Bayesian generalized linear models and an appropriate default prior

Andrew Gelman, Aleks Jakulin, Maria Grazia Pittau, and Yu-Sung Su Columbia University

14 August 2008

イロト イヨト イヨト イヨト

Classical logistic regression The problem of separation Bayesian solution

Logistic regression



イロン 不同と 不同と 不同と

æ

Classical logistic regression The problem of separation Bayesian solution

A clean example



Gelman, Jakulin, Pittau, Su Bayesian generalized linear models and an appropriate default

イロン イヨン イヨン イヨン

æ

Classical logistic regression The problem of separation Bayesian solution

The problem of separation



slope = infinity?

Gelman, Jakulin, Pittau, Su Bayesian generalized linear models and an appropriate default

イロト イヨト イヨト イヨト

Classical logistic regression The problem of separation

Separation is no joke!

glm (vote ~ female + black + income, family=binomial(link="logit"))

1960

coef.est	coef.se
-0.14	0.23
0.24	0.14
-1.03	0.36
0.03	0.06
	coef.est -0.14 0.24 -1.03 0.03

1964

	coef.est	coef.se
(Intercept)	-1.15	0.22
female	-0.09	0.14
black	-16.83	420.40
income	0.19	0.06

1968

	coef.est	coef.se
(Intercept)	0.47	0.24
female	-0.01	0.15
black	-3.64	0.59
income	-0.03	0.07
1972		
	coef.est	coef.se
(Intercept)	0.67	0.18
female	-0.25	0.12
black	-2.63	0.27
income	0.09	0.05

・ロト ・回ト ・ヨト ・ヨト

æ

Gelman, Jakulin, Pittau, Su Bayesian generalized linear models and an appropriate default

Classical logistic regression The problem of separation Bayesian solution

bayesglm()

Bayesian logistic regression

- In the arm (Applied Regression and Multilevel modeling) package
- Replaces glm(), estimates are more numerically and computationally stable
- Student-t prior distributions for regression coefs
- Use EM-like algorithm
- We went inside glm.fit to augment the iteratively weighted least squares step
- Default choices for tuning parameters (we'll get back to this!)

・ロト ・回ト ・ヨト ・ヨト

3

Classical logistic regression The problem of separation Bayesian solution

bayesglm()

Bayesian logistic regression

- In the arm (Applied Regression and Multilevel modeling) package
- Replaces glm(), estimates are more numerically and computationally stable
- Student-t prior distributions for regression coefs
- Use EM-like algorithm
- We went inside glm.fit to augment the iteratively weighted least squares step
- Default choices for tuning parameters (we'll get back to this!)

・ロト ・回ト ・ヨト ・ヨト

3

Classical logistic regression The problem of separation Bayesian solution



- Bayesian logistic regression
- In the arm (Applied Regression and Multilevel modeling) package
- Replaces glm(), estimates are more numerically and computationally stable
- Student-t prior distributions for regression coefs
- Use EM-like algorithm
- We went inside glm.fit to augment the iteratively weighted least squares step
- Default choices for tuning parameters (we'll get back to this!)

(ロ) (同) (E) (E) (E)

Classical logistic regression The problem of separation Bayesian solution



- Bayesian logistic regression
- In the arm (Applied Regression and Multilevel modeling) package
- Replaces glm(), estimates are more numerically and computationally stable
- Student-t prior distributions for regression coefs
- Use EM-like algorithm
- We went inside glm.fit to augment the iteratively weighted least squares step
- Default choices for tuning parameters (we'll get back to this!)

・ロト ・回ト ・ヨト ・ヨト

3

Classical logistic regression The problem of separation Bayesian solution



- Bayesian logistic regression
- In the arm (Applied Regression and Multilevel modeling) package
- Replaces glm(), estimates are more numerically and computationally stable
- Student-t prior distributions for regression coefs
- Use EM-like algorithm
- We went inside glm.fit to augment the iteratively weighted least squares step
- Default choices for tuning parameters (we'll get back to this!)

Classical logistic regression The problem of separation Bayesian solution



- Bayesian logistic regression
- In the arm (Applied Regression and Multilevel modeling) package
- Replaces glm(), estimates are more numerically and computationally stable
- Student-t prior distributions for regression coefs
- Use EM-like algorithm
- We went inside glm.fit to augment the iteratively weighted least squares step
- Default choices for tuning parameters (we'll get back to this!)

Classical logistic regression The problem of separation Bayesian solution



- Bayesian logistic regression
- In the arm (Applied Regression and Multilevel modeling) package
- Replaces glm(), estimates are more numerically and computationally stable
- Student-t prior distributions for regression coefs
- Use EM-like algorithm
- We went inside glm.fit to augment the iteratively weighted least squares step
- Default choices for tuning parameters (we'll get back to this!)

Classical logistic regression The problem of separation Bayesian solution



- Bayesian logistic regression
- In the arm (Applied Regression and Multilevel modeling) package
- Replaces glm(), estimates are more numerically and computationally stable
- Student-t prior distributions for regression coefs
- Use EM-like algorithm
- We went inside glm.fit to augment the iteratively weighted least squares step
- Default choices for tuning parameters (we'll get back to this!)

イロン イヨン イヨン イヨン

Classical logistic regression The problem of separation Bayesian solution

Regularization in action!



Gelman, Jakulin, Pittau, Su

Bayesian generalized linear models and an appropriate default p

Classical logistic regression The problem of separation Bayesian solution

What else is out there?

- glm (maximum likelihood): fails under separation, gives noisy answers for sparse data
- Augment with prior "successes" and "failures": doesn't work well for multiple predictors
- brlr (Jeffreys-like prior distribution): computationally unstable
- brglm (improvement on brlr): doesn't do enough smoothing
- BBR (Laplace prior distribution): OK, not quite as good as bayesglm
- Non-Bayesian machine learning algorithms: understate uncertainty in predictions

Classical logistic regression The problem of separation Bayesian solution

What else is out there?

- glm (maximum likelihood): fails under separation, gives noisy answers for sparse data
- Augment with prior "successes" and "failures": doesn't work well for multiple predictors
- brlr (Jeffreys-like prior distribution): computationally unstable
- brglm (improvement on brlr): doesn't do enough smoothing
- BBR (Laplace prior distribution): OK, not quite as good as bayesglm
- Non-Bayesian machine learning algorithms: understate uncertainty in predictions

Classical logistic regression The problem of separation Bayesian solution

What else is out there?

