Towards Robust Inference in Graphical Models

Corné van Daalen

Electronic Systems Laboratory (ESL)
Department of Electrical and Electronic Engineering

http://staff.ee.sun.ac.za/cvdaalen/
cvdaalen@sun.ac.za

17 May 2019

Beforehand

- This is ongoing work
  - Speculative statements: *
  - Unsubstantiated statements: *
- Informal development
- Toy problems
- Feedback and collaboration welcome

Outline

Outline

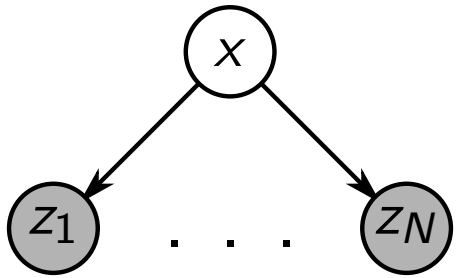
Inference in Graphical Models

- 2 components: model and evidence
- Modelling:
  - Define random variables and statistical relationships between them
  - Directed graph (Bayes network) or indirected graph (Markov network)
  - General framework
- Evidence
  - Typically measurements
- Inference: model + evidence $\rightarrow$ answers to questions
  - 2 main flavours: variational and sampling-based
  - We focus on message passing (variational)

Regression Example

Model:

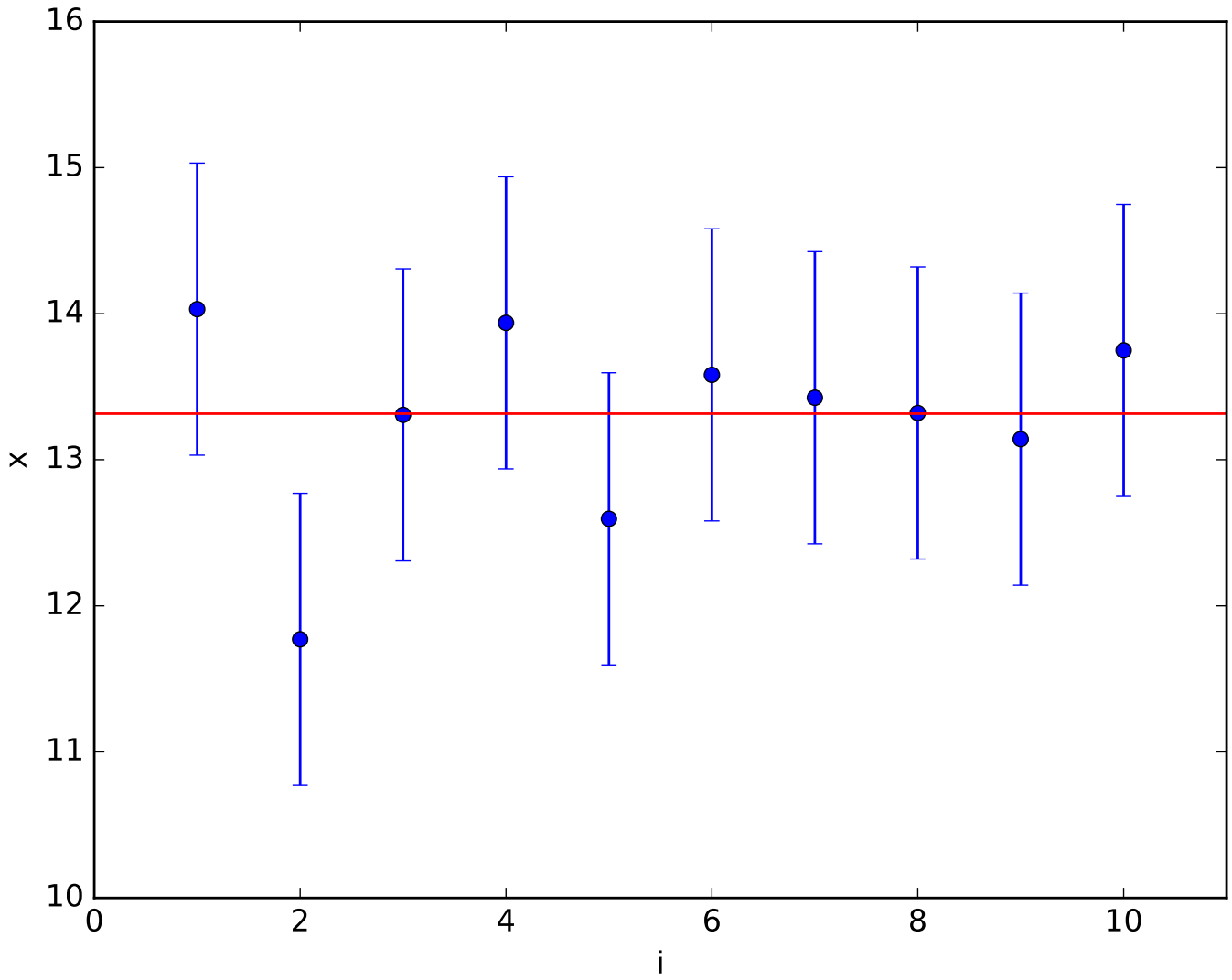
$$p(x) = \mathcal{N}\left(\mu_x, \sigma_x^2\right)$$
$$z_i = x + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}\left(0, \sigma_z^2\right), \quad i = 1, \ldots, N$$



