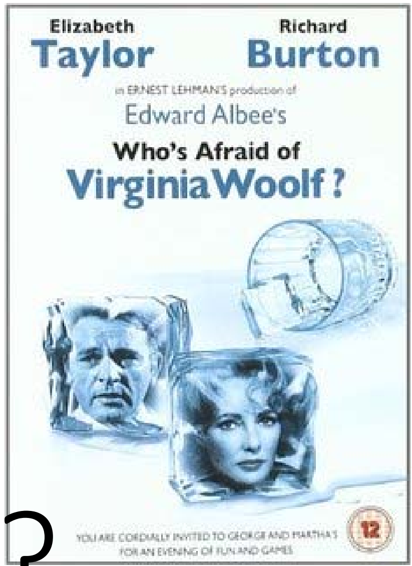


Who‘s afraid of …
Bayesian non-collapsibility?

Georg Heinze and Angelika Geroldinger

Medical University of Vienna

An example

- Simple 2 x 2 table:

	X=0	X=1	
Y=0	7	1	*8*
Y=1	2	4	*6*
	9	*5*	*14*

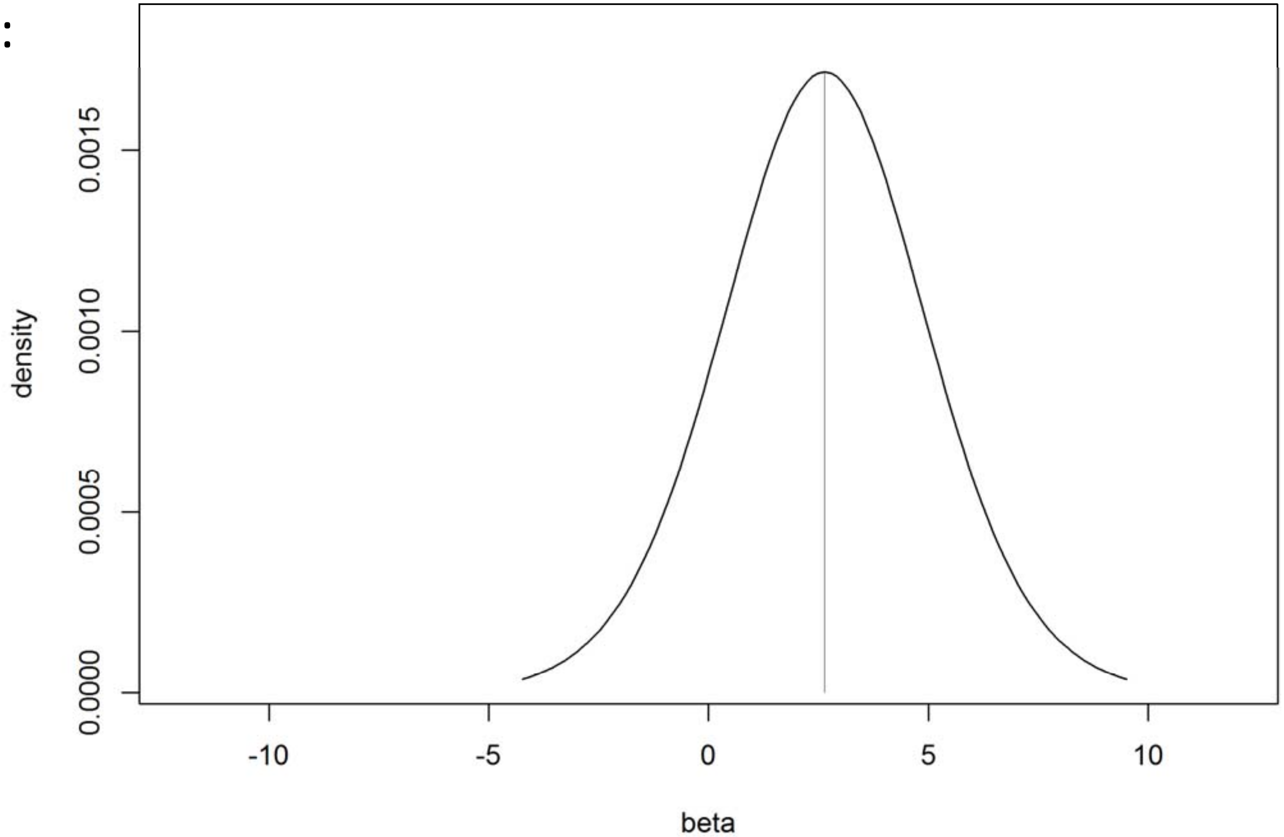
- Suppose we are interested in the log odds ratio of X:
- $\beta_1 = \log \frac{n_{11}/n_{10}}{n_{01}/n_{00}} = \log \frac{4/1}{2/7} = 2.6$