- glm (maximum likelihood): fails under separation, gives noisy answers for sparse data
- Augment with prior "successes" and "failures": doesn't work well for multiple predictors
- brlr (Jeffreys-like prior distribution): computationally
 unstable
- brglm (improvement on brlr): doesn't do enough smoothing
- BBR (Laplace prior distribution): OK, not quite as good as bayesglm
- Non-Bayesian machine learning algorithms: understate uncertainty in predictions

Classical logistic regression The problem of separation Bayesian solution

What else is out there?

- glm (maximum likelihood): fails under separation, gives noisy answers for sparse data
- Augment with prior "successes" and "failures": doesn't work well for multiple predictors
- brlr (Jeffreys-like prior distribution): computationally unstable
- brglm (improvement on brlr): doesn't do enough smoothing
- BBR (Laplace prior distribution): OK, not quite as good as bayesglm
- Non-Bayesian machine learning algorithms: understate uncertainty in predictions

・ロン ・回 と ・ 回 と ・ 回 と

Classical logistic regression The problem of separation Bayesian solution

What else is out there?

- glm (maximum likelihood): fails under separation, gives noisy answers for sparse data
- Augment with prior "successes" and "failures": doesn't work well for multiple predictors
- brlr (Jeffreys-like prior distribution): computationally unstable
- brglm (improvement on brlr): doesn't do enough smoothing
- BBR (Laplace prior distribution): OK, not quite as good as bayesglm
- Non-Bayesian machine learning algorithms: understate uncertainty in predictions

(ロ) (同) (E) (E) (E)

Classical logistic regression The problem of separation Bayesian solution

What else is out there?

- glm (maximum likelihood): fails under separation, gives noisy answers for sparse data
- Augment with prior "successes" and "failures": doesn't work well for multiple predictors
- brlr (Jeffreys-like prior distribution): computationally unstable
- brglm (improvement on brlr): doesn't do enough smoothing
- BBR (Laplace prior distribution): OK, not quite as good as bayesglm
- Non-Bayesian machine learning algorithms: understate uncertainty in predictions

・ロン ・回 と ・ 回 と ・ 回 と

Classical logistic regression The problem of separation Bayesian solution

What else is out there?

- glm (maximum likelihood): fails under separation, gives noisy answers for sparse data
- Augment with prior "successes" and "failures": doesn't work well for multiple predictors
- brlr (Jeffreys-like prior distribution): computationally unstable
- brglm (improvement on brlr): doesn't do enough smoothing
- BBR (Laplace prior distribution): OK, not quite as good as bayesglm
- Non-Bayesian machine learning algorithms: understate uncertainty in predictions

(ロ) (同) (E) (E) (E)

Prior information

Who's the real conservative? Evaluation using a corpus of datasets Other generalized linear models

Information in prior distributions

Informative prior dist

- A full generative model for the data
- Noninformative prior dist

Weakly informative prior dist

・ロン ・回 とくほど ・ ほとう

Prior information

Who's the real conservative? Evaluation using a corpus of datasets Other generalized linear models

Information in prior distributions

Informative prior dist

- A full generative model for the data
- Noninformative prior dist
 - Let the data speak Coal: valid informatic for an
- ▶ Weakly informative prior dist

・ロン ・回 と ・ ヨン ・ ヨン

Prior information

Who's the real conservative? Evaluation using a corpus of datasets Other generalized linear models

Information in prior distributions

Informative prior dist

- A full generative model for the data
- Noninformative prior dist
 - Let the data speak.
 - Goal: valid inference for any heta
- Weakly informative prior dist

・ロン ・四マ ・ヨマ ・ヨマ

Prior information

Who's the real conservative? Evaluation using a corpus of datasets Other generalized linear models

Information in prior distributions

- Informative prior dist
 - A full generative model for the data
- Noninformative prior dist
 - Let the data speak
 - Goal: valid inference for any θ
- Weakly informative prior dist

イロン イヨン イヨン イヨン

Prior information

Who's the real conservative? Evaluation using a corpus of datasets Other generalized linear models

Information in prior distributions

- Informative prior dist
 - A full generative model for the data
- Noninformative prior dist
 - Let the data speak
 - Goal: valid inference for any θ
- Weakly informative prior dist
 - Purposely include less information than we actually have Goal: regularization, stabilization

イロト イヨト イヨト イヨト

Prior information

Who's the real conservative? Evaluation using a corpus of datasets Other generalized linear models

Information in prior distributions

- Informative prior dist
 - A full generative model for the data
- Noninformative prior dist
 - Let the data speak
 - Goal: valid inference for any θ
- Weakly informative prior dist
 - Purposely include less information than we actually have
 - Goal: regularization, stabilization

イロン 不同と 不同と 不同と

Prior information

Who's the real conservative? Evaluation using a corpus of datasets Other generalized linear models

Information in prior distributions

- Informative prior dist
 - A full generative model for the data
- Noninformative prior dist
 - Let the data speak
 - Goal: valid inference for any θ
- Weakly informative prior dist
 - Purposely include less information than we actually have
 - Goal: regularization, stabilization

イロン イヨン イヨン イヨン

Prior information

Who's the real conservative? Evaluation using a corpus of datasets Other generalized linear models

Information in prior distributions

- Informative prior dist
 - A full generative model for the data
- Noninformative prior dist
 - Let the data speak
 - Goal: valid inference for any θ
- Weakly informative prior dist
 - Purposely include less information than we actually have
 - Goal: regularization, stabilization

イロト イヨト イヨト イヨト

Prior information

Who's the real conservative? Evaluation using a corpus of datasets Other generalized linear models

Information in prior distributions

- Informative prior dist
 - A full generative model for the data
- Noninformative prior dist
 - Let the data speak
 - Goal: valid inference for any θ
- Weakly informative prior dist
 - Purposely include less information than we actually have
 - Goal: regularization, stabilization

イロン イヨン イヨン イヨン

Prior information Who's the real conservative? Evaluation using a corpus of datasets Other generalized linear models

Weakly informative priors for logistic regression coefficients

- Separation in logistic regression
- ▶ Some prior info: logistic regression coefs are almost always between -5 and 5:
 - 5 on the logit scale takes you from 0.01 to 0.50
 - or from 0.50 to 0.99
 - Smoking and lung cancer
- Independent Cauchy prior dists with center 0 and scale 2.5
- Rescale each predictor to have mean 0 and sd ¹/₂
- ▶ Fast implementation using EM; easy adaptation of glm

< ロ > < 回 > < 回 > < 回 > < 回 > <

Prior information Who's the real conservative? Evaluation using a corpus of datasets Other generalized linear models

