Challenges on the use of Imprecise Prior

for Imprecise Inference
on Poisson Sampling Models

Chel Hee Lee[1] Miķelis Bickis[2]

[1]Clinical Research Support Unit
Community Health and Epidemiology
College of Medicine
University of Saskatchewan

[2]Department of Mathematics and Statistics
University of Saskatchewan

9th WPMSIIP, Durham, England
4 September 2016

General Framework

P. I. N.	Count
X8213	0
X8222	2
X8223	1
X8224	1
X8225	3
X8226	2
X8227	7
X8227	7
X8227	7

General Framework

P. I. N.	Count
X8213	0
X8222	2
X8223	1
X8224	1
X8225	3
X8226	2
X8227	7
X8227	7
X8227	7

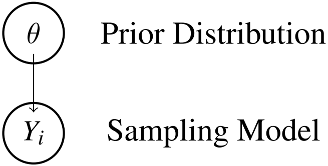
General Framework

P. I. N.	Count
X8213	0
X8222	2
X8223	1
X8224	1
X8225	3
X8226	2
X8227	7
X8227	7
X8227	7

Y_i Sampling Model

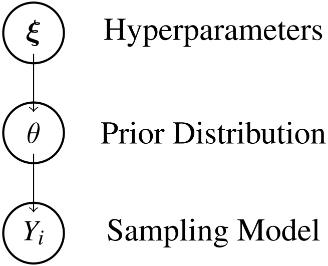
General Framework

P. I. N.	Count
X8213	0
X8222	2
X8223	1
X8224	1
X8225	3
X8226	2
X8227	7
X8227	7
X8227	7



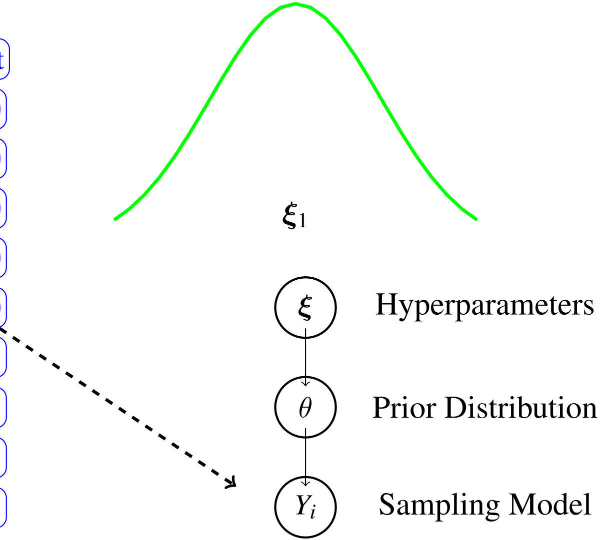
General Framework

P. I. N.	Count
X8213	0
X8222	2
X8223	1
X8224	1
X8225	3
X8226	2
X8227	7
X8227	7
X8227	7

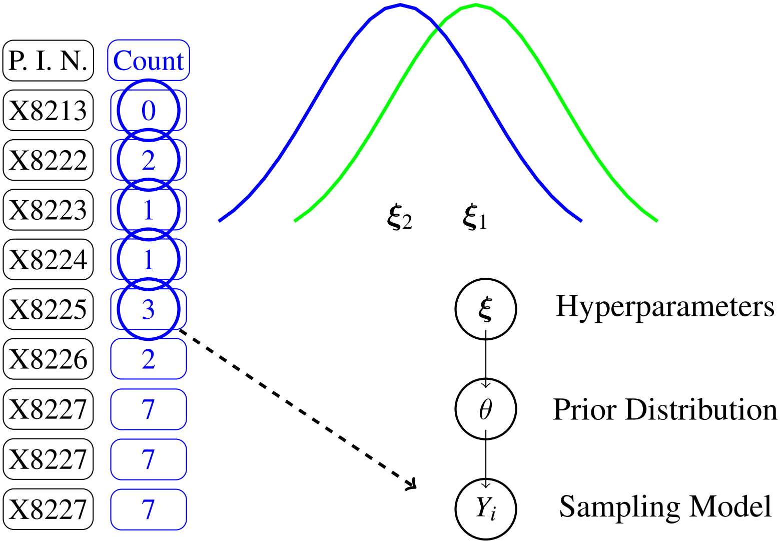


General Framework

P. I. N.	Count
X8213	0
X8222	2
X8223	1
X8224	1
X8225	3
X8226	2
X8227	7
X8227	7
X8227	7

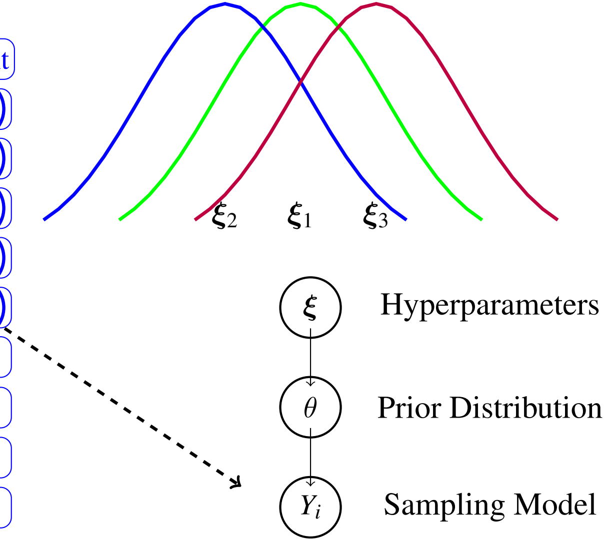


General Framework

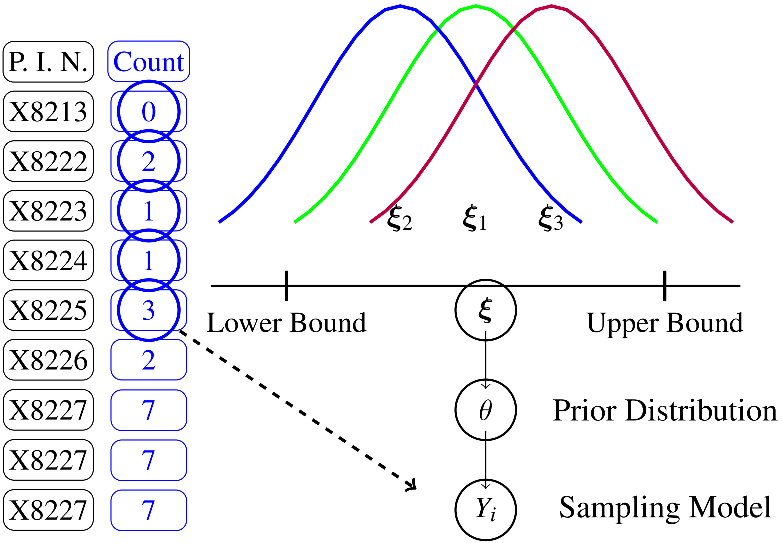


General Framework

P. I. N.	Count
X8213	0
X8222	2
X8223	1
X8224	1
X8225	3
X8226	2
X8227	7
X8227	7
X8227	7



General Framework



Robust Bayesian Analysis (Berger et al., 1994, pp. 24–25)

A prior distribution should be

- easy to elicit and interpret,
- easy to handle computationally,
- reasonable to reflect uncertainty,
- extensible to higher dimensions, and
- adaptable to incorporate constraints

HOW?