Question:

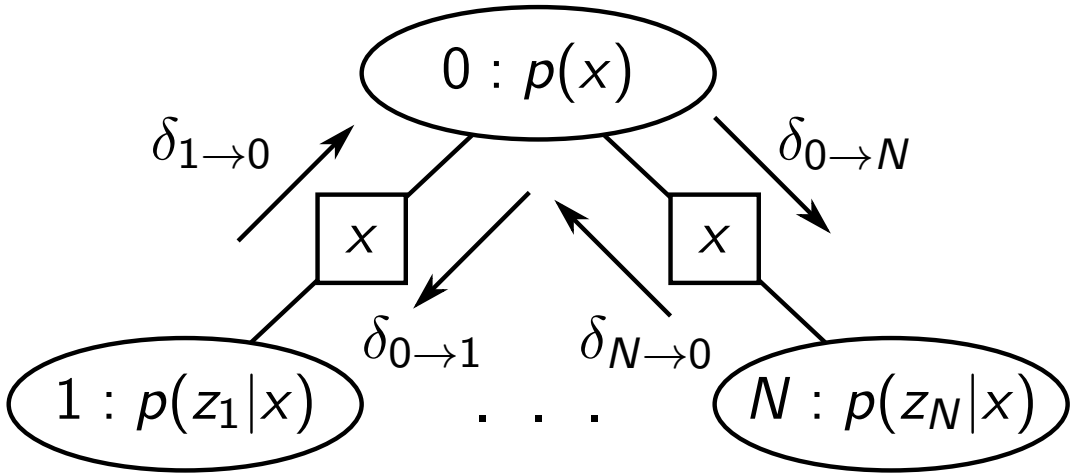
$$p\left(x | z_1, \ldots, z_N\right)?$$

Regression Example

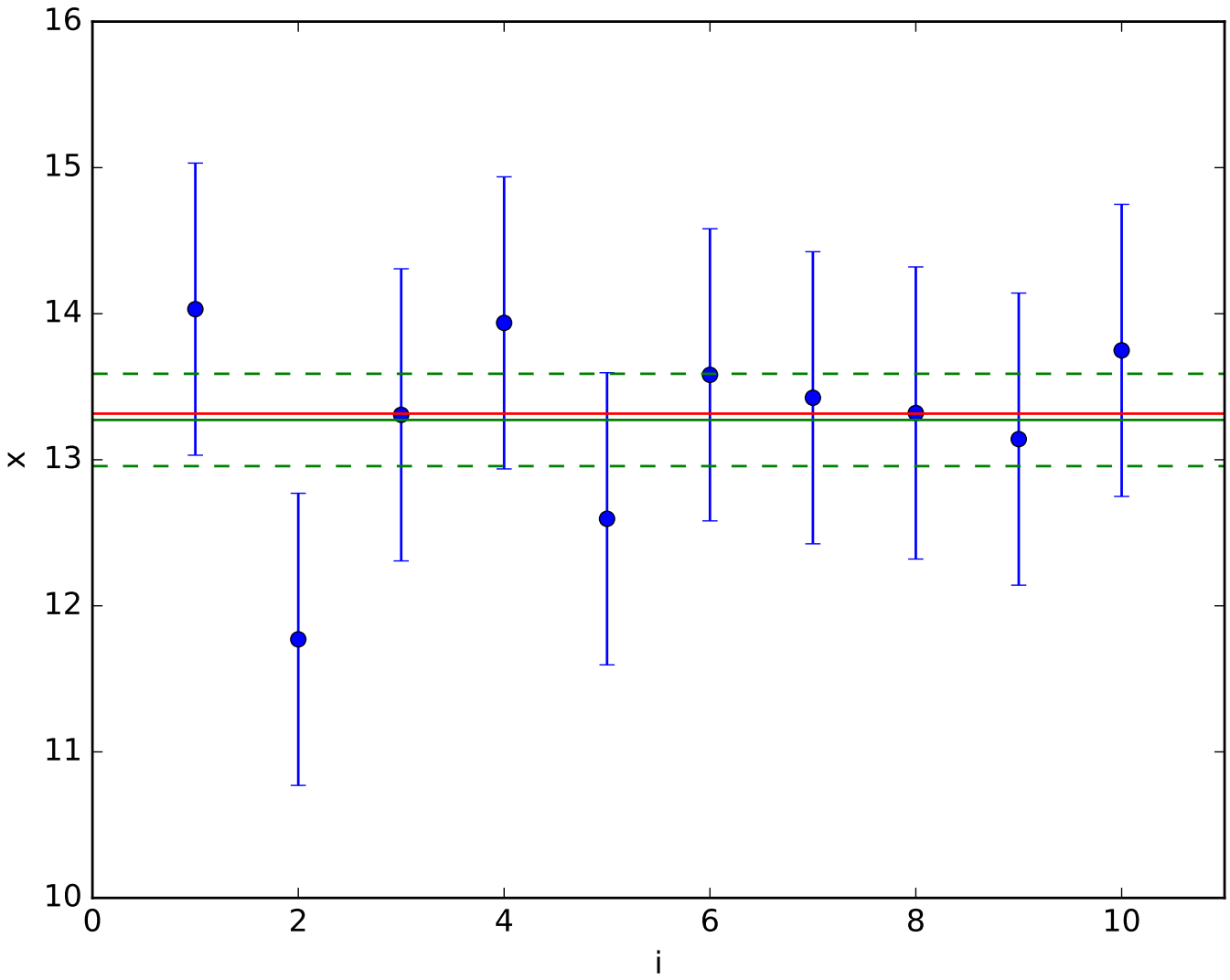


Regression Example

Inference:



