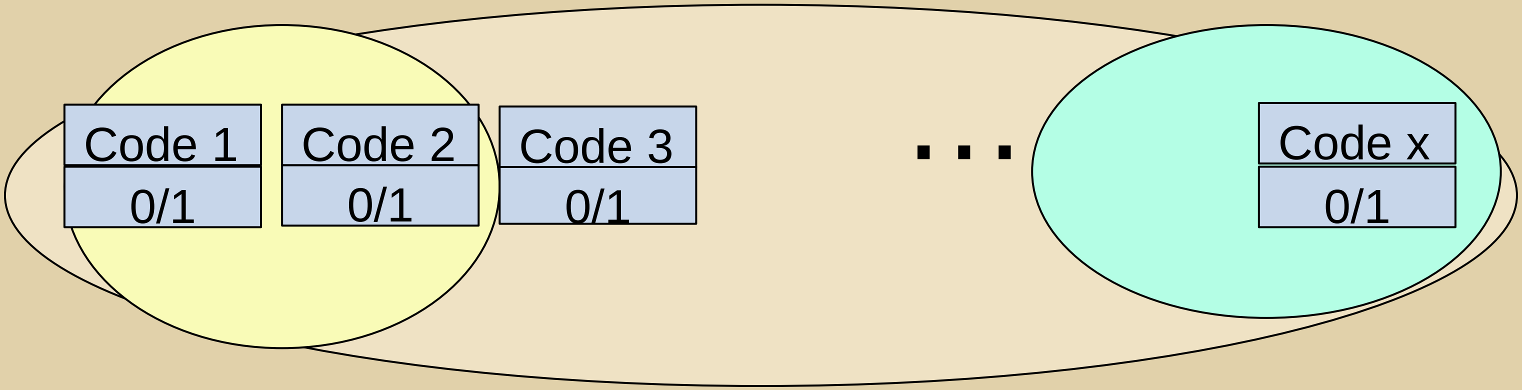
International Classification of Diseases Injury Severity Score



$$DSP_i = \frac{\#\ survival\ individuals\ with\ code_i}{\#\ individuals\ with\ code_i}$$