Imprecise Inferential Framework

Canonical Representation

Consider a family of probability measures P_θ whose density with respect to μ:

$$dP_\theta(y) = \exp\{\theta \cdot t(y) - A(\theta)\}d\mu(y),$$

where $t : R^m \to R^k$ is a measurable function of y and the cumulant transform

$$A(\theta) = \ln \int \exp\{\theta \cdot t(y)\}d\mu(y)$$

serves to normalize the measure P_θ.

Imprecise Inferential Framework

Conjugate Prior Formulation

We consider the following family of prior measures for P_θ with respect to Lebesgue measure:

$$d\pi_{\xi_2,\xi_1,\xi_0}(\theta) \propto \exp\{-\xi_2\theta^2 + \theta\xi_1 - \xi_0 A(\theta) - M(\xi_2, \xi_1, \xi_0)\}d\theta,$$

where $\xi = (\xi_2, \xi_1, \xi_0)$ are hyperparameters and

$$M(\xi_2, \xi_1, \xi_0) = \ln \int_{-\infty}^{\infty} \exp\{-\xi_2\theta^2 + \theta\xi_1 - \xi_0 A(\theta)\}d\theta < +\infty$$

is the cumulant transform of ξ producing the densities $\pi_\xi(\theta)$.

Illustration (based on Poisson samples)

Consider the problem of parameter estimation

- (Scenario 1) when a prior is conjugate to a likelihood
  - using a log-gamma prior distribution
- (Scenario 2) when a prior is not conjugate to a likelihood
  - using a normal prior distribution
- (Scenario 3) under the generalized linear model setting
  - having only intercept
  - incorporating a single predictor (with an intercept)

Scenario I

Using Log-Gamma Priors

If $\mu \sim Gamma(\alpha, \beta)$ and $\theta = \log(\mu)$,

$$\pi_{\alpha,\beta}(\theta) \propto e^{\alpha\theta - \beta e^{\theta}}$$

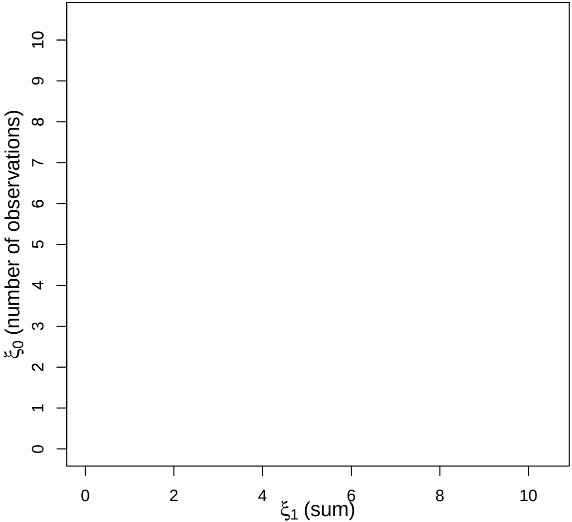
We see the following

$$\begin{aligned} p(\theta|y) &\propto e^{(y\theta - e^{\theta})} e^{(\alpha\theta - \beta e^{\theta})} \\ &= e^{(\alpha+y)\theta - (\beta+1)e^{\theta}} \end{aligned}$$

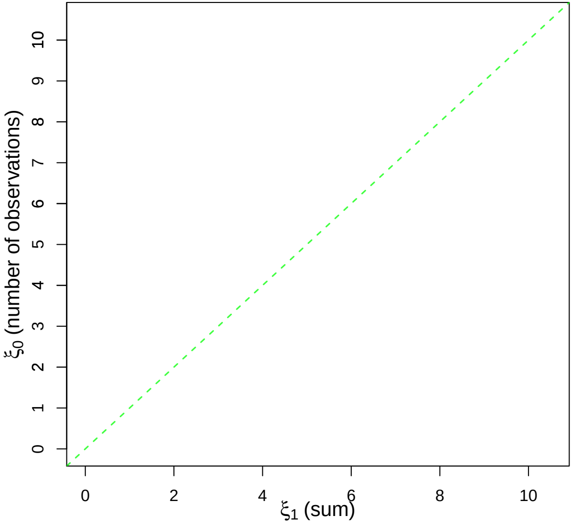
which has the form $p(\theta|y) \propto \exp(-\xi_2\theta^2 + \xi_1\theta - \xi_0 e^{\theta})$ with hyperparameters

$$\xi_2 = 0, \quad \xi_1 = \alpha_1 + y, \quad \xi_0 = \beta + 1, \tag{1}$$

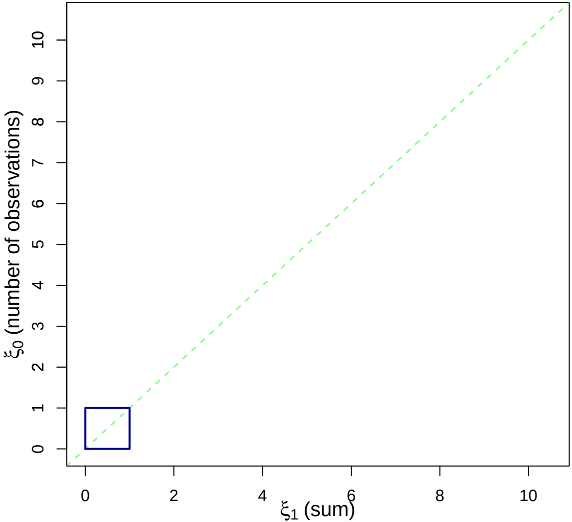
Set Basic Analytical Frame



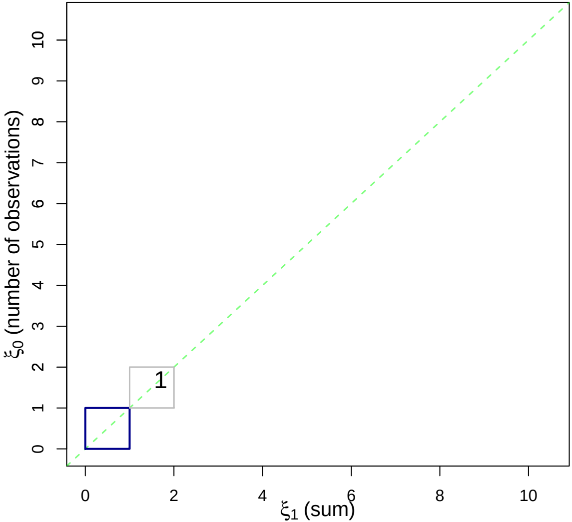
Assume Poisson Mean Parameter ($\mu = 1$)