- The likelihood for the example:
- $L(\beta|x,y) =$

$$\pi(X=1)^{n_{11}}$$
$$\left(1-\pi(X=1)\right)^{n_{10}}$$
$$\pi(X=0)^{n_{01}}$$
$$\left(1-\pi(X=0)\right)^{n_{00}}$$

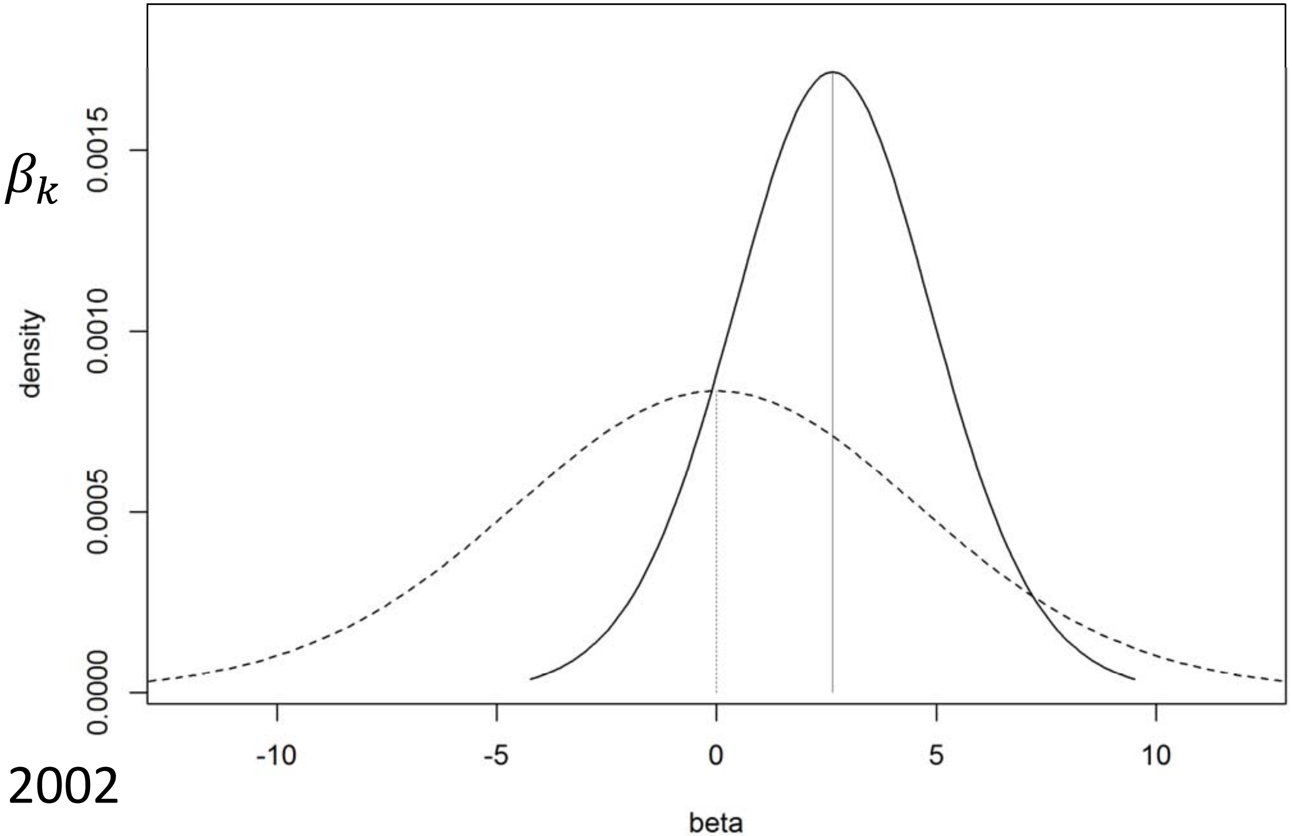
- with