Weakly informative priors for logistic regression coefficients

- Separation in logistic regression
- ▶ Some prior info: logistic regression coefs are almost always between -5 and 5:
 - 5 on the logit scale takes you from 0.01 to 0.50 or from 0.50 to 0.00
 - Smoking and lung cancer
- Independent Cauchy prior dists with center 0 and scale 2.5
- Rescale each predictor to have mean 0 and sd ¹/₂
- Fast implementation using EM; easy adaptation of glm

Prior information Who's the real conservative? Evaluation using a corpus of datasets Other generalized linear models

Weakly informative priors for logistic regression coefficients

- Separation in logistic regression
- ► Some prior info: logistic regression coefs are almost always between -5 and 5:
 - 5 on the logit scale takes you from 0.01 to 0.50 or from 0.50 to 0.99
 - Smoking and lung cancer
- Independent Cauchy prior dists with center 0 and scale 2.5
- Rescale each predictor to have mean 0 and sd ¹/₂
- Fast implementation using EM; easy adaptation of glm

・ロン ・四マ ・ヨマ ・ヨマ

Prior information Who's the real conservative? Evaluation using a corpus of datasets Other generalized linear models

Weakly informative priors for logistic regression coefficients

- Separation in logistic regression
- ► Some prior info: logistic regression coefs are almost always between -5 and 5:
 - 5 on the logit scale takes you from 0.01 to 0.50 or from 0.50 to 0.99
 - Smoking and lung cancer
- Independent Cauchy prior dists with center 0 and scale 2.5
- Rescale each predictor to have mean 0 and sd $\frac{1}{2}$
- Fast implementation using EM; easy adaptation of glm

Prior information Who's the real conservative? Evaluation using a corpus of datasets Other generalized linear models

Weakly informative priors for logistic regression coefficients

- Separation in logistic regression
- ► Some prior info: logistic regression coefs are almost always between -5 and 5:
 - 5 on the logit scale takes you from 0.01 to 0.50 or from 0.50 to 0.99
 - Smoking and lung cancer
- Independent Cauchy prior dists with center 0 and scale 2.5
- Rescale each predictor to have mean 0 and sd $\frac{1}{2}$
- Fast implementation using EM; easy adaptation of glm

Prior information Who's the real conservative? Evaluation using a corpus of datasets Other generalized linear models

Weakly informative priors for logistic regression coefficients

- Separation in logistic regression
- ► Some prior info: logistic regression coefs are almost always between -5 and 5:
 - 5 on the logit scale takes you from 0.01 to 0.50 or from 0.50 to 0.99
 - Smoking and lung cancer
- Independent Cauchy prior dists with center 0 and scale 2.5
- Rescale each predictor to have mean 0 and sd ¹/₂
- Fast implementation using EM; easy adaptation of glm
Prior information Who's the real conservative? Evaluation using a corpus of datasets Other generalized linear models

Weakly informative priors for logistic regression coefficients

- Separation in logistic regression
- ► Some prior info: logistic regression coefs are almost always between -5 and 5:
 - 5 on the logit scale takes you from 0.01 to 0.50 or from 0.50 to 0.99
 - Smoking and lung cancer
- Independent Cauchy prior dists with center 0 and scale 2.5
- Rescale each predictor to have mean 0 and sd $\frac{1}{2}$
- Fast implementation using EM; easy adaptation of glm

・ロト ・回ト ・ヨト ・ヨト

Prior information Who's the real conservative? Evaluation using a corpus of datasets Other generalized linear models

Weakly informative priors for logistic regression coefficients

- Separation in logistic regression
- ► Some prior info: logistic regression coefs are almost always between -5 and 5:
 - 5 on the logit scale takes you from 0.01 to 0.50 or from 0.50 to 0.99
 - Smoking and lung cancer
- Independent Cauchy prior dists with center 0 and scale 2.5
- Rescale each predictor to have mean 0 and sd $\frac{1}{2}$
- Fast implementation using EM; easy adaptation of glm

・ロン ・回と ・ヨン ・ヨン

Prior information

Who's the real conservative? Evaluation using a corpus of datasets Other generalized linear models

Prior distributions



< □ > < □ > < □ > < □ > < □ > .

æ

Prior information

Who's the real conservative? Evaluation using a corpus of datasets Other generalized linear models

Another example

Dose	#deaths/ $#$ animals
-0.86	0/5
-0.30	1/5
-0.05	3/5
0.73	5/5

- ▶ Slope of a logistic regression of Pr(death) on dose:
 - Maximum likelihood est is 7.8 ± 4.9
 - With weakly-informative prior: Bayes est is 4.4 ± 1.9
- Which is truly conservative?
- The sociology of shrinkage

・ロト ・回ト ・ヨト ・ヨト

3

Prior information Who's the real conserv

Evaluation using a corpus of datasets Other generalized linear models

Another example

Dose	#deaths/ $#$ animals
-0.86	0/5
-0.30	1/5
-0.05	3/5
0.73	5/5

Slope of a logistic regression of Pr(death) on dose:

- Maximum likelihood est is 7.8 ± 4.9
- With weakly-informative prior: Bayes est is 4.4 ± 1.9
- Which is truly conservative?
- The sociology of shrinkage

・ロン ・回 と ・ ヨン ・ ヨン

3

Prior information Who's the real conservative?

Evaluation using a corpus of datasets Other generalized linear models

Another example

Dose	#deaths/ $#$ animals
-0.86	0/5
-0.30	1/5
-0.05	3/5
0.73	5/5

- Slope of a logistic regression of Pr(death) on dose:
 - Maximum likelihood est is 7.8 ± 4.9
 - With weakly-informative prior: Bayes est is 4.4 ± 1.9
- Which is truly conservative?
- The sociology of shrinkage

・ロン ・四マ ・ヨマ ・ヨマ

Prior information Who's the real conservative?

Evaluation using a corpus of datasets Other generalized linear models

Another example

Dose	#deaths/ $#$ animals
-0.86	0/5
-0.30	1/5
-0.05	3/5
0.73	5/5

- Slope of a logistic regression of Pr(death) on dose:
 - Maximum likelihood est is 7.8 ± 4.9
 - \blacktriangleright With weakly-informative prior: Bayes est is 4.4 ± 1.9
- Which is truly conservative?
- The sociology of shrinkage

・ロット (四) (日) (日)

Prior information Who's the real conservative?

