# Prior-Data Conflict: a brief introduction

Gero Walter

Institut für Statistik
Ludwig-Maximilians-Universität München

September 7th, 2010

# Prior-Data Conflict

Prior-Data Conflict $\hat{=}$ situation in which...

- ...*informative prior beliefs* and *trusted data* (sampling model correct, no outliers, etc.) are in conflict
- "... the prior [places] its mass primarily on distributions in the sampling model for which the observed data is surprising." (Evans & Moshonov, 2006)

(...and there are not enough data to overrule prior beliefs!)

# Example: Dirichlet-Multinomial-Model

| | | | | |
|---|---|---|---|---|
| Data: | $\mathbf{k}$ | $\sim$ | $\mathrm{M}(\boldsymbol{\theta})$ | $(\sum k_j = n)$ |
| conjugate prior: | $\boldsymbol{\theta}$ | $\sim$ | $\mathrm{Dir}(\boldsymbol{\alpha})$ | $(\sum \theta_j = 1)$ |
| posterior: | $\boldsymbol{\theta} \mid \mathbf{k}$ | $\sim$ | $\mathrm{Dir}(\boldsymbol{\alpha} + \mathbf{k})$ | |

$$\mathbb{E}[\theta_j] = \frac{\alpha_j}{\sum \alpha_i} \qquad \mathbb{V}(\theta_j) = \frac{\mathbb{E}[\theta_j](1 - \mathbb{E}[\theta_j])}{\sum \alpha_i + 1}$$

# Example: Dirichlet-Multinomial-Model

| | | | | |
|---|---|---|---|---|
| Data: | $\mathbf{k}$ | $\sim$ | $\mathrm{M}(\boldsymbol{\theta})$ | $(\sum k_j = n)$ |
| conjugate prior: | $\boldsymbol{\theta}$ | $\sim$ | $\mathrm{Dir}(\boldsymbol{\alpha})$ | $(\sum \theta_j = 1)$ |
| posterior: | $\boldsymbol{\theta} \mid \mathbf{k}$ | $\sim$ | $\mathrm{Dir}(\boldsymbol{\alpha} + \mathbf{k})$ | |

$$\mathbb{E}[\theta_j] = \frac{\alpha_j}{\sum \alpha_i} =: y_j^{(0)} \quad \mathbb{V}(\theta_j) = \frac{\mathbb{E}[\theta_j](1 - \mathbb{E}[\theta_j])}{\sum \alpha_i + 1} = \frac{y_j^{(0)}(1 - y_j^{(0)})}{n^{(0)} + 1}$$

| | | | | |
|---|---|---|---|---|
| Data : | $\mathbf{k}$ | $\sim$ | $\mathrm{M}(\boldsymbol{\theta})$ | |
| conjugate prior: | $\boldsymbol{\theta}$ | $\sim$ | $\mathrm{Dir}(n^{(0)}, \mathbf{y}^{(0)})$ | $n^{(0)} = \sum \alpha_i$ |
| posterior: | $\boldsymbol{\theta} \mid \mathbf{k}$ | $\sim$ | $\mathrm{Dir}(n^{(1)}, \mathbf{y}^{(1)})$ | |

# Example: Dirichlet-Multinomial-Model

$$\begin{array}{r|ccl} \text{Data:} & \mathbf{k} & \sim & \mathrm{M}(\boldsymbol{\theta}) \\ \text{conjugate prior:} & \boldsymbol{\theta} & \sim & \mathrm{Dir}(\boldsymbol{\alpha}) \\ \hline \text{posterior:} & \boldsymbol{\theta} \mid \mathbf{k} & \sim & \mathrm{Dir}(\boldsymbol{\alpha} + \mathbf{k}) \end{array} \qquad \begin{array}{l} (\sum k_j = n) \\ (\sum \theta_j = 1) \\ \\ \end{array}$$