$$\pi(X=x) =$$
$$1/(1+\exp(-\beta_0-\beta_1 x)$$



Jeffreys prior: $p(\beta) = |I(\beta)|^{1/2}$

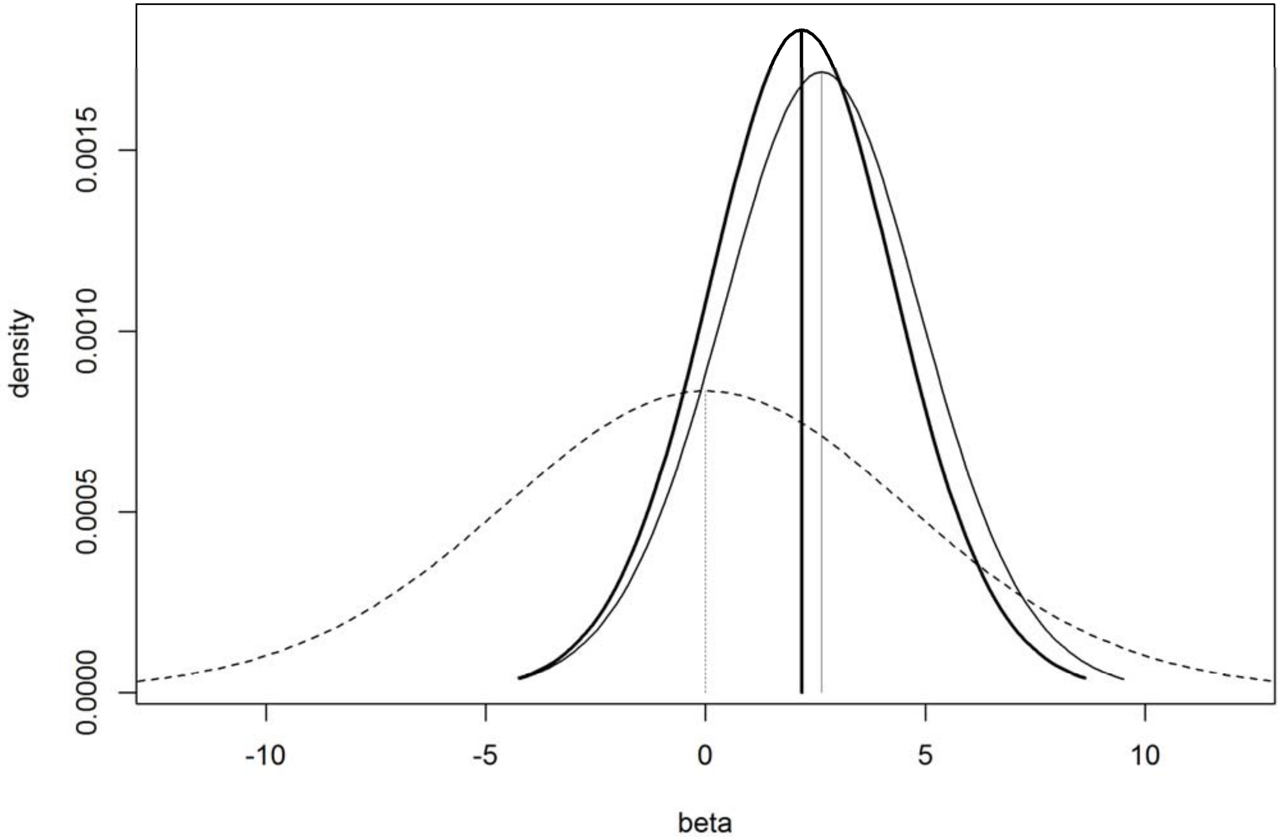
$$I(\beta)_{jk} = -\partial \log L(\beta) / \partial\beta_j \partial\beta_k$$