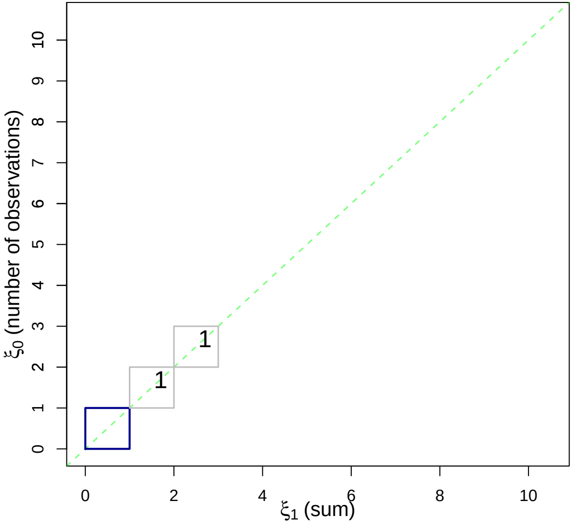
Before Seeing Data



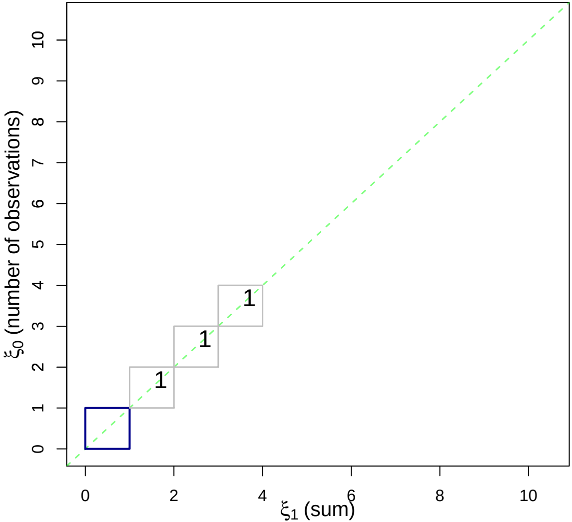
After Seeing One Observation



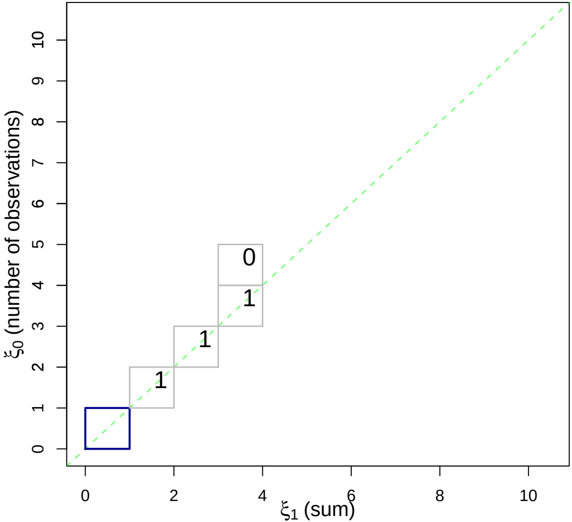
One More Observation



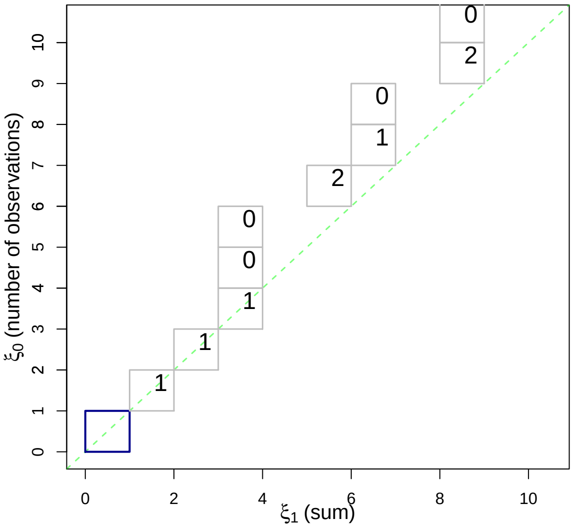
See What Data Tell Us



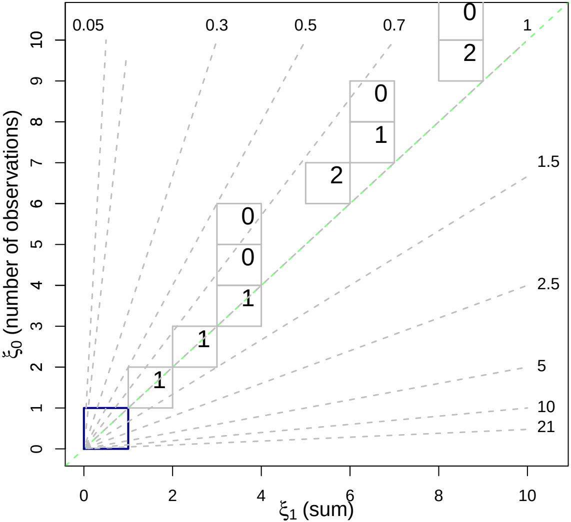
Continue to Watch!



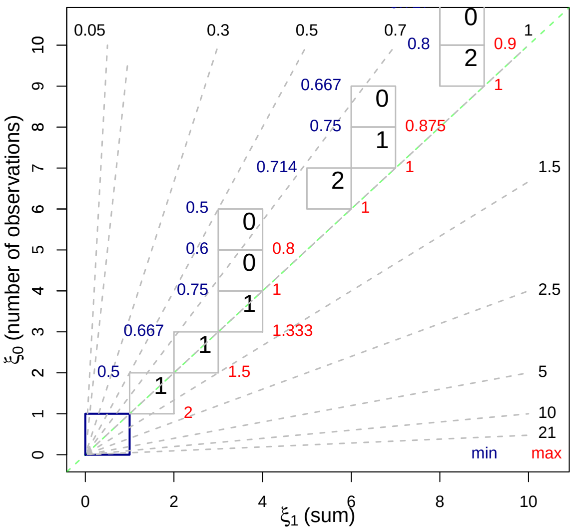
Learning Process from Ten Observations



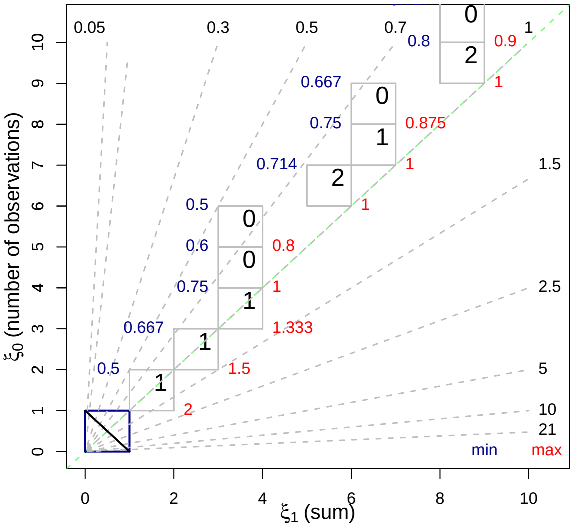
Contour Levels of Prior Expectation E[Y]