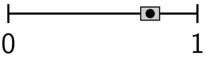
$$\mathbb{E}[\theta_j] = \frac{\alpha_j}{\sum \alpha_i} =: y_j^{(0)} \qquad \mathbb{V}(\theta_j) = \frac{\mathbb{E}[\theta_j](1 - \mathbb{E}[\theta_j])}{\sum \alpha_i + 1} = \frac{y_j^{(0)}(1 - y_j^{(0)})}{n^{(0)} + 1}$$

$$\begin{array}{r|ccl} \text{Data :} & \mathbf{k} & \sim & \mathrm{M}(\boldsymbol{\theta}) \\ \text{conjugate prior:} & \boldsymbol{\theta} & \sim & \mathrm{Dir}(n^{(0)}, \mathbf{y}^{(0)}) \\ \hline \text{posterior:} & \boldsymbol{\theta} \mid \mathbf{k} & \sim & \mathrm{Dir}(n^{(1)}, \mathbf{y}^{(1)}) \end{array} \qquad n^{(0)} = \sum \alpha_i$$

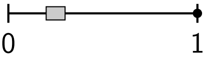
$$y_j^{(1)} = \frac{n^{(0)}}{n^{(0)} + n} \cdot y_j^{(0)} + \frac{n}{n^{(0)} + n} \cdot \frac{k_j}{n} \qquad n^{(1)} = n^{(0)} + n$$

# Example: Dirichlet-Multinomial-Model / IDM

Case (i): $y_j^{(0)} \in [0.7, 0.8]$, $(n^{(0)} = 8)$ $k_j/n = 0.75$ $(n = 16)$

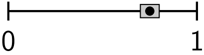



Case (ii): $y_j^{(0)} \in [0.2, 0.3]$, $(n^{(0)} = 8)$ $k_j/n = 1$ $(n = 16)$

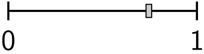


# Example: Dirichlet-Multinomial-Model / IDM

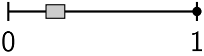
Case (i): $y_j^{(0)} \in [0.7, 0.8]\,,$ $({\color{blue}n^{(0)}} = 8)$ $\quad k_j/n = 0.75$ $(n = 16)$




→ $y_j^{(1)} \in [0.73, 0.76]$ $({\color{blue}n^{(0)}} = 24)$

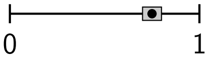



Case (ii): $y_j^{(0)} \in [0.2, 0.3]\,,$ $({\color{blue}n^{(0)}} = 8)$ $\quad k_j/n = 1$ $(n = 16)$

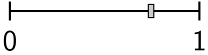


# Example: Dirichlet-Multinomial-Model / IDM

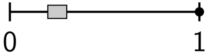
Case (i): $y_j^{(0)} \in [0.7, 0.8]\,,$ $(n^{(0)} = 8)$ $\quad k_j/n = 0.75$ $(n = 16)$



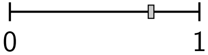

→ $y_j^{(1)} \in [0.73, 0.76]$ $(n^{(0)} = 24)$




Case (ii): $y_j^{(0)} \in [0.2, 0.3]\,,$ $(n^{(0)} = 8)$ $\quad k_j/n = 1$ $(n = 16)$

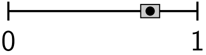



→ $y_j^{(1)} \in [0.73, 0.76]$ $(n^{(0)} = 24)$

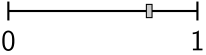


# Example: Dirichlet-Multinomial-Model / IDM

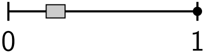
Case (i): $y_j^{(0)} \in [0.7, 0.8]$, $(n^{(0)} = 8)$   $k_j/n = 0.75$ $(n = 16)$



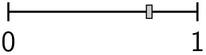

→ $y_j^{(1)} \in [0.73, 0.76]$ $(n^{(0)} = 24)$




Case (ii): $y_j^{(0)} \in [0.2, 0.3]$, $(n^{(0)} = 8)$   $k_j/n = 1$ $(n = 16)$




→ $y_j^{(1)} \in [0.73, 0.76]$ $(n^{(0)} = 24)$




[ $\mathbb{V}(\theta_j) \in [0.0178, 0.0233]$ → $\mathbb{V}(\theta_j|\mathbf{k}) \in [0.0072, 0.0078]$ ]

⚠ Posterior inferences do not reflect uncertainty due to unexpected observations! ⚠