Weakly informative prior
Automatic solution
Nice properties

Chen et al JASA 2008,
Firth Biometrika 1993,
Heinze and Schemper StatMed 2002

- The posterior:
$$p(\beta|x,y) = p(\beta)L(\beta|x,y)$$
- As expected,
the posterior ‚collapses‘ prior and likelihood



Using priors in practice

• General prior	Derive posterior by simulation (MCMC)
• Ridge regression Firth‘s method/Jeffreys	Prior can be expressed as likelihood penalty
• Conjugate prior	Such that posterior has same algebraic form as prior, can be expressed as pseudo-observations *(data augmentation prior)*

- In special cases Jeffreys prior reduces to data augmentation

Firth Biometrika 1993; Jeffreys JASA 1946

An example

- Augmented 2 x 2 table:

	X=0	X=1	
Y=0	7.5	1.5	*9*
Y=1	2.5	4.5	*7*
	10	*6*	*16*

- Maximization of the likelihood of augmented table is now equivalent to finding the posterior mode with original data and Jeffreys prior

Example of Greenland 2010

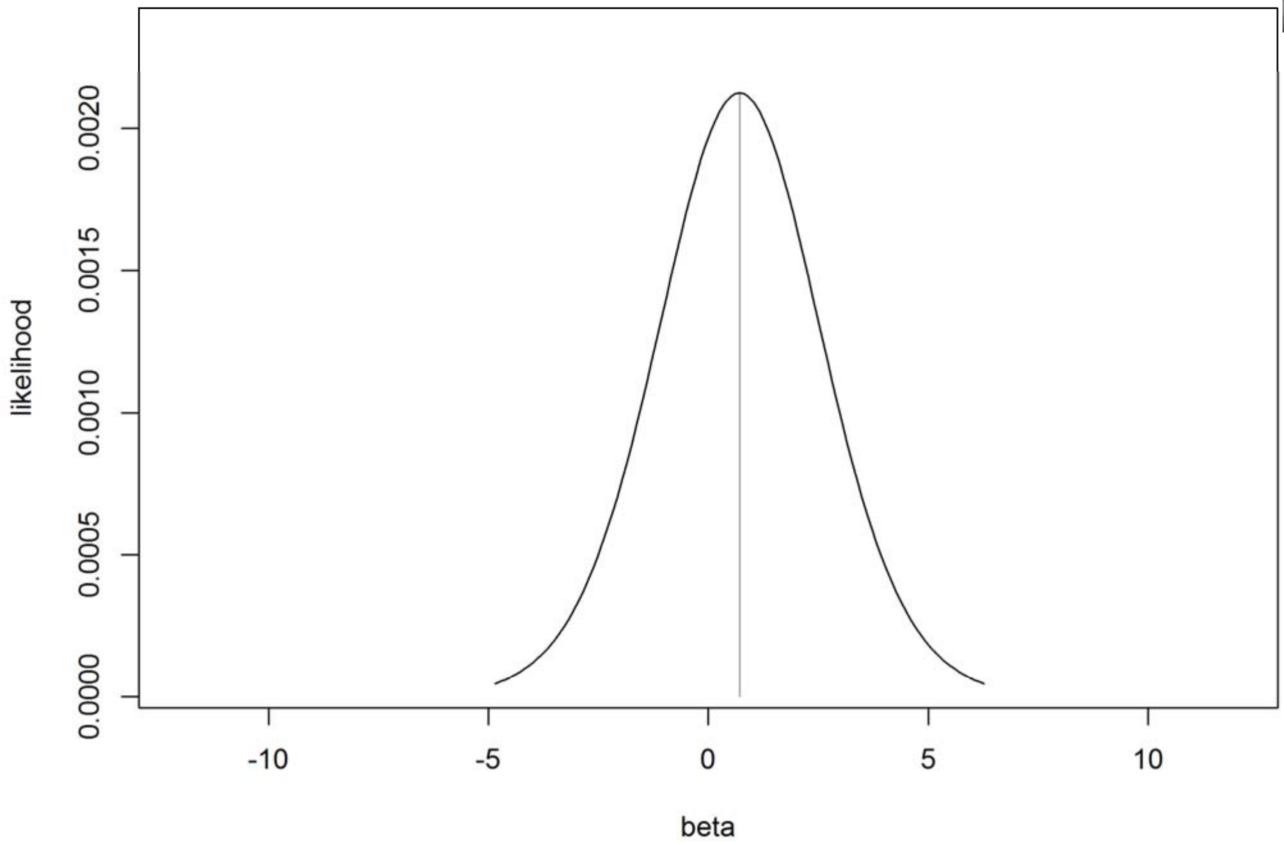
- 2x2 table

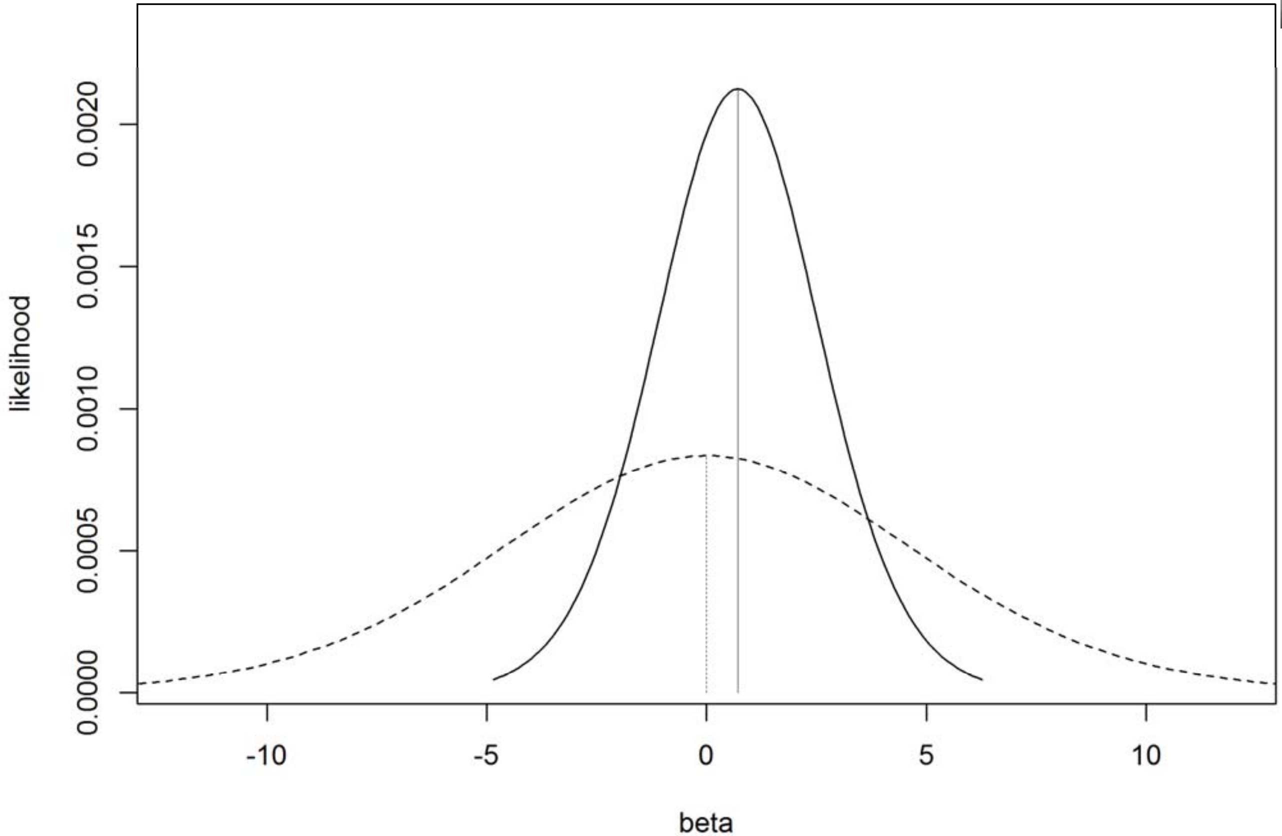
	X=0	X=1	
Y=0	315	5	*320*
Y=1	31	1	*32*
	346	*6*	*352*