$ICISS_j$ = product of DSP_i for each code i present for indivdual j

ICD 10

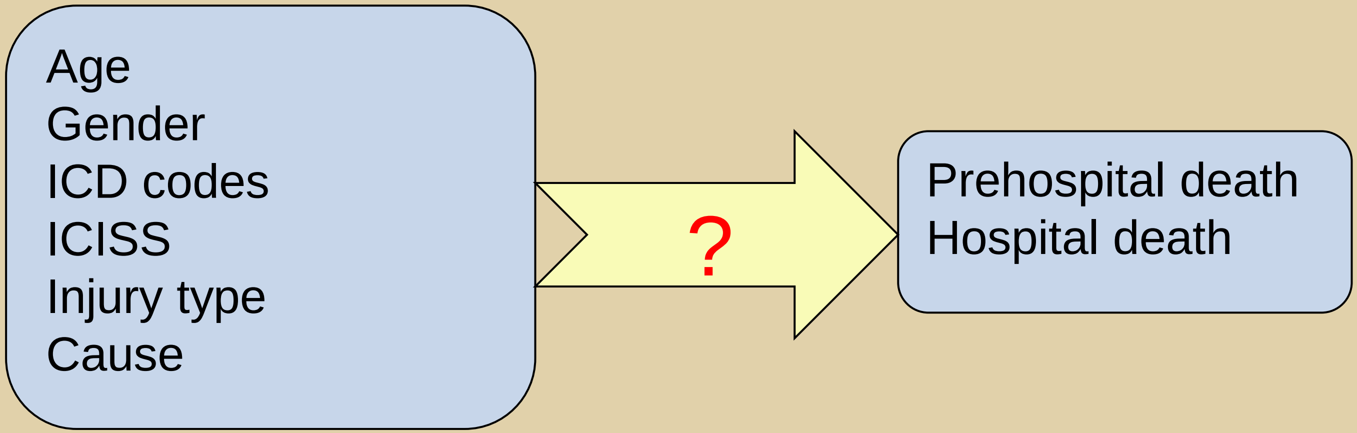


Cause
Cut/pierce
Fire/flame
Firearm
Struck by/against
Suffocation
other

Injury type
Head
Thoracic
Abdominal
Head/thorax
Head/abdomen

UPPSALA UNIVERSITET

Model



AKADEMISKA SJUKHUSET

UCR
UPPSALA CLINICAL RESEARCH CENTER

UPPSALA UNIVERSITET

Logistic regression

$$y_i \sim Be(p_i)$$

$$\log\left(\frac{p_i}{1-p_i}\right) = \mathrm{x}\beta$$

Estimate β by Maximum Likelihood

AKADEMISKA SJUKHUSET

UCR UPPSALA CLINICAL RESEARCH CENTER

Logistic regression

For prediction purposes, use some kind of shrinkage

- (Multiply each β with a factor c < 1)

- Estimate β_i with ML but with constraints

$$\sum_j \beta_j^2 \leq t$$

$$\sum_j \left|\beta_j\right| \leq t$$

UPPSALA UNIVERSITET

AKADEMISKA SJUKHUSET

UCR UPPSALA CLINICAL RESEARCH CENTER

LASSO – Least Absolute Shrinkage and Selection Operator

Tibshirani 1996

$$\beta = \arg\max \left\{ l\left(\beta \middle| data\right) - \lambda \sum \left|\beta_j\right| \right\}$$

UPPSALA UNIVERSITET

AKADEMISKA SJUKHUSET

UCR UPPSALA CLINICAL RESEARCH CENTER

Bayes – in general

$$apriori: \quad \pi(\beta)$$

$$likelihood: \quad L(\beta|data)$$

$$aposteriori: \quad f(\beta|data) \propto \pi(\beta)L(\beta|data)$$

$$\log f \propto \log \pi(\beta) + l(\beta|data)$$

Lasso as Bayes estimate

$$\beta = \arg\max \left\{ l\left(\beta \middle| data\right) - \lambda \sum \left|\beta_j\right| \right\}$$

?

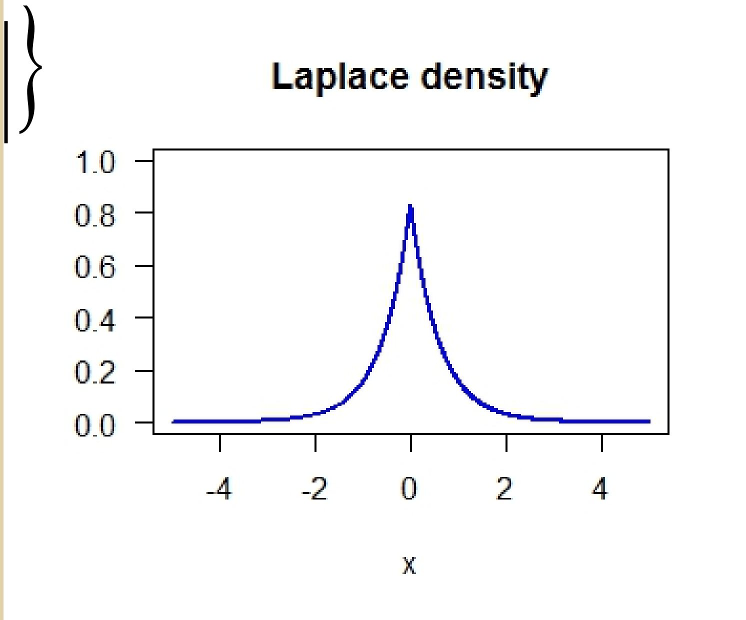
$$\log f \propto \log \pi\left(\beta\right) + l\left(\beta \middle| data\right)$$

Lasso as Bayes estimate

$$\beta = \arg\max\left\{ l\left(\beta \middle| data\right) - \lambda \sum \left|\beta_j\right| \right\}$$

$$\log f \propto \log \pi\left(\beta\right) + l\left(\beta \middle| data\right)$$