# Some Thoughts on Prior-Data Conflict

- genuinely Bayesian, but today not limited to: we consider also *data-data conflict*

# Some Thoughts on Prior-Data Conflict

- genuinely Bayesian, but today not limited to:
  we consider also *data-data conflict*
- precise models:
  - most conjugate models tend to ignore prior-data conflict
  - some do react, but often not satisfactorily
  - diagnosis schemes: presence, and when safely ignorable
  - if present, no general strategy (if and) how to do inference

# Some Thoughts on Prior-Data Conflict

- genuinely Bayesian, but today not limited to: we consider also *data-data conflict*
- precise models:
  - most conjugate models tend to ignore prior-data conflict
  - some do react, but often not satisfactorily
  - diagnosis schemes: presence, and when safely ignorable
  - if present, no general strategy (if and) how to do inference
- imprecise probability models:
  - some share same drawbacks as precise models
  - opportunity to encode *precision* of probability statements
  - relation of (amount/quality of) information and imprecision

# Some Thoughts on Prior-Data Conflict

- genuinely Bayesian, but today not limited to:
  we consider also *data-data conflict*
- precise models:
  - most conjugate models tend to ignore prior-data conflict
  - some do react, but often not satisfactorily
  - diagnosis schemes: presence, and when safely ignorable
  - if present, no general strategy (if and) how to do inference
- imprecise probability models:
  - some share same drawbacks as precise models
  - opportunity to encode *precision* of probability statements
  - relation of (amount/quality of) information and imprecision
- relations to *learning* in general?
- adjusting background information in the light of unexpected observations (Frank Hampel, 2007)

# Bernoulli Data

## Bernoulli Data Scenario

Prior model expects quite certainly 5 successes out of 20,
but data says 18 successes out of 20.

# Bernoulli Data

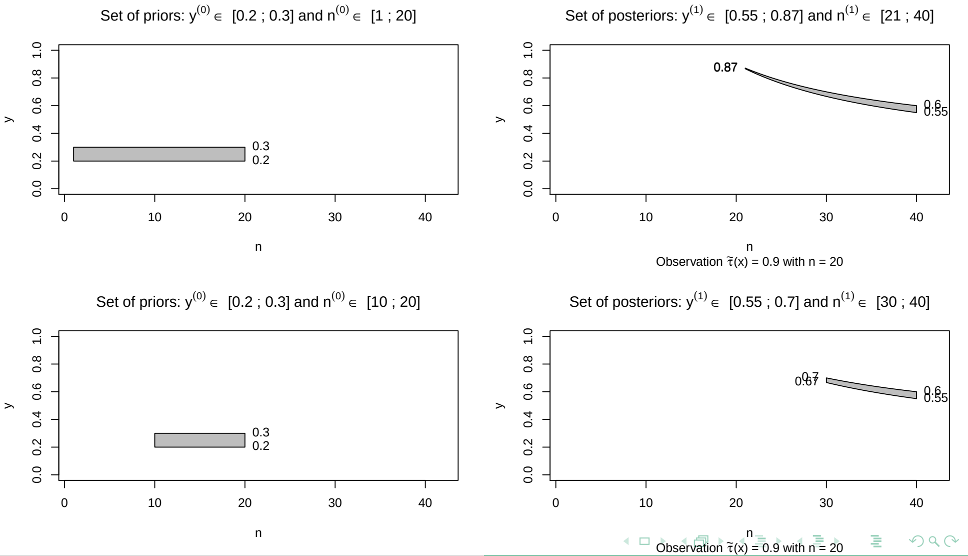
## Bernoulli Data Scenario

Prior model expects quite certainly 5 successes out of 20,
but data says 18 successes out of 20.