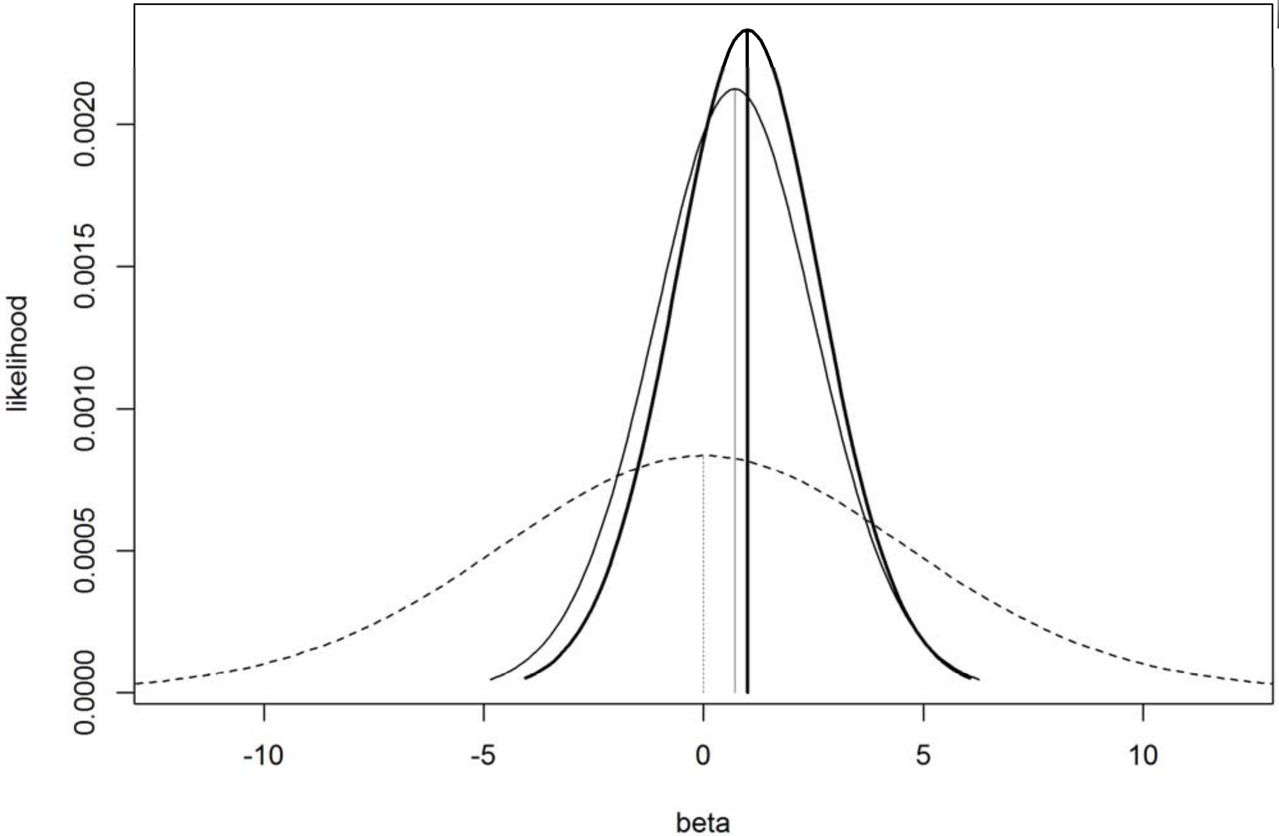
← Rare events

↑ Rare exposure

Greenland, AmStat 2010


likelihood
0.0000
0.0005
0.0010
0.0015
0.0020
-10
-5
0
5
10
beta


likelihood
0.0000
0.0005
0.0010
0.0015
0.0020
-10
-5
0
5
10
beta