Regression Example

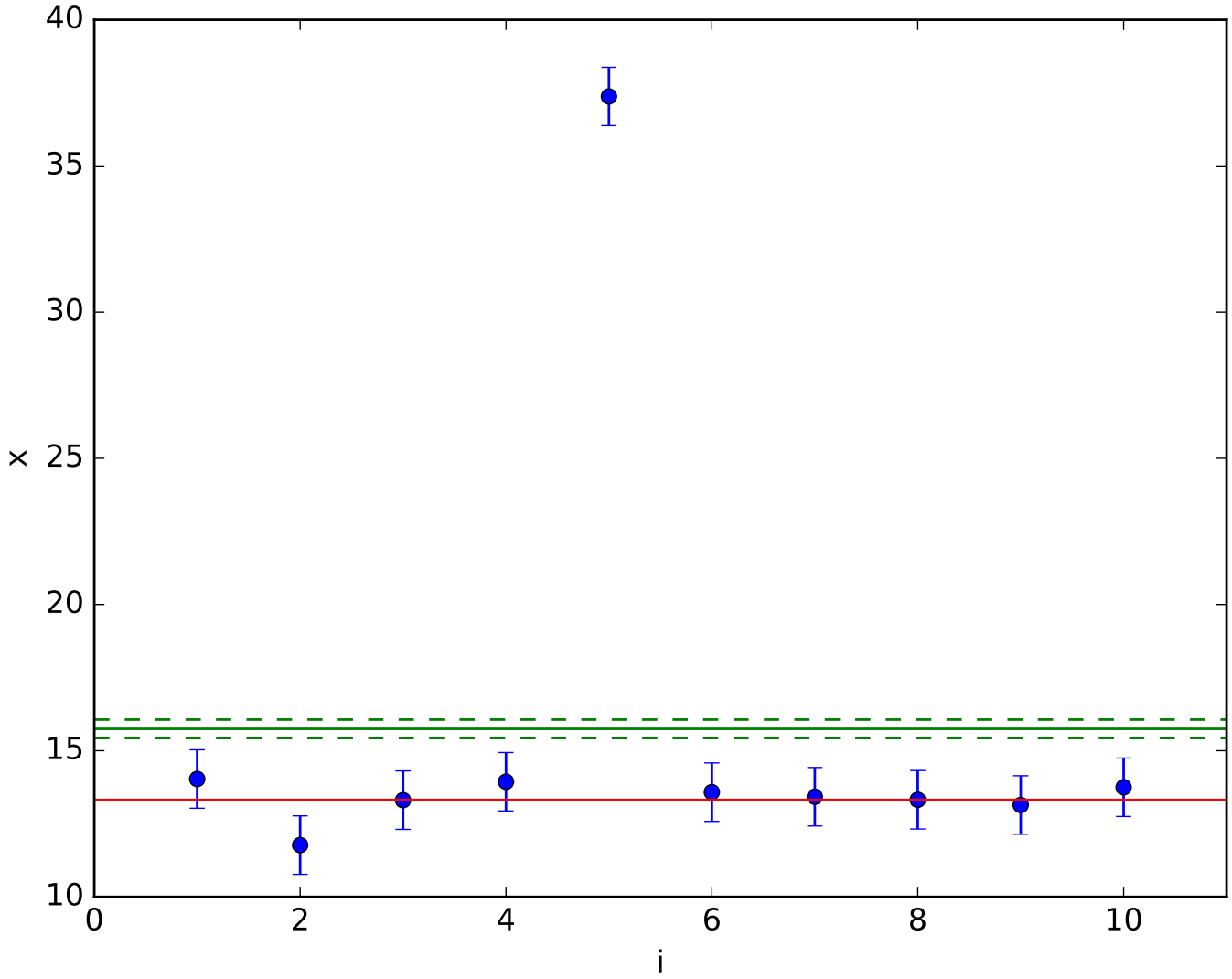


Outline

So what is the problem?

- Inference never produces correct answers
  - Models are never correct*
  - Tractable inference require approximations
- Result:
  - Rubbish answers
  - Things break*
- Existing approaches:
  - Full Bayesian approach
  - Model selection
- Disadvantages of existing approaches:
  - Require complex models
  - Expensive
  - They do not solve problem*