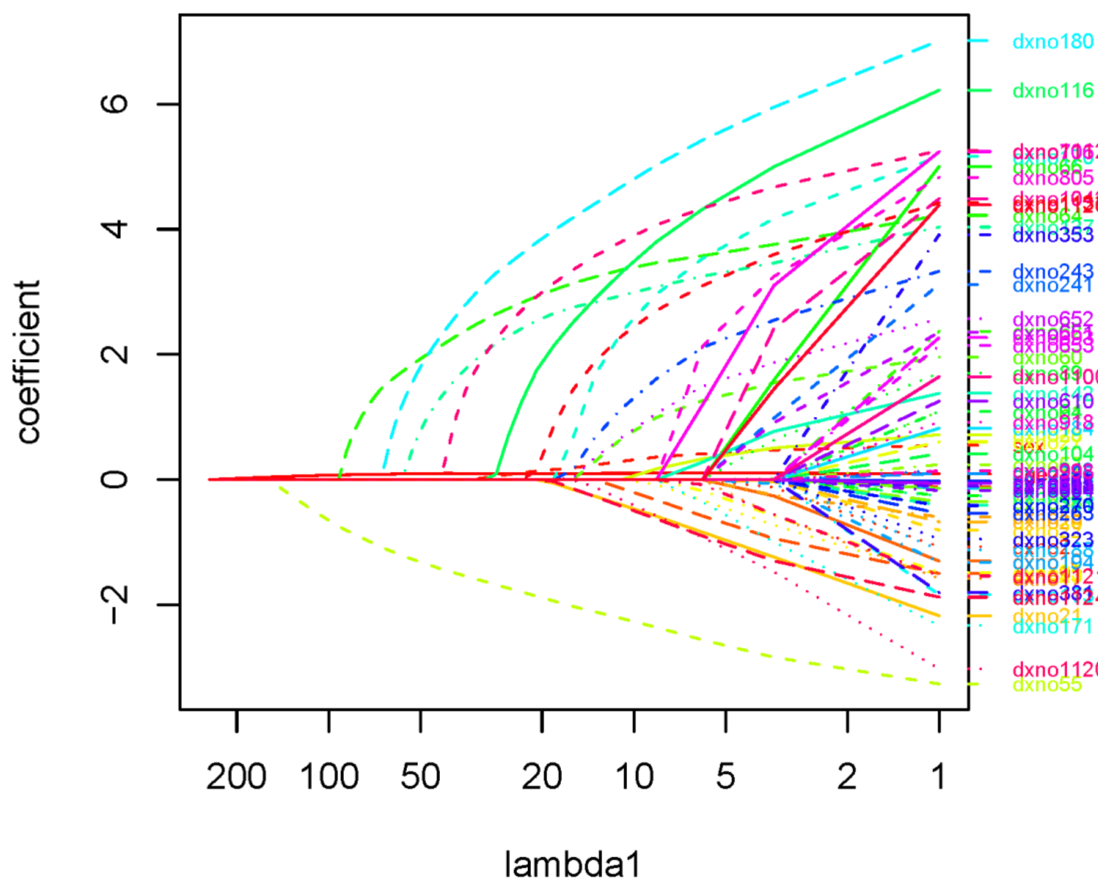
$$\pi\left(\beta\right) = \frac{\lambda}{2} \exp\left(-\lambda\left|\beta\right|\right)$$



Lasso estimate equals Maximum Posterior Mode from Bayesian logistic regression if Laplace distribution used as prior

UPPSALA UNIVERSITET

Lasso regression, model 3



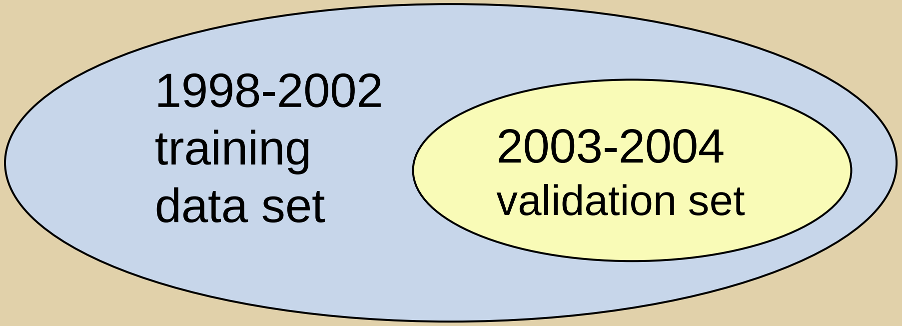
AKADEMISKA SJUKHUSET

UCR
UPPSALA CLINICAL RESEARCH CENTER

BBRtrain & BBRclassify

Genkin, Lewis , Madigan (Large-Scale Bayesian Logistic Regression for Text Categorization, 2007)