Evaluation using a corpus of datasets Other generalized linear models

Another example

Dose	#deaths/ $#$ animals
-0.86	0/5
-0.30	1/5
-0.05	3/5
0.73	5/5

- Slope of a logistic regression of Pr(death) on dose:
 - Maximum likelihood est is 7.8 ± 4.9
 - \blacktriangleright With weakly-informative prior: Bayes est is 4.4 ± 1.9
- Which is truly conservative?
- The sociology of shrinkage

・ロット (四) (日) (日)

Prior information Who's the real conservative?

Evaluation using a corpus of datasets Other generalized linear models

Another example

Dose	#deaths/ $#$ animals
-0.86	0/5
-0.30	1/5
-0.05	3/5
0.73	5/5

- Slope of a logistic regression of Pr(death) on dose:
 - Maximum likelihood est is 7.8 ± 4.9
 - \blacktriangleright With weakly-informative prior: Bayes est is 4.4 ± 1.9
- Which is truly conservative?
- The sociology of shrinkage

(日) (同) (E) (E) (E)

Prior information Who's the real conservative? Evaluation using a corpus of datasets Other generalized linear models

Maximum likelihood and Bayesian estimates



æ

Prior information Who's the real conservative? Evaluation using a corpus of datasets Other generalized linear models

Conservatism of Bayesian inference

- Problems with maximum likelihood when data show separation:
 - ► Coefficient estimate of −∞
 - Estimated predictive probability of 0 for new cases
- Is this conservative?
- Not if evaluated by log score or predictive log-likelihood

・ロン ・回 と ・ ヨ と ・ ヨ と

Prior information Who's the real conservative? Evaluation using a corpus of datasets Other generalized linear models

Conservatism of Bayesian inference

Problems with maximum likelihood when data show separation:

- ► Coefficient estimate of -∞
- Estimated predictive probability of 0 for new cases
- Is this conservative?
- Not if evaluated by log score or predictive log-likelihood

・ロン ・回 と ・ ヨ と ・ ヨ と

Prior information Who's the real conservative? Evaluation using a corpus of datasets Other generalized linear models

Conservatism of Bayesian inference

- Problems with maximum likelihood when data show separation:
 - Coefficient estimate of $-\infty$
 - Estimated predictive probability of 0 for new cases
- Is this conservative?
- Not if evaluated by log score or predictive log-likelihood.

イロン 不同と 不同と 不同と

Prior information Who's the real conservative? Evaluation using a corpus of datasets Other generalized linear models

Conservatism of Bayesian inference

- Problems with maximum likelihood when data show separation:
 - Coefficient estimate of $-\infty$
 - Estimated predictive probability of 0 for new cases
- Is this conservative?
- Not if evaluated by log score or predictive log-likelihood

イロン 不同と 不同と 不同と

Prior information Who's the real conservative? Evaluation using a corpus of datasets Other generalized linear models

Conservatism of Bayesian inference

- Problems with maximum likelihood when data show separation:
 - Coefficient estimate of $-\infty$
 - Estimated predictive probability of 0 for new cases
- Is this conservative?
- ▶ Not if evaluated by log score or predictive log-likelihood

<ロ> (日) (日) (日) (日) (日)

Prior information Who's the real conservative? Evaluation using a corpus of datasets Other generalized linear models

Conservatism of Bayesian inference

- Problems with maximum likelihood when data show separation:
 - Coefficient estimate of $-\infty$
 - Estimated predictive probability of 0 for new cases
- Is this conservative?
- Not if evaluated by log score or predictive log-likelihood

イロト イヨト イヨト イヨト

Prior information Who's the real conservative? Evaluation using a corpus of datasets Other generalized linear models

Which one is conservative?



Gelman, Jakulin, Pittau, Su Bayesian generalized linear models and an appropriate default

æ

Prior information Who's the real conservative? Evaluation using a corpus of datasets Other generalized linear models

Prior as population distribution

- Consider many possible datasets
- The "true prior" is the distribution of β 's across these datasets
- Fit one dataset at a time
- A "weakly informative prior" has less information (wider variance) than the true prior
- Open question: How to formalize the tradeoffs from using different priors?

・ロン ・回と ・ヨン ・ヨン

Prior information Who's the real conservative? Evaluation using a corpus of datasets Other generalized linear models

Prior as population distribution

Consider many possible datasets

- The "true prior" is the distribution of β 's across these datasets
- Fit one dataset at a time
- A "weakly informative prior" has less information (wider variance) than the true prior
- Open question: How to formalize the tradeoffs from using different priors?

イロン 不同と 不同と 不同と

Prior information Who's the real conservative? Evaluation using a corpus of datasets Other generalized linear models

Prior as population distribution

- Consider many possible datasets
- The "true prior" is the distribution of β 's across these datasets
- Fit one dataset at a time
- A "weakly informative prior" has less information (wider variance) than the true prior
- Open question: How to formalize the tradeoffs from using different priors?

Prior information Who's the real conservative? Evaluation using a corpus of datasets Other generalized linear models

Prior as population distribution

- Consider many possible datasets
- The "true prior" is the distribution of β 's across these datasets
- Fit one dataset at a time
- A "weakly informative prior" has less information (wider variance) than the true prior
- Open question: How to formalize the tradeoffs from using different priors?

・ロン ・回と ・ヨン ・ヨン

Prior information Who's the real conservative? Evaluation using a corpus of datasets Other generalized linear models

Prior as population distribution

- Consider many possible datasets
- The "true prior" is the distribution of β 's across these datasets
- Fit one dataset at a time
- A "weakly informative prior" has less information (wider variance) than the true prior
- Open question: How to formalize the tradeoffs from using different priors?

・ロン ・回と ・ヨン ・ヨン

Prior information Who's the real conservative? Evaluation using a corpus of datasets Other generalized linear models

Prior as population distribution

- Consider many possible datasets
- The "true prior" is the distribution of β 's across these datasets
- Fit one dataset at a time
- A "weakly informative prior" has less information (wider variance) than the true prior
- Open question: How to formalize the tradeoffs from using different priors?

イロト イヨト イヨト イヨト

Prior information Who's the real conservative? Evaluation using a corpus of datasets Other generalized linear models

Evaluation using a corpus of datasets

- Compare classical glm to Bayesian estimates using various prior distributions
- Evaluate using 5-fold cross-validation and average predictive error
- The optimal prior distribution for β 's is (approx) Cauchy (0, 1)
- Our Cauchy (0, 2.5) prior distribution is weakly informative!