➡ generalized iLUCK-model: $y^{(0)} \approx \frac{5}{20} = \frac{1}{4}$, $n^{(0)} \leq 20$

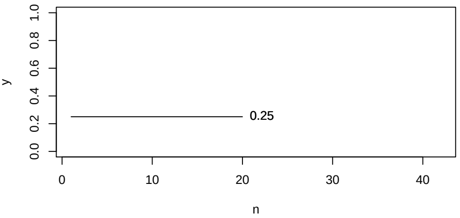
- $y^{(0)} \in [0.2, 0.3]$, $n^{(0)} \in [1, 20]$
- $y^{(0)} \in [0.2, 0.3]$, $n^{(0)} \in [10, 20]$
- $y^{(0)} = 0.25$, $n^{(0)} \in [1, 20]$
- $y^{(0)} = 0.25$, $n^{(0)} \in [10, 20]$

# Bernoulli Data

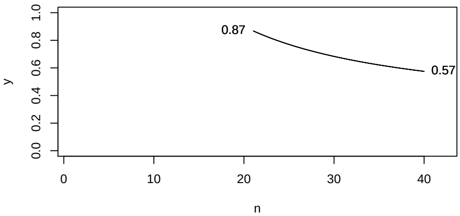


# Bernoulli Data

Set of priors: $y^{(0)} = 0.25$ and $n^{(0)} \in [1 ; 20]$

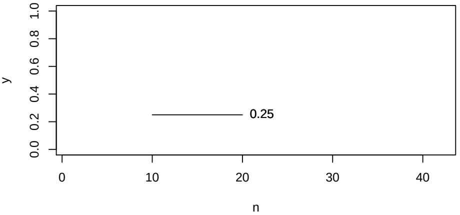



Set of posteriors: $y^{(1)} \in [0.57 ; 0.87]$ and $n^{(1)} \in [21 ; 40]$

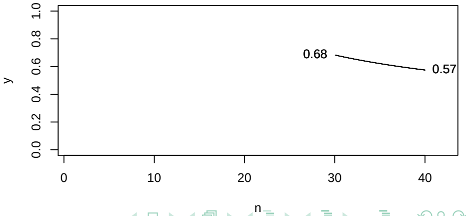



Observation $\tilde{\tau}(x) = 0.9$ with $n = 20$

Set of priors: $y^{(0)} = 0.25$ and $n^{(0)} \in [10 ; 20]$




Set of posteriors: $y^{(1)} \in [0.57 ; 0.68]$ and $n^{(1)} \in [30 ; 40]$



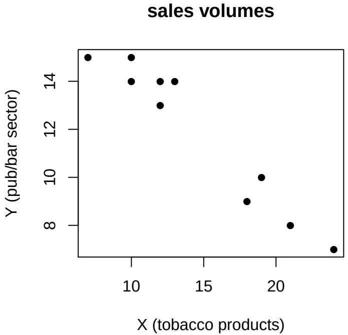

Observation $\tilde{\tau}(x) = 0.9$ with $n = 20$

# Regression/Correlation Data

Before the introduction of general smoking bans for public areas, it was thought that an introduction of such a law would seriously harm pubs and bars.

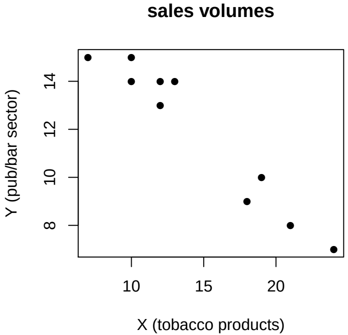
# Regression/Correlation Data

Before the introduction of general smoking bans for public areas, it was thought that an introduction of such a law would seriously harm pubs and bars.



# Regression/Correlation Data

Before the introduction of general smoking bans for public areas, it was thought that an introduction of such a law would seriously harm pubs and bars.