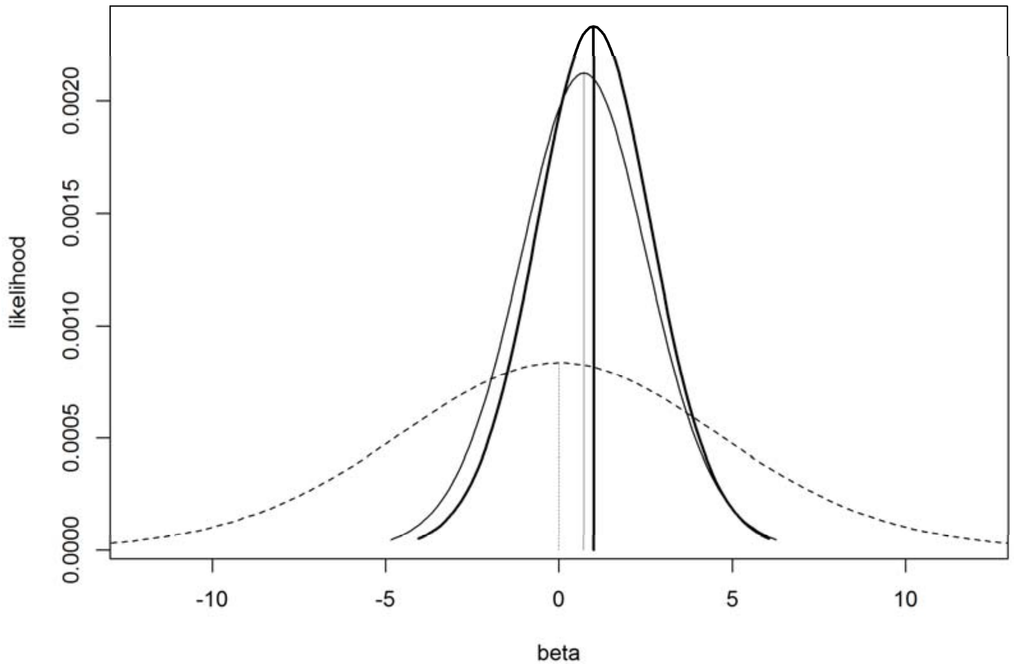


likelihood
0.0000
0.0005
0.0010
0.0015
0.0020
-10
-5
0
5
10
beta

Bayesian non-collapsibility:

- Prior and likelihood do not ‚collapse'
- The posterior mode is more extreme than the ML estimate

- How can that happen???