イロン イヨン イヨン イヨン

Prior information Who's the real conservative? Evaluation using a corpus of datasets Other generalized linear models

Evaluation using a corpus of datasets

- Compare classical glm to Bayesian estimates using various prior distributions
- Evaluate using 5-fold cross-validation and average predictive error
- The optimal prior distribution for β 's is (approx) Cauchy (0, 1)
- Our Cauchy (0, 2.5) prior distribution is weakly informative!

イロン イヨン イヨン イヨン

Prior information Who's the real conservative? Evaluation using a corpus of datasets Other generalized linear models

Evaluation using a corpus of datasets

- Compare classical glm to Bayesian estimates using various prior distributions
- Evaluate using 5-fold cross-validation and average predictive error
- The optimal prior distribution for β 's is (approx) Cauchy (0, 1)
- Our Cauchy (0,2.5) prior distribution is weakly informative!

・ロン ・回 と ・ ヨ と ・ ヨ と

Prior information Who's the real conservative? Evaluation using a corpus of datasets Other generalized linear models

Evaluation using a corpus of datasets

- Compare classical glm to Bayesian estimates using various prior distributions
- Evaluate using 5-fold cross-validation and average predictive error
- The optimal prior distribution for β 's is (approx) Cauchy(0,1)
- Our Cauchy (0,2.5) prior distribution is weakly informative!

・ロット (四) (日) (日)

Prior information Who's the real conservative? Evaluation using a corpus of datasets Other generalized linear models

Evaluation using a corpus of datasets

- Compare classical glm to Bayesian estimates using various prior distributions
- Evaluate using 5-fold cross-validation and average predictive error
- The optimal prior distribution for β 's is (approx) Cauchy (0, 1)
- Our Cauchy (0, 2.5) prior distribution is weakly informative!

・ロン ・回と ・ヨン ・ヨン

Prior information Who's the real conservative? Evaluation using a corpus of datasets Other generalized linear models

Expected predictive loss, avg over a corpus of datasets



< E

Prior information Who's the real conservative? Evaluation using a corpus of datasets Other generalized linear models

Priors for other regression models

- Probit
- Ordered logit/probit
- Poisson
- Linear regression with normal errors

・ロト ・回ト ・ヨト ・ヨト

3

Prior information Who's the real conservative? Evaluation using a corpus of datasets Other generalized linear models

Priors for other regression models

Probit

- Ordered logit/probit
- Poisson
- Linear regression with normal errors

・ロン ・回 と ・ ヨ と ・ ヨ と

3

Prior information Who's the real conservative? Evaluation using a corpus of datasets Other generalized linear models

Priors for other regression models

- Probit
- Ordered logit/probit
- Poisson
- Linear regression with normal errors

Prior information Who's the real conservative? Evaluation using a corpus of datasets Other generalized linear models

Priors for other regression models

- Probit
- Ordered logit/probit
- Poisson
- Linear regression with normal errors

Prior information Who's the real conservative? Evaluation using a corpus of datasets Other generalized linear models

Priors for other regression models

- Probit
- Ordered logit/probit
- Poisson
- Linear regression with normal errors

Prior information Who's the real conservative? Evaluation using a corpus of datasets Other generalized linear models

Other examples of weakly informative priors

- Variance parameters
- Covariance matrices
- Population variation in a physiological model
- Mixture models
- Intentional underpooling in hierarchical models

イロン イヨン イヨン イヨン

Prior information Who's the real conservative? Evaluation using a corpus of datasets Other generalized linear models

Other examples of weakly informative priors

Variance parameters

- Covariance matrices
- Population variation in a physiological model
- Mixture models
- Intentional underpooling in hierarchical models

イロン イヨン イヨン イヨン
Prior information Who's the real conservative? Evaluation using a corpus of datasets Other generalized linear models

Other examples of weakly informative priors

- Variance parameters
- Covariance matrices
- Population variation in a physiological model
- Mixture models
- Intentional underpooling in hierarchical models

Prior information Who's the real conservative? Evaluation using a corpus of datasets Other generalized linear models

Other examples of weakly informative priors

- Variance parameters
- Covariance matrices
- Population variation in a physiological model
- Mixture models
- Intentional underpooling in hierarchical models

Prior information Who's the real conservative? Evaluation using a corpus of datasets Other generalized linear models

Other examples of weakly informative priors

- Variance parameters
- Covariance matrices
- Population variation in a physiological model
- Mixture models
- Intentional underpooling in hierarchical models

Prior information Who's the real conservative? Evaluation using a corpus of datasets Other generalized linear models

Other examples of weakly informative priors

- Variance parameters
- Covariance matrices
- Population variation in a physiological model
- Mixture models
- Intentional underpooling in hierarchical models

Conclusions Extra stuff

Conclusions

- "Noninformative priors" are actually weakly informative
- "Weakly informative" is a more general and useful concept
 Regularization

Why use weakly informative priors rather than informative priors?

・ロト ・回ト ・ヨト ・ヨト

3

Conclusions Extra stuff

Conclusions

- "Noninformative priors" are actually weakly informative
- "Weakly informative" is a more general and useful concept
- Regularization
 - Better inferences
 - Stability of computation (bayesglm)
- Why use weakly informative priors rather than informative priors?

・ロン ・回 と ・ ヨ と ・ ヨ と

3

Conclusions Extra stuff

Conclusions

- "Noninformative priors" are actually weakly informative
- "Weakly informative" is a more general and useful concept
- Regularization
 - Better inferences
 - Stability of computation (bayesglm)
- Why use weakly informative priors rather than informative priors?

(ロ) (同) (E) (E) (E)

Conclusions Extra stuff

Conclusions

- "Noninformative priors" are actually weakly informative
- "Weakly informative" is a more general and useful concept
- Regularization
 - Better inferences
 - Stability of computation (bayesglm)

Why use weakly informative priors rather than informative priors?