- Fast algorithm for finding posterior mode



UPPSALA UNIVERSITET

Models

Model	Variables	No of variables original	No of variables final
1	Gender, age	2	2
2	model 1 + iciss	3	3
3	model 1 + ICD	599	105 (128)
4	model 3 + 2way ICD interactions	5773	(177)
5	model 3 + iciss	600	(59)
6	model 2 + head + thorax + abdomen + head&thorax + head&abdomen	8	(8)
7	model 6 + cause	14	(14)
8	model 2 + cause	9	(9)

AKADEMISKA SJUKHUSET

UCR UPPSALA CLINICAL RESEARCH CENTER

Validation measures

Using point estimates (MAP)

- AUC
- HL
- Brier Score
- Scaled Brier Score
- Intercept
- Slope

What did we aim for?

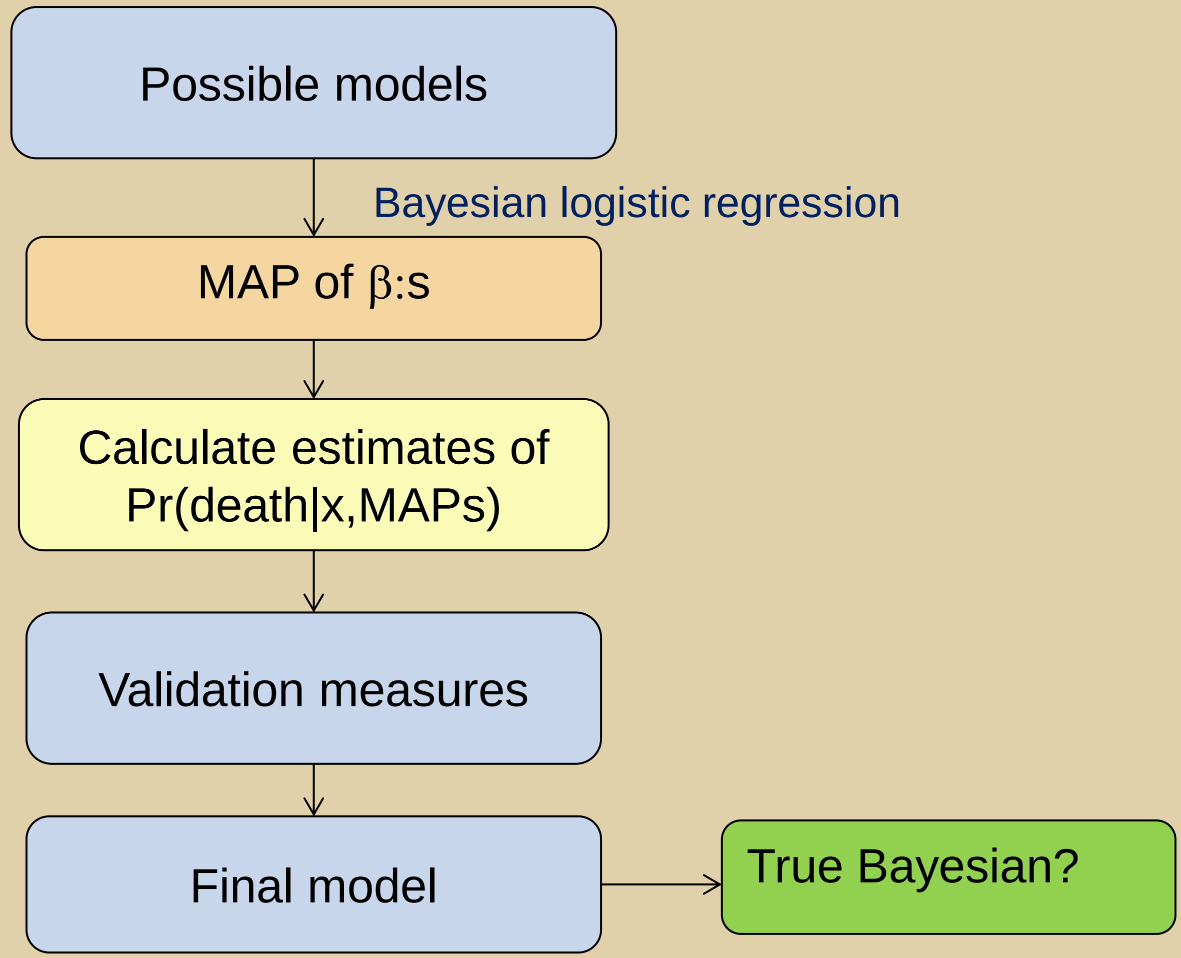
A prediction model with age, gender, iciss and perhaps other predictors that