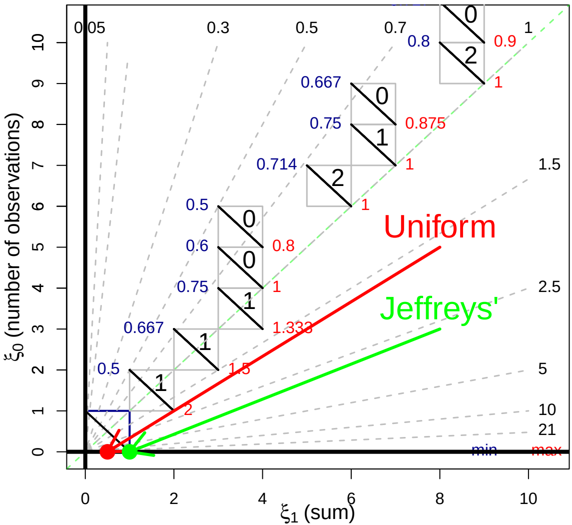
Computational Efficiency



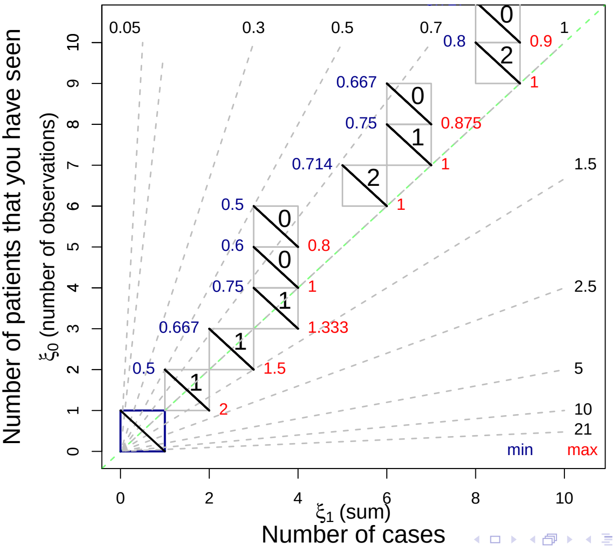
Elicitation Under Almost Complete Prior Ignorance



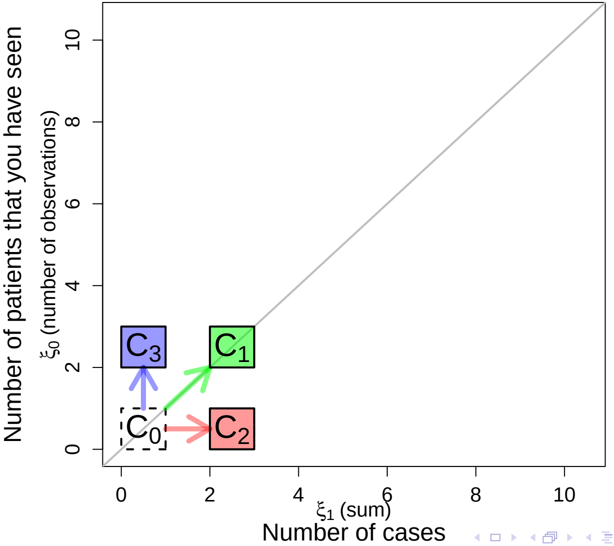
Elicitation – Improper Priors



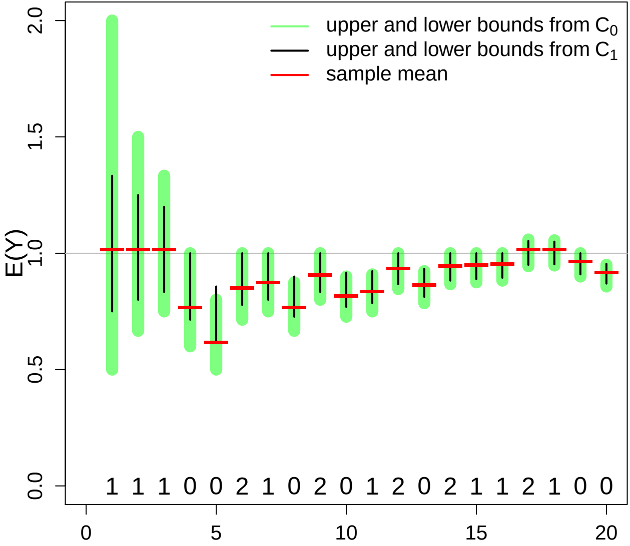
Meaningful Interpretation for Communication