・ロト ・回ト ・ヨト ・ヨト

Conclusions Extra stuff

Conclusions

- "Noninformative priors" are actually weakly informative
- "Weakly informative" is a more general and useful concept
- Regularization
 - Better inferences
 - Stability of computation (bayesglm)
- Why use weakly informative priors rather than informative priors?
 - Conformity with statistical culture ("conservatism") Labor-saving device
 - Robustness

Conclusions Extra stuff

Conclusions

- "Noninformative priors" are actually weakly informative
- "Weakly informative" is a more general and useful concept
- Regularization
 - Better inferences
 - Stability of computation (bayesglm)
- Why use weakly informative priors rather than informative priors?
 - Conformity with statistical culture ("conservatism")
 - Labor-saving device
 - Robustness

(ロ) (同) (E) (E) (E)

Conclusions Extra stuff

Conclusions

- "Noninformative priors" are actually weakly informative
- "Weakly informative" is a more general and useful concept
- Regularization
 - Better inferences
 - Stability of computation (bayesglm)
- Why use weakly informative priors rather than informative priors?
 - Conformity with statistical culture ("conservatism")
 - Labor-saving device
 - Robustness

・ロト ・回ト ・ヨト ・ヨト

Conclusions Extra stuff

Conclusions

- "Noninformative priors" are actually weakly informative
- "Weakly informative" is a more general and useful concept
- Regularization
 - Better inferences
 - Stability of computation (bayesglm)
- Why use weakly informative priors rather than informative priors?
 - Conformity with statistical culture ("conservatism")
 - Labor-saving device
 - Robustness

Conclusions Extra stuff

Conclusions

- "Noninformative priors" are actually weakly informative
- "Weakly informative" is a more general and useful concept
- Regularization
 - Better inferences
 - Stability of computation (bayesglm)
- Why use weakly informative priors rather than informative priors?
 - Conformity with statistical culture ("conservatism")
 - Labor-saving device
 - Robustness

Conclusions Extra stuff

Conclusions

- "Noninformative priors" are actually weakly informative
- "Weakly informative" is a more general and useful concept
- Regularization
 - Better inferences
 - Stability of computation (bayesglm)
- Why use weakly informative priors rather than informative priors?
 - Conformity with statistical culture ("conservatism")
 - Labor-saving device
 - Robustness

Conclusions Extra stuff

Other examples of weakly informative priors

- Variance parameters
- Covariance matrices
- Population variation in a physiological model
- Mixture models
- Intentional underpooling in hierarchical models

Conclusions Extra stuff

Other examples of weakly informative priors

Variance parameters

- Covariance matrices
- Population variation in a physiological model
- Mixture models
- Intentional underpooling in hierarchical models

Conclusions Extra stuff

Other examples of weakly informative priors

- Variance parameters
- Covariance matrices
- Population variation in a physiological model
- Mixture models
- Intentional underpooling in hierarchical models

Conclusions Extra stuff

Other examples of weakly informative priors

- Variance parameters
- Covariance matrices
- Population variation in a physiological model
- Mixture models
- Intentional underpooling in hierarchical models

Conclusions Extra stuff

Other examples of weakly informative priors

- Variance parameters
- Covariance matrices
- Population variation in a physiological model
- Mixture models
- Intentional underpooling in hierarchical models

Conclusions Extra stuff

Other examples of weakly informative priors

- Variance parameters
- Covariance matrices
- Population variation in a physiological model
- Mixture models
- Intentional underpooling in hierarchical models

- 4 同 6 4 日 6 4 日 6

Conclusions Extra stuff

Weakly informative priors for variance parameter

- Basic hierarchical model
- Traditional inverse-gamma(0.001, 0.001) prior can be highly informative (in a bad way)!
- Noninformative uniform prior works better
- ▶ But if #groups is small (J = 2, 3, even 5), a weakly informative prior helps by shutting down huge values of τ

Conclusions Extra stuff

Weakly informative priors for variance parameter

Basic hierarchical model

- Traditional inverse-gamma(0.001, 0.001) prior can be highly informative (in a bad way)!
- Noninformative uniform prior works better
- ▶ But if #groups is small (J = 2, 3, even 5), a weakly informative prior helps by shutting down huge values of τ

Conclusions Extra stuff

Weakly informative priors for variance parameter

- Basic hierarchical model
- Traditional inverse-gamma(0.001, 0.001) prior can be highly informative (in a bad way)!
- Noninformative uniform prior works better
- ▶ But if #groups is small (J = 2, 3, even 5), a weakly informative prior helps by shutting down huge values of τ

・ロト ・回ト ・ヨト ・ヨト

Conclusions Extra stuff

Weakly informative priors for variance parameter

- Basic hierarchical model
- Traditional inverse-gamma(0.001, 0.001) prior can be highly informative (in a bad way)!
- Noninformative uniform prior works better
- ▶ But if #groups is small (J = 2, 3, even 5), a weakly informative prior helps by shutting down huge values of τ

・ロン ・回 と ・ ヨ と ・ ヨ と

Conclusions Extra stuff

Weakly informative priors for variance parameter

- Basic hierarchical model
- Traditional inverse-gamma(0.001, 0.001) prior can be highly informative (in a bad way)!
- Noninformative uniform prior works better
- ▶ But if #groups is small (J = 2, 3, even 5), a weakly informative prior helps by shutting down huge values of τ

Conclusions Extra stuff

Priors for variance parameter: J = 8 groups



Conclusions Extra stuff

Priors for variance parameter: J = 3 groups



(4月) (4日) (4日)

Conclusions Extra stuff

Weakly informative priors for covariance matrices

- Inverse-Wishart has problems
- Correlations can be between 0 and 1
- Set up models so prior expectation of correlations is 0
- Goal: to be weakly informative about correlations and variances
- Scaled inverse-Wishart model uses redundant parameterization

Conclusions Extra stuff

Weakly informative priors for covariance matrices

Inverse-Wishart has problems

- Correlations can be between 0 and 1
- Set up models so prior expectation of correlations is 0
- Goal: to be weakly informative about correlations and variances
- Scaled inverse-Wishart model uses redundant parameterization

<ロ> (日) (日) (日) (日) (日)

Conclusions Extra stuff

Weakly informative priors for covariance matrices

- Inverse-Wishart has problems
- Correlations can be between 0 and 1
- Set up models so prior expectation of correlations is 0
- Goal: to be weakly informative about correlations and variances
- Scaled inverse-Wishart model uses redundant parameterization

Conclusions Extra stuff

Weakly informative priors for covariance matrices

- Inverse-Wishart has problems
- Correlations can be between 0 and 1
- Set up models so prior expectation of correlations is 0
- Goal: to be weakly informative about correlations and variances
- Scaled inverse-Wishart model uses redundant parameterization

Conclusions Extra stuff

Weakly informative priors for covariance matrices

- Inverse-Wishart has problems
- Correlations can be between 0 and 1
- Set up models so prior expectation of correlations is 0
- Goal: to be weakly informative about correlations and variances
- Scaled inverse-Wishart model uses redundant parameterization

Conclusions Extra stuff

Weakly informative priors for covariance matrices

- Inverse-Wishart has problems
- Correlations can be between 0 and 1
- Set up models so prior expectation of correlations is 0
- Goal: to be weakly informative about correlations and variances
- Scaled inverse-Wishart model uses redundant parameterization