- Is easy to use (parsimonious)
- Gives estimate of probability of death (not of β)
- Precision of the estimates

→ model 7: age, gender, iciss, cause, injury type (14 predictors)

UPPSALA UNIVERSITET

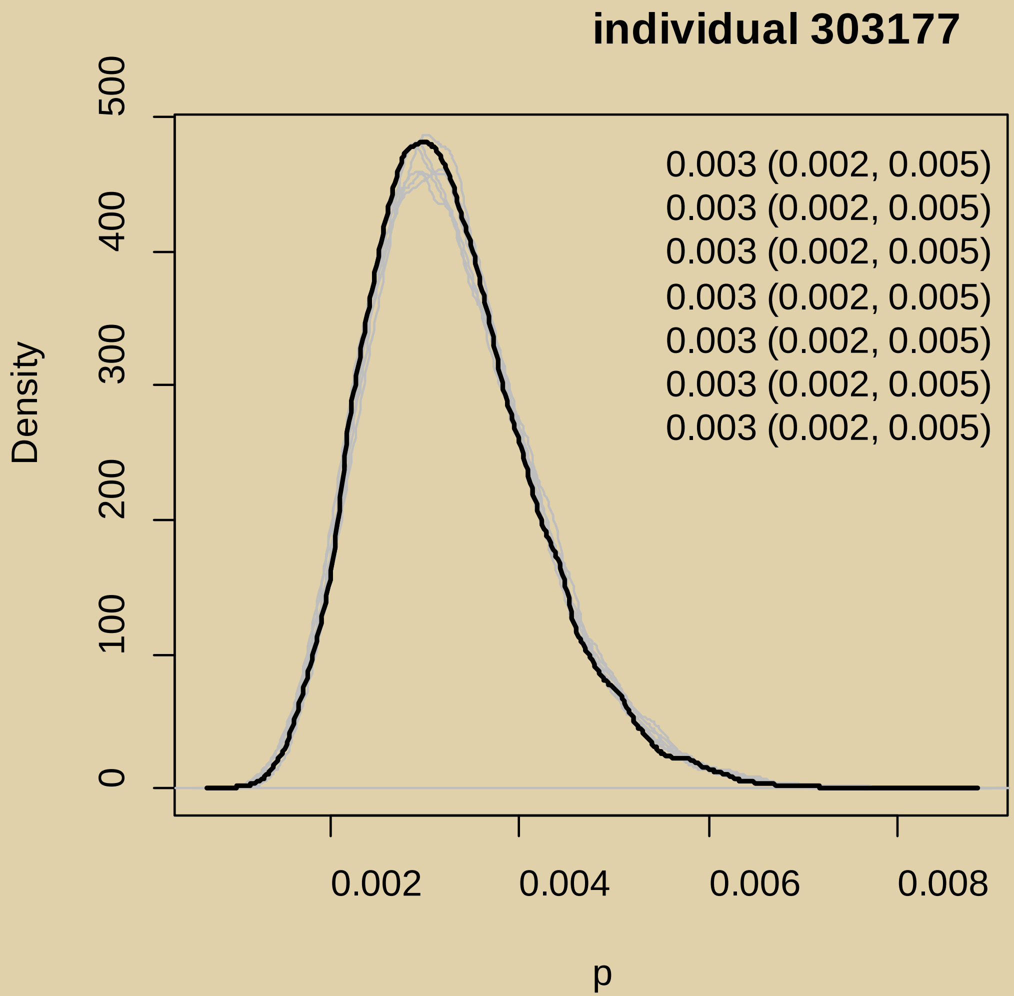
How did we do it?



AKADEMISKA SJUKHUSET

UCR
UPPSALA CLINICAL RESEARCH CENTER

True Bayesian, model 7



Credibility intervals

Bayesian logistic regression

- Necessary asumptions formalized in priors
- Possibility to incorporate prior knowledge
- Easy to implement using MCMC
- Estimates of parameters of interest
- Precision of estimates through credibility intervals which are easy to interpret