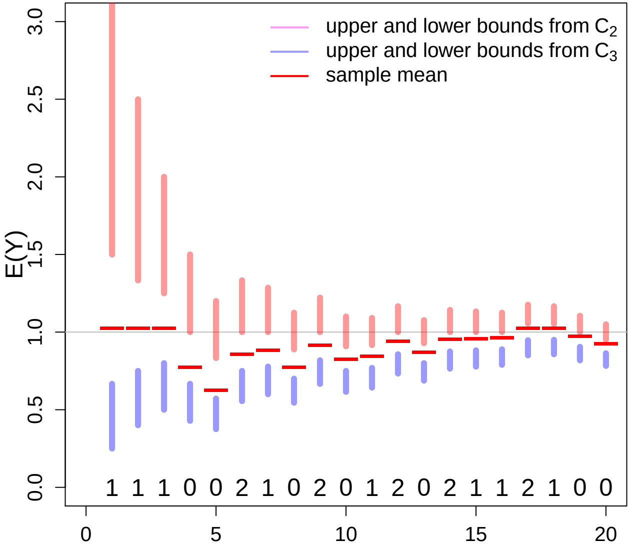
All Priors Are Informative In Some Way



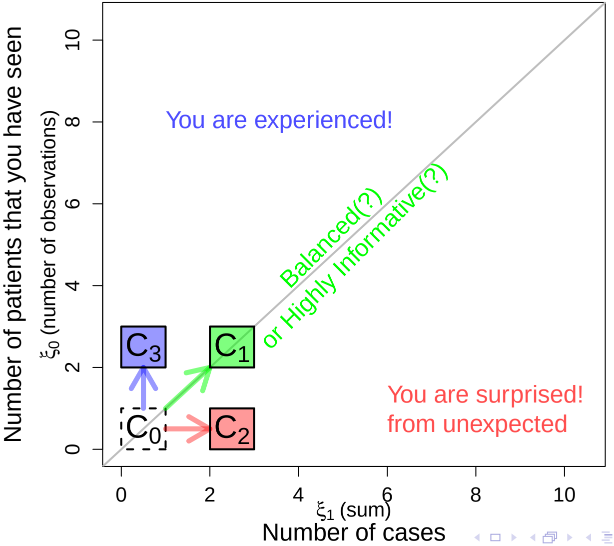
Learning Curve of Imprecise Prior Capturing True



Learning Curve of Imprecise Prior Not Capturing True



Some Priors Are More Informative Than The Others?



Scenario II

Using Normal Priors

If $\mu \sim LN(\log(\nu), \tau^2)$ and $\theta = \log(\mu)$,

$$\pi_{\nu,\tau}(\theta) \propto \exp\left(-\frac{1}{2\tau^2}\theta^2 + \frac{\nu}{\tau^2}\theta\right), \tag{2}$$

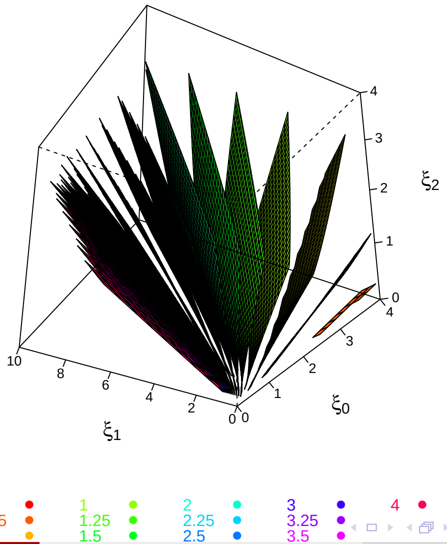
We see the following

$$p(\theta|y) \propto \exp(y\theta - e^{\theta})\ \exp\left(-\frac{1}{2\tau^2}\theta^2 + \frac{\nu}{\tau^2}\theta\right). \tag{3}$$

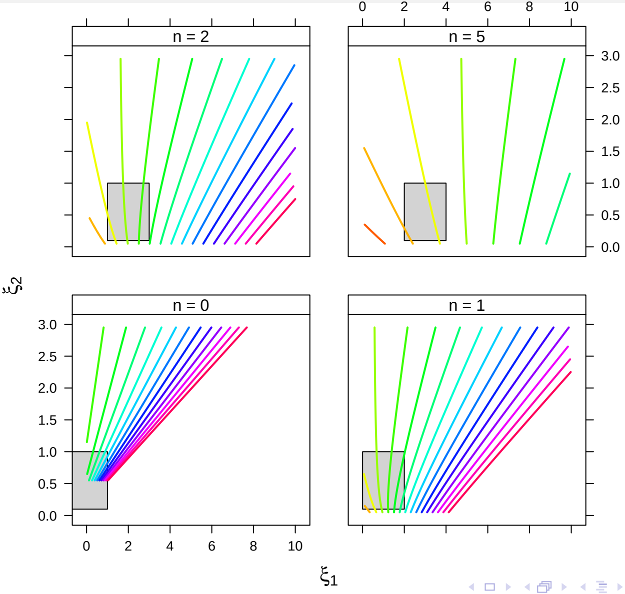
which has the form $p(\theta|y) \propto \exp(-\xi_2\theta^2 + \xi_1\theta - \xi_0 e^{\theta})$ with hyperparameters

$$\xi_2 = \frac{1}{2\tau^2}, \quad \xi_1 = \frac{\nu}{\tau^2} + y, \quad \xi_0 = 1, \tag{4}$$

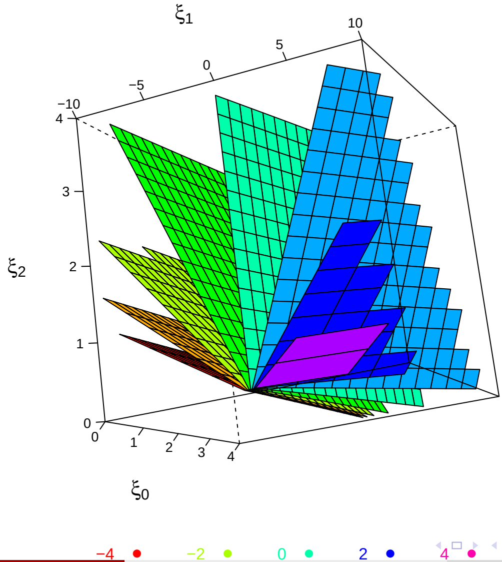
Level Sets of Prior Expectation E[Y]



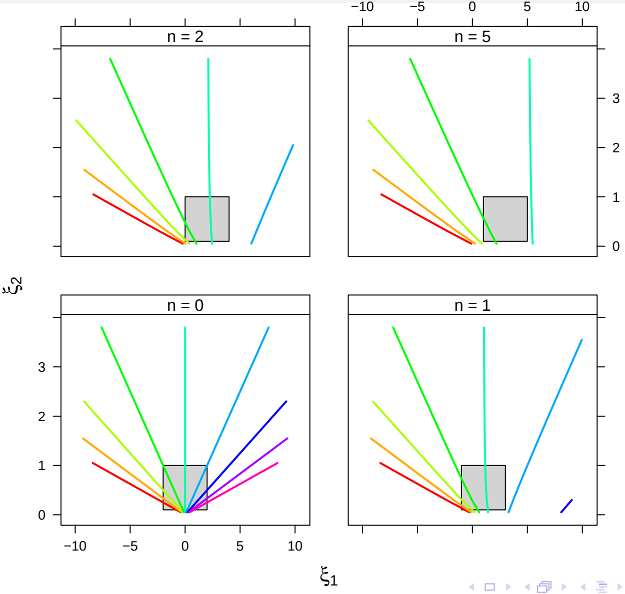
Level Sets of Prior Expectation E[Y]



Level Sets of Prior Expectation E[θ] (Intercept Model)