Conclusions Extra stuff

Weakly informative priors for population variation in a physiological model

- Pharamcokinetic parameters such as the "Michaelis-Menten coefficient"
- Wide uncertainty: prior guess for θ is 15 with a factor of 100 of uncertainty, log θ ∼ N(log(15), log(10)²)
- Population model: data on several people j, log θ_j ~ N(log(15), log(10)²) ????
- Hierarchical prior distribution:

Weakly informative

・ロン ・回と ・ヨン・

Conclusions Extra stuff

Weakly informative priors for population variation in a physiological model

- Pharamcokinetic parameters such as the "Michaelis-Menten coefficient"
- Wide uncertainty: prior guess for θ is 15 with a factor of 100 of uncertainty, log θ ∼ N(log(15), log(10)²)
- Population model: data on several people j, log θ_j ~ N(log(15), log(10)²) ????
- Hierarchical prior distribution:

Weakly informative

・ロン ・回と ・ヨン ・ヨン

Conclusions Extra stuff

Weakly informative priors for population variation in a physiological model

- Pharamcokinetic parameters such as the "Michaelis-Menten coefficient"
- ► Wide uncertainty: prior guess for θ is 15 with a factor of 100 of uncertainty, log θ ~ N(log(15), log(10)²)
- ▶ Population model: data on several people j, log θ_j ~ N(log(15), log(10)²) ????
- Hierarchical prior distribution:

Weakly informative

・ロン ・回 と ・ ヨ と ・ ヨ と
Conclusions Extra stuff

Weakly informative priors for population variation in a physiological model

- Pharamcokinetic parameters such as the "Michaelis-Menten coefficient"
- Wide uncertainty: prior guess for θ is 15 with a factor of 100 of uncertainty, log θ ∼ N(log(15), log(10)²)
- ▶ Population model: data on several people j, log θ_j ~ N(log(15), log(10)²) ????
- Hierarchical prior distribution:
 - ▶ $\mu \sim N(\log(15), \log(10)^2)$
- Weakly informative

Conclusions Extra stuff

Weakly informative priors for population variation in a physiological model

- Pharamcokinetic parameters such as the "Michaelis-Menten coefficient"
- Wide uncertainty: prior guess for θ is 15 with a factor of 100 of uncertainty, log θ ∼ N(log(15), log(10)²)
- Population model: data on several people j, log θ_j ∼ N(log(15), log(10)²) ????
- Hierarchical prior distribution:
 - $\log \theta_j \sim \mathsf{N}(\mu, \sigma^2), \ \sigma \approx \log(2)$
 - $\mu \sim \mathsf{N}(\mathsf{log}(15),\mathsf{log}(10)^2)$
- Weakly informative

Conclusions Extra stuff

Weakly informative priors for population variation in a physiological model

- Pharamcokinetic parameters such as the "Michaelis-Menten coefficient"
- ► Wide uncertainty: prior guess for θ is 15 with a factor of 100 of uncertainty, log θ ~ N(log(15), log(10)²)
- Population model: data on several people j, log θ_j ∼ N(log(15), log(10)²) ????
- Hierarchical prior distribution:
 - $\log \theta_j \sim N(\mu, \sigma^2), \ \sigma \approx \log(2)$
 - $\mu \sim N(\log(15), \log(10)^2)$
- Weakly informative

・ 同 ト ・ ヨ ト ・ ヨ ト

Conclusions Extra stuff

Weakly informative priors for population variation in a physiological model

- Pharamcokinetic parameters such as the "Michaelis-Menten coefficient"
- ► Wide uncertainty: prior guess for θ is 15 with a factor of 100 of uncertainty, log θ ~ N(log(15), log(10)²)
- ▶ Population model: data on several people j, log θ_j ~ N(log(15), log(10)²) ????
- Hierarchical prior distribution:
 - $\log \theta_j \sim N(\mu, \sigma^2), \ \sigma \approx \log(2)$
 - $\mu \sim N(\log(15), \log(10)^2)$
- Weakly informative

・ 同 ト ・ ヨ ト ・ ヨ ト

Conclusions Extra stuff

Weakly informative priors for population variation in a physiological model

- Pharamcokinetic parameters such as the "Michaelis-Menten coefficient"
- ► Wide uncertainty: prior guess for θ is 15 with a factor of 100 of uncertainty, log θ ~ N(log(15), log(10)²)
- ▶ Population model: data on several people j, log θ_j ~ N(log(15), log(10)²) ????
- Hierarchical prior distribution:
 - $\log \theta_j \sim N(\mu, \sigma^2), \ \sigma \approx \log(2)$
 - $\mu \sim N(\log(15), \log(10)^2)$
- Weakly informative

・ 同 ト ・ ヨ ト ・ ヨ ト …

Conclusions Extra stuff

Weakly informative priors for mixture models

- Well-known problem of fitting the mixture model likelihood
- The maximum likelihood fits are weird, with a single point taking half the mixture
- Bayes with flat prior is just as bad
- These solutions don't "look" like mixtures
- There must be additional prior information—or, to put it another way, regularization
- Simple constraints, for example, a prior dist on the variance ratio
- Weakly informative

・ロト ・回ト ・ヨト ・ヨト

Conclusions Extra stuff

Weakly informative priors for mixture models

Well-known problem of fitting the mixture model likelihood

- The maximum likelihood fits are weird, with a single point taking half the mixture
- Bayes with flat prior is just as bad
- These solutions don't "look" like mixtures
- There must be additional prior information—or, to put it another way, regularization
- Simple constraints, for example, a prior dist on the variance ratio
- Weakly informative

・ロン ・回 と ・ ヨ と ・ ヨ と

Conclusions Extra stuff

Weakly informative priors for mixture models

- Well-known problem of fitting the mixture model likelihood
- The maximum likelihood fits are weird, with a single point taking half the mixture
- Bayes with flat prior is just as bad
- These solutions don't "look" like mixtures
- There must be additional prior information—or, to put it another way, regularization
- Simple constraints, for example, a prior dist on the variance ratio
- Weakly informative

・ロン ・回 と ・ 回 と ・ 回 と

3

Conclusions Extra stuff

Weakly informative priors for mixture models

- Well-known problem of fitting the mixture model likelihood
- The maximum likelihood fits are weird, with a single point taking half the mixture
- Bayes with flat prior is just as bad
- These solutions don't "look" like mixtures
- There must be additional prior information—or, to put it another way, regularization
- Simple constraints, for example, a prior dist on the variance ratio
- Weakly informative