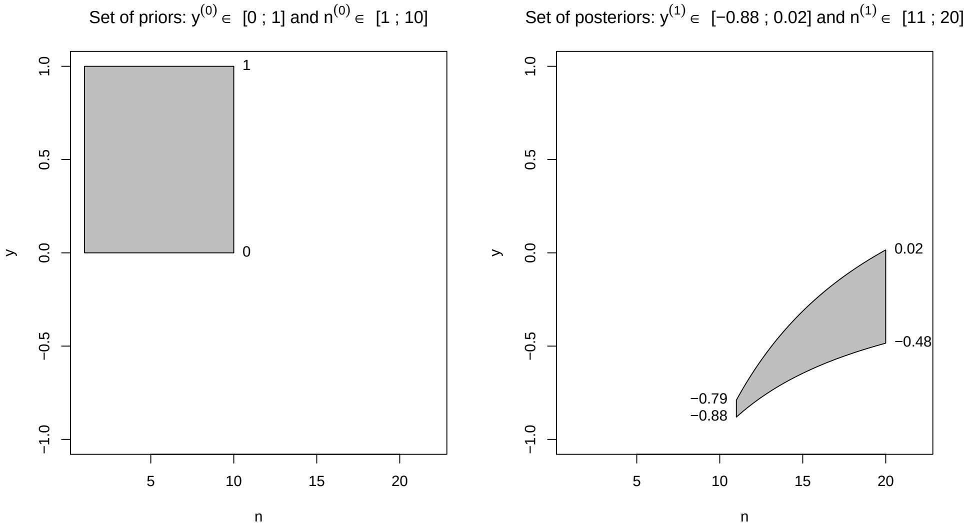

## Regression/Correlation Data Scenario

Prior model expects quite certainly a positive slope,
data suggests instead a negative slope.

# Regression/Correlation Data

- standardize $x$ and $y$: $\tilde{x}$, $\tilde{y}$
- no intercept necessary, $\beta_1 = \rho(x, y)$
- posterior expectation of $\beta_1$ is weighted average of prior expectation and LS estimate $(= -0.9687)$

# Regression/Correlation Data




Observation $\tilde{\tau}(x) = -0.97$ with $n = 10$

# "Strong Happiness"

Data situation: Bernoulli sampling (observe 0 or 1).

Idea: Choose sample size $n_1$ such that $I^{(1)}$ low enough under certain threshold $I$ such that – even if we got another sample $n_2$ in conflict to $\mathcal{Y}^{(1)}$ – $I^{(2)}$ is still below $I$!

With a sample of such a size $n_1$ we attain "strong happiness", because whatever we would see as a following sample, we would never get more imprecise than $I$!

(Is only possible because degree of prior-data conflict is bounded due to Bernoulli sampling!).

# "Strong Happiness"

$$I^{(2)} = \frac{\overline{n}^{(0)} I^{(0)}}{\overline{n}^{(0)} + n_1 + n_2} + \frac{\overline{n}^{(0)} - \underline{n}^{(0)}}{\overline{n}^{(0)} + n_1 + n_2} \left( \frac{n_1}{\underline{n}^{(0)} + n_1} \Delta\left(\frac{c_1}{n_1}, \mathcal{Y}^{(0)}\right) + \frac{n_2}{\underline{n}^{(0)} + n_1 + n_2} \Delta\left(\frac{c_2}{n_2}, \mathcal{Y}^{(1)}\right) \right)$$

$$\overset{!}{\leq} I \quad \forall\, (k_2,\ n_2)$$

# "Strong Happiness"

$$I^{(2)} = \frac{\overline{n}^{(0)} I^{(0)}}{\overline{n}^{(0)} + n_1 + n_2} + \frac{\overline{n}^{(0)} - \underline{n}^{(0)}}{\overline{n}^{(0)} + n_1 + n_2} \Bigg( \frac{n_1}{\underline{n}^{(0)} + n_1} \Delta\Big(\frac{c_1}{n_1}, \mathcal{Y}^{(0)}\Big) + \frac{n_2}{\underline{n}^{(0)} + n_1 + n_2} \Delta\Big(\frac{c_2}{n_2}, \mathcal{Y}^{(1)}\Big) \Bigg)$$

$$\overset{!}{\leq} I \quad \forall\, (k_2,\, n_2)$$

If data from second sample in conflict with $\mathcal{Y}^{(1)}$ (and conflict strong enough), then imprecision $I^{(2)}$ should first increase and then decrease in $n_2$.

→ find $n_2$ that maximizes $I^{(2)}$ for maximal possible conflict with current $\mathcal{Y}^{(1)}$, plug into formula for $I^{(2)}$ and give $n_1$ as a function of $I^{(0)}$ (or $\mathcal{Y}^{(0)}$) and $I$ (and $\mathcal{N}^{(0)}$).