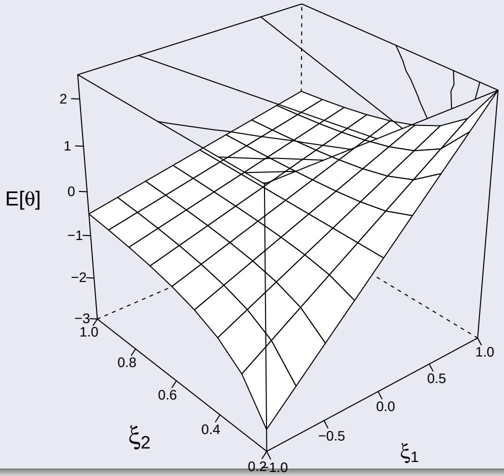


Level Sets of Prior Expectation E[θ] (Intercept Model)



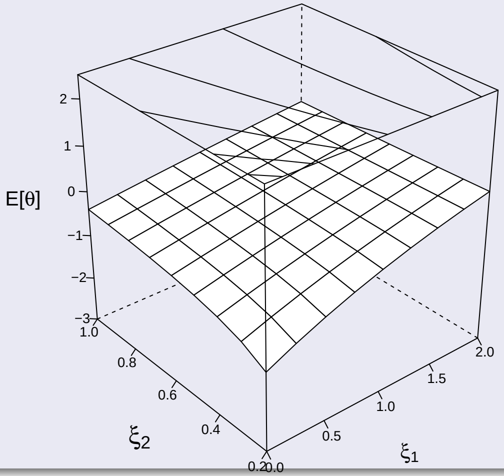
Soft-Linearity

$n = 0$



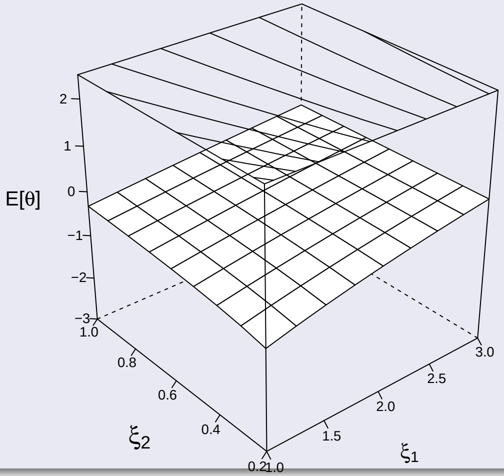
Soft-Linearity

n = 1



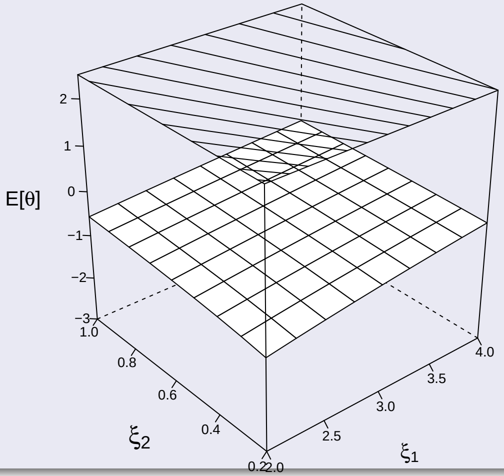
Soft-Linearity

n = 2

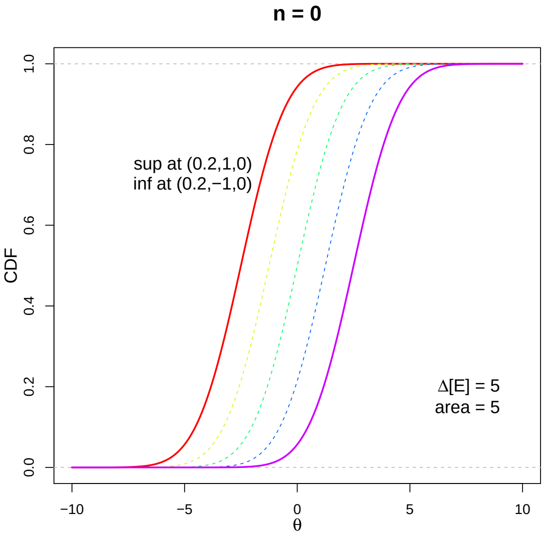


Soft-Linearity

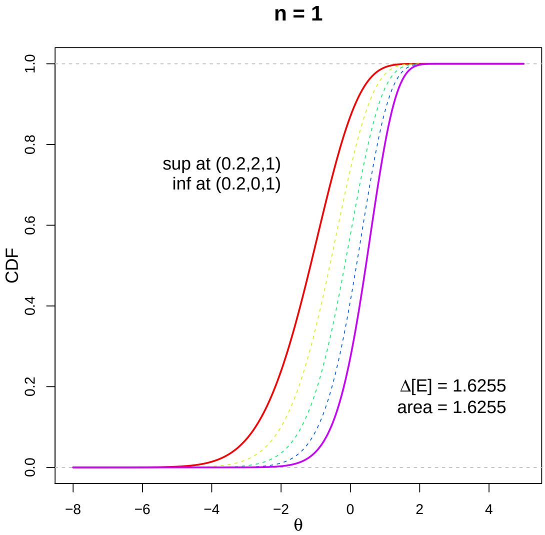
$n = 5$



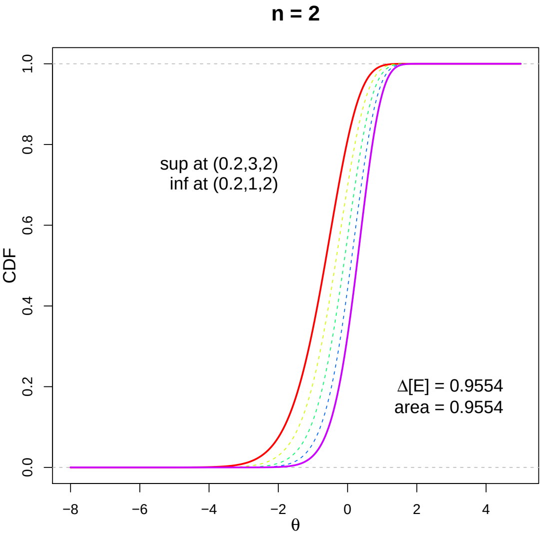
Focusing Feature



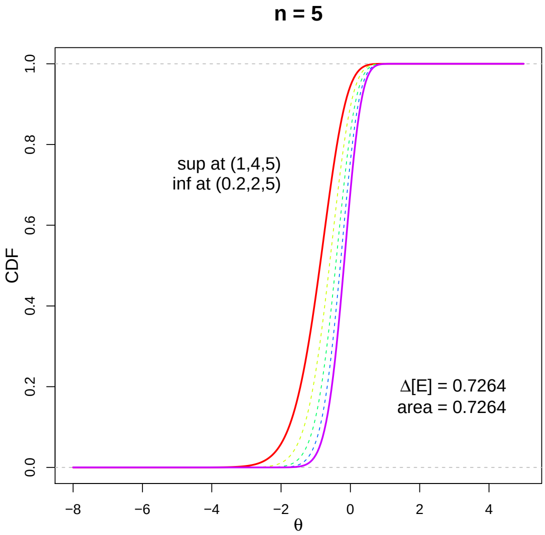
Focusing Feature



Focusing Feature



Focusing Feature



Graphical Demonstration

Simulation Design

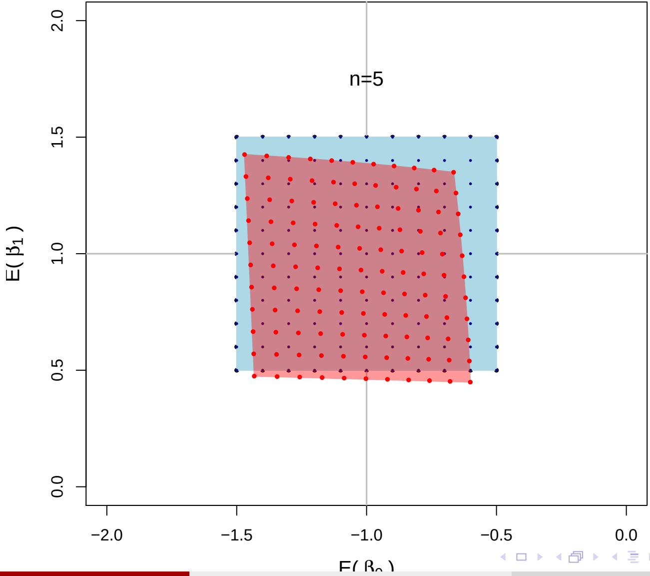