An even more extreme example from Greenland 2010

- 2x2 table

	X=0	X=1	
Y=0	25	5	*30*
Y=1	5	1	*6*
	30	*6*	*36*

- Here we immediately see that the odds ratio = 1 ($\beta_1 = 0$)

- But the estimate from augmented data: odds ratio = 1.26 (try it out!)

Greenland, AmStat 2010

Reason for Bayesian non-collapsilibity

- We look at the association of X and Y

- We could treat the origin of data as a ‚ghost factor‘ G
- G=0 for original table
- G=1 for pseudo data

- We ignore that the conditional association of X and Y given G is different from the marginal association

Simulating the example of Greenland

- We should distinguish BNC in a single data set from a systematic increase in bias of a method (in simulations)

- (This is only of interest to frequentists)

- Simulation of the example:
- Fixed groups x=0 and x=1, P(Y=1|X) as observed in example
- True log OR=0.709

	X=0	X=1	
Y=0	315	5	*320*
Y=1	31	1	*32*
	346	*6*	*352*

Simulating the example of Greenland

- True value: log OR = 0.709

Parameter	ML	Jeffreys-Firth	
Bias β_1	*	+18%	
RMSE β_1	*	0.86	
Bayesian non-collapsibility $\boldsymbol{\beta_2}$		**63.7%**	

* Separation causes β_1 be undefined ($-\infty$) in 31.7% of the cases

Simulating the example of Greenland

- To overcome Bayesian non-collapsibility, Greenland and Mansournia (2015) have proposed not to impose a prior on the intercept

- They suggest a logF(1,1) prior for all other regression coefficients

- The method can be used with standard software because it is a data augmentation prior

Greenland and Mansournia, StatMed 2015

Simulating the example of Greenland

- Re-running the simulation with the logF(1,1) method yields:

Parameter	ML	Jeffreys-Firth	logF(1,1)
Bias β_1	*	+18%	
RMSE β_1	*	0.86	
Bayesian non-collapsibility $\boldsymbol{\beta_2}$		**63.7%**	**0%**

* Separation causes β_1 be undefined ($-\infty$) in 31.7% of the cases

Simulating the example of Greenland

- Re-running the simulation with the logF(1,1) method yields:

Parameter	ML	Jeffreys-Firth	logF(1,1)
Bias β_1	*	+18%	-52%
RMSE β_1	*	0.86	1.05
Bayesian non-collapsibility $\boldsymbol{\beta_2}$		**63.7%**	**0%**

* Separation causes β_1 be undefined ($-\infty$) in 31.7% of the cases

Other, more subtle occurrences of Bayesian non-collapsibility

- Ridge regression: normal prior around 0
- usually implies bias towards zero,
- But:
- With correlated predictors with different effect sizes, for some predictors the bias can be away from zero

Simulation of bivariable log reg models

- $X_1, X_2 \sim \text{Bin}(0.5)$ with correlation $r = 0.8, n = 50$
- $\beta_1 = 1.5, \beta_2 = 0.1$

Parameter	ML	Ridge (CV λ)	logF(1,1)	Jeffreys-Firth
Bias β_1	+40% (+9%*)	-26%	-2.5%	+1.2%
RMSE β_1	3.04 (1.02*)	1.01	0.73	0.79
Bias β_2	-451% (+16%*)	+48%	+77%	+16%
RMSE β_2	2.95 (0.81*)	0.73	0.68	0.76
Bayesian non-collapsibility β_2		**25%**	**28%**	**23%**

*excluding 2.7% separated samples

Conclusion

Bayesian:

- Bayesian non-collapsibility is usually unexpected
- Unintended in data analyses
- Can be avoided in univariable models,
  but no general rule to avoid it in multivariable models

Frequentist:

- Frequentist looks at repeated-sampling properties (bias, RMSE)
- Likelihood penalization can often decrease RMSE (even with BNC)
- Likelihood penalization ≠ guaranteed shrinkage
- Appropriate coverage of CI? Needs unbiased estimates!