Conclusions Extra stuff

Weakly informative priors for mixture models

- Well-known problem of fitting the mixture model likelihood
- The maximum likelihood fits are weird, with a single point taking half the mixture
- Bayes with flat prior is just as bad
- These solutions don't "look" like mixtures
- There must be additional prior information—or, to put it another way, regularization
- Simple constraints, for example, a prior dist on the variance ratio
- Weakly informative

Conclusions Extra stuff

Weakly informative priors for mixture models

- Well-known problem of fitting the mixture model likelihood
- The maximum likelihood fits are weird, with a single point taking half the mixture
- Bayes with flat prior is just as bad
- These solutions don't "look" like mixtures
- There must be additional prior information—or, to put it another way, regularization
- Simple constraints, for example, a prior dist on the variance ratio
- Weakly informative

Conclusions Extra stuff

Weakly informative priors for mixture models

- Well-known problem of fitting the mixture model likelihood
- The maximum likelihood fits are weird, with a single point taking half the mixture
- Bayes with flat prior is just as bad
- These solutions don't "look" like mixtures
- There must be additional prior information—or, to put it another way, regularization
- Simple constraints, for example, a prior dist on the variance ratio
- Weakly informative

Conclusions Extra stuff

Weakly informative priors for mixture models

- Well-known problem of fitting the mixture model likelihood
- The maximum likelihood fits are weird, with a single point taking half the mixture
- Bayes with flat prior is just as bad
- These solutions don't "look" like mixtures
- There must be additional prior information—or, to put it another way, regularization
- Simple constraints, for example, a prior dist on the variance ratio
- Weakly informative

イロン イ部ン イヨン イヨン 三日

Conclusions Extra stuff

Intentional underpooling in hierarchical models

Basic hierarchical model:

- Data y_i on parameters θ_i
- Group-level model $\theta_j \sim N(\mu, \tau^2)$
- ▶ No-pooling estimate $\hat{\theta}_i = y_i$
- \succ Bayesian partial-pooling estimate $E(\theta_i|y)$
- Weak Bayes estimate: same as Bayes, but replacing au with 2 au
- An example of the "incompatible Gibbs" algorithm
- Why would we do this??

・ロト ・回ト ・ヨト ・ヨト

Conclusions Extra stuff

Intentional underpooling in hierarchical models

Basic hierarchical model:

- Data y_j on parameters θ_j
- Group-level model $\theta_j \sim N(\mu, \tau^2)$
- No-pooling estimate $\hat{\theta}_j = y_j$
- Bayesian partial-pooling estimate $E(\theta_j|y)$
- \blacktriangleright Weak Bayes estimate: same as Bayes, but replacing au with 2 au
- An example of the "incompatible Gibbs" algorithm
- Why would we do this??

イロン イヨン イヨン イヨン

Conclusions Extra stuff

Intentional underpooling in hierarchical models

Basic hierarchical model:

- Data y_j on parameters θ_j
- Group-level model $\theta_j \sim N(\mu, \tau^2)$
- No-pooling estimate $\hat{\theta}_j = y_j$
- Bayesian partial-pooling estimate $E(\theta_j|y)$
- \blacktriangleright Weak Bayes estimate: same as Bayes, but replacing au with 2 au
- An example of the "incompatible Gibbs" algorithm
- Why would we do this??

イロン イヨン イヨン イヨン

Conclusions Extra stuff

Intentional underpooling in hierarchical models

- Basic hierarchical model:
 - Data y_j on parameters θ_j
 - Group-level model $heta_j \sim N(\mu, \tau^2)$
 - No-pooling estimate $\hat{\theta}_j = y_j$
 - Bayesian partial-pooling estimate $E(\theta_j|y)$
- Weak Bayes estimate: same as Bayes, but replacing au with 2 au
- An example of the "incompatible Gibbs" algorithm
- Why would we do this??

イロン イヨン イヨン イヨン

Conclusions Extra stuff

Intentional underpooling in hierarchical models

- Basic hierarchical model:
 - Data y_j on parameters θ_j
 - Group-level model $\theta_j \sim N(\mu, \tau^2)$
 - No-pooling estimate $\hat{\theta}_j = y_j$
 - Bayesian partial-pooling estimate $E(\theta_j|y)$
- Weak Bayes estimate: same as Bayes, but replacing au with 2 au
- An example of the "incompatible Gibbs" algorithm
- Why would we do this??

・ロン ・回と ・ヨン・

Conclusions Extra stuff

Intentional underpooling in hierarchical models

- Basic hierarchical model:
 - Data y_j on parameters θ_j
 - Group-level model $\theta_j \sim N(\mu, \tau^2)$
 - No-pooling estimate $\hat{\theta}_j = y_j$
 - Bayesian partial-pooling estimate $E(\theta_j|y)$
- Weak Bayes estimate: same as Bayes, but replacing au with 2 au
- An example of the "incompatible Gibbs" algorithm
- Why would we do this??

イロト イポト イヨト イヨト

Conclusions Extra stuff

Intentional underpooling in hierarchical models

- Basic hierarchical model:
 - Data y_j on parameters θ_j
 - Group-level model $\theta_j \sim N(\mu, \tau^2)$
 - No-pooling estimate $\hat{\theta}_j = y_j$
 - Bayesian partial-pooling estimate $E(\theta_j|y)$
- Weak Bayes estimate: same as Bayes, but replacing au with 2 au
- An example of the "incompatible Gibbs" algorithm
- Why would we do this??

イロト イポト イヨト イヨト

Conclusions Extra stuff

Intentional underpooling in hierarchical models

- Basic hierarchical model:
 - Data y_j on parameters θ_j
 - Group-level model $\theta_j \sim N(\mu, \tau^2)$
 - No-pooling estimate $\hat{\theta}_j = y_j$
 - Bayesian partial-pooling estimate $E(\theta_j|y)$
- Weak Bayes estimate: same as Bayes, but replacing au with 2 au
- An example of the "incompatible Gibbs" algorithm
- Why would we do this??

・ 同 ト ・ ヨ ト ・ ヨ ト

Conclusions Extra stuff

Intentional underpooling in hierarchical models

- Basic hierarchical model:
 - Data y_j on parameters θ_j
 - Group-level model $\theta_j \sim N(\mu, \tau^2)$
 - No-pooling estimate $\hat{\theta}_j = y_j$
 - Bayesian partial-pooling estimate $E(\theta_j|y)$
- Weak Bayes estimate: same as Bayes, but replacing au with 2 au
- An example of the "incompatible Gibbs" algorithm
- Why would we do this??

(1) マン・ション・