$$E(\boldsymbol{\beta}|\mathbf{y},\mathbf{X}) = \frac{\int \boldsymbol{\beta} f(\mathbf{y}|\mathbf{X},\boldsymbol{\beta})\pi(\boldsymbol{\beta})\, d\boldsymbol{\beta}}{\int f(\mathbf{y}|\mathbf{X},\boldsymbol{\beta})\pi(\boldsymbol{\beta})\, d\boldsymbol{\beta}}$$

where $\mathbf{X} \sim MVN_2\left(\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0 & 0\\0 & 1\end{bmatrix}\right)$, $\pi(\boldsymbol{\beta}) \sim MVN_2\left(\mathbf{b}, \mathbf{B}\right)$ such as

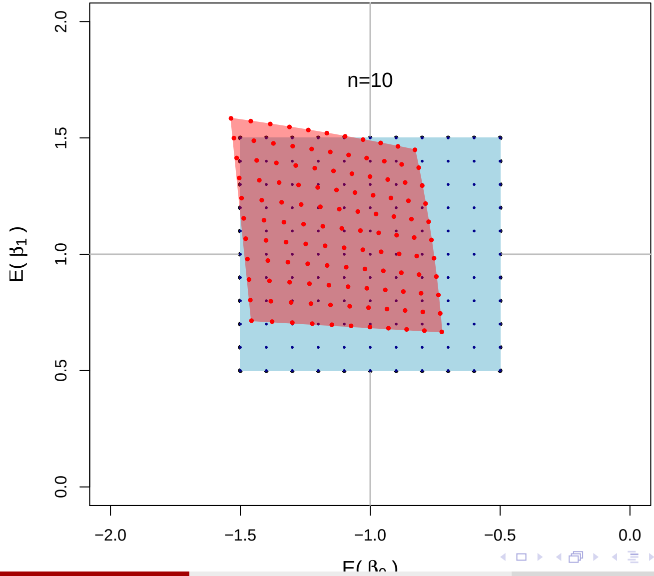
$$\mathbf{b} = \begin{bmatrix}\mu_0\\ \mu_1\end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix}\sigma_0 & 0\\ 0 & \sigma_1\end{bmatrix}\begin{bmatrix}1 & \rho\\ \rho & 1\end{bmatrix}\begin{bmatrix}\sigma_0 & 0\\ 0 & \sigma_1\end{bmatrix}$$

Metropolis-Hasting algorithm, Laplace approximation, Importance sampling methods are used for approximation.

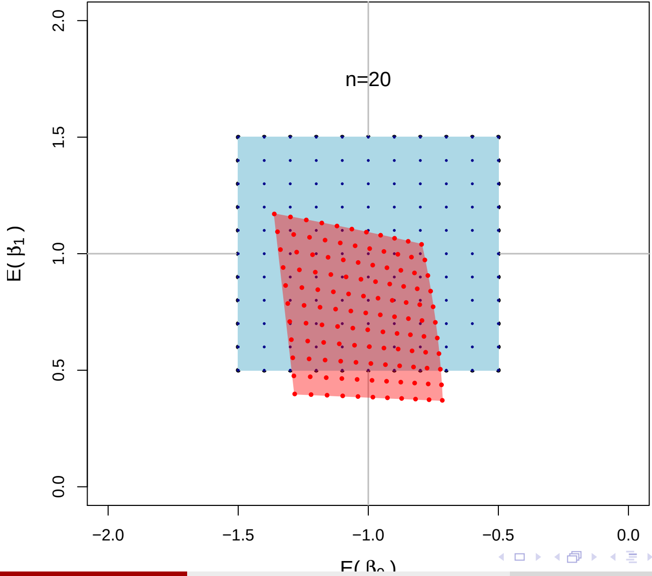
Focusing Feature



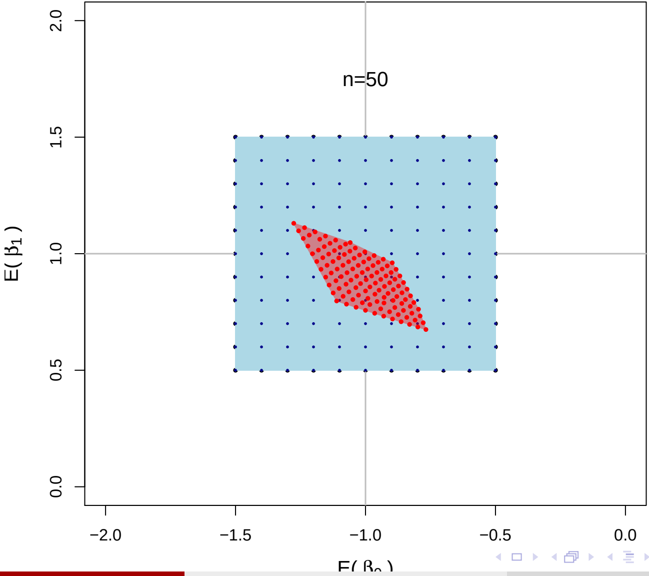
Focusing Feature



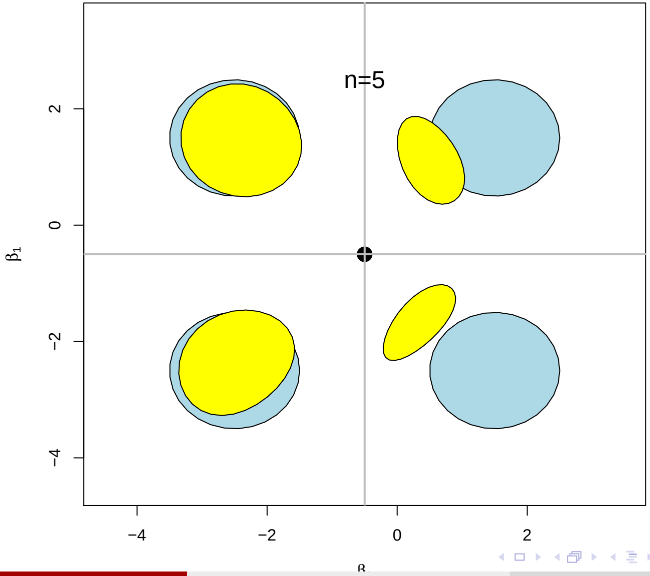
Focusing Feature



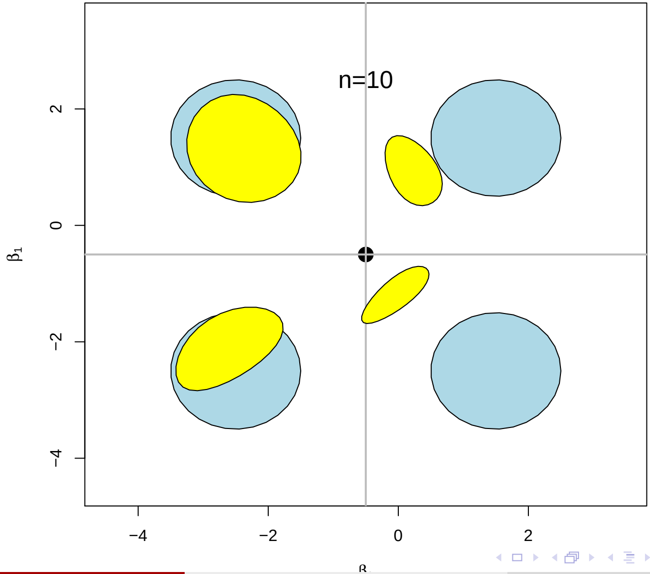
Focusing Feature



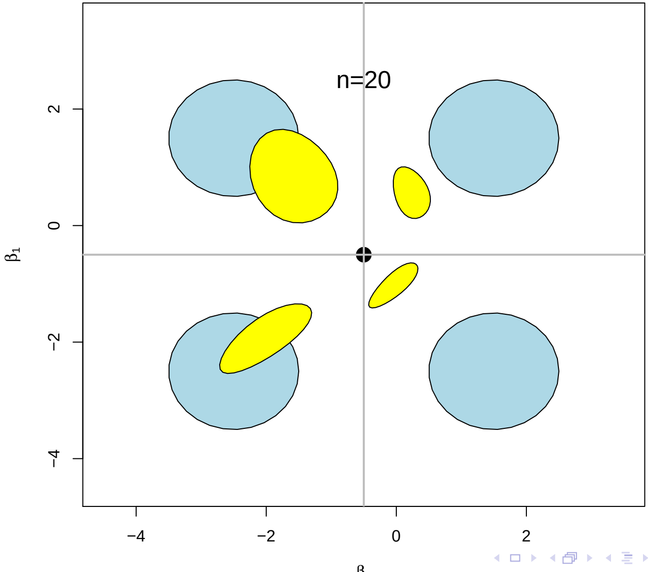
Agreement Between Intentional Units



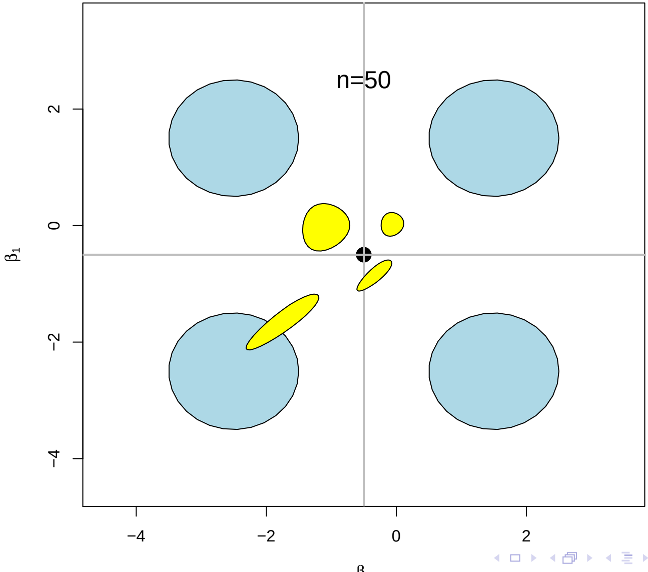
Agreement Between Intentional Units



Agreement Between Intentional Units



Agreement Between Intentional Units



Working in Progress

- Extention to other sampling models
  - Binomial distribution
  - Geometric distribution
  - Exponential distribution
  - Normal distribution (with known variance)
  - Normal distribution (with known mean)
- Incorporation of Kullback-Leibler divergence measure
- Imprecise inference for comparing two groups
- Dicision problem in practice
- Software development

Discussion

Advantage of this approach

- Extension to generalized linear model
- Easy implementation in software

Study Design

Imprecision on Data

- (systematic) incomplete data
- misclassified data
- missing data
- partially informative

Model imprecision

- Midspecified model
- Partially correct model

References I

Berger, J., Moreno, E., Pericchi, L., Bayarri, M., Bernardo, J., Cano, J., De la Horra, J., Martín, J., Ríos-Insúa, D., Betrò, B., Dasgupta, A., Gustafson, P., Wasserman, L., Kadane, J., Srinivasan, C., Lavine, M., O'Hagan, A., Polasek, W., Robert, C., Goutis, C., Ruggeri, F., Salinetti, G., and Sivaganesan, S. (1994). An overview of robust Bayesian analysis. Test, 3(1):5–124.

Diaconis, P. and Ylvisaker, D. (1979). Conjugate Priors for Exponential Families. Ann. Statist., 7(2):269–281.

Irony, T. and Singpurwalla, N. (1997). Noninformative priors do not exist: a Discussion with Jose M. Bernardo. Journal of Statistical Inference and Planning, 65:159–189.

Walley, P. (1991). Statistical reasoning with imprecise probabilities. Chapman and Hall, London;.