Regression Example



Outline

Quantifying Model Inconsistency

- Bayesian surprise:

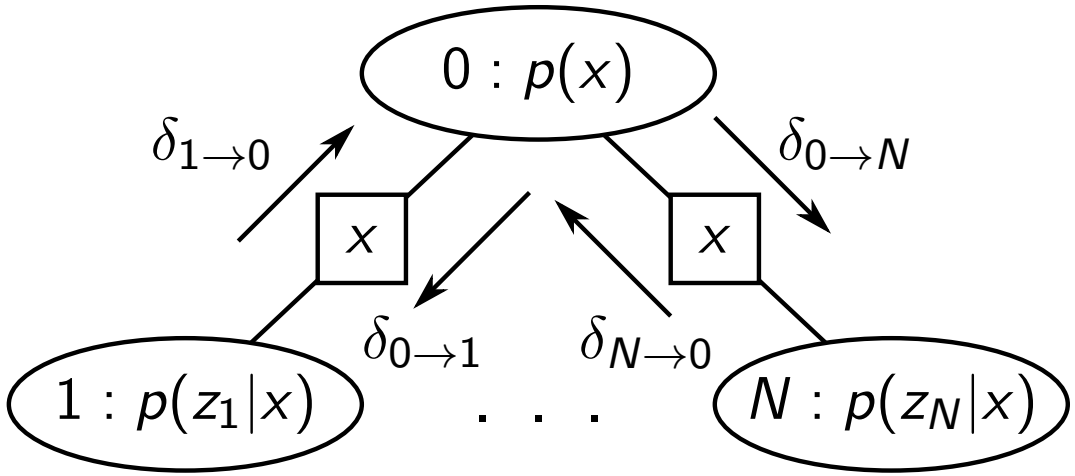
$$S(E, p(x)) \triangleq D_{KL}(p(x|E)||p(x))$$

  - *How much the evidence changed the prior belief*
- Properties:
  - Number between 0 and ∞
  - Lower measurement likelihood $\rightarrow$ higher surprise
  - Defined for most parametric distributions
  - For Gaussian measurement – scaled Mahalanobis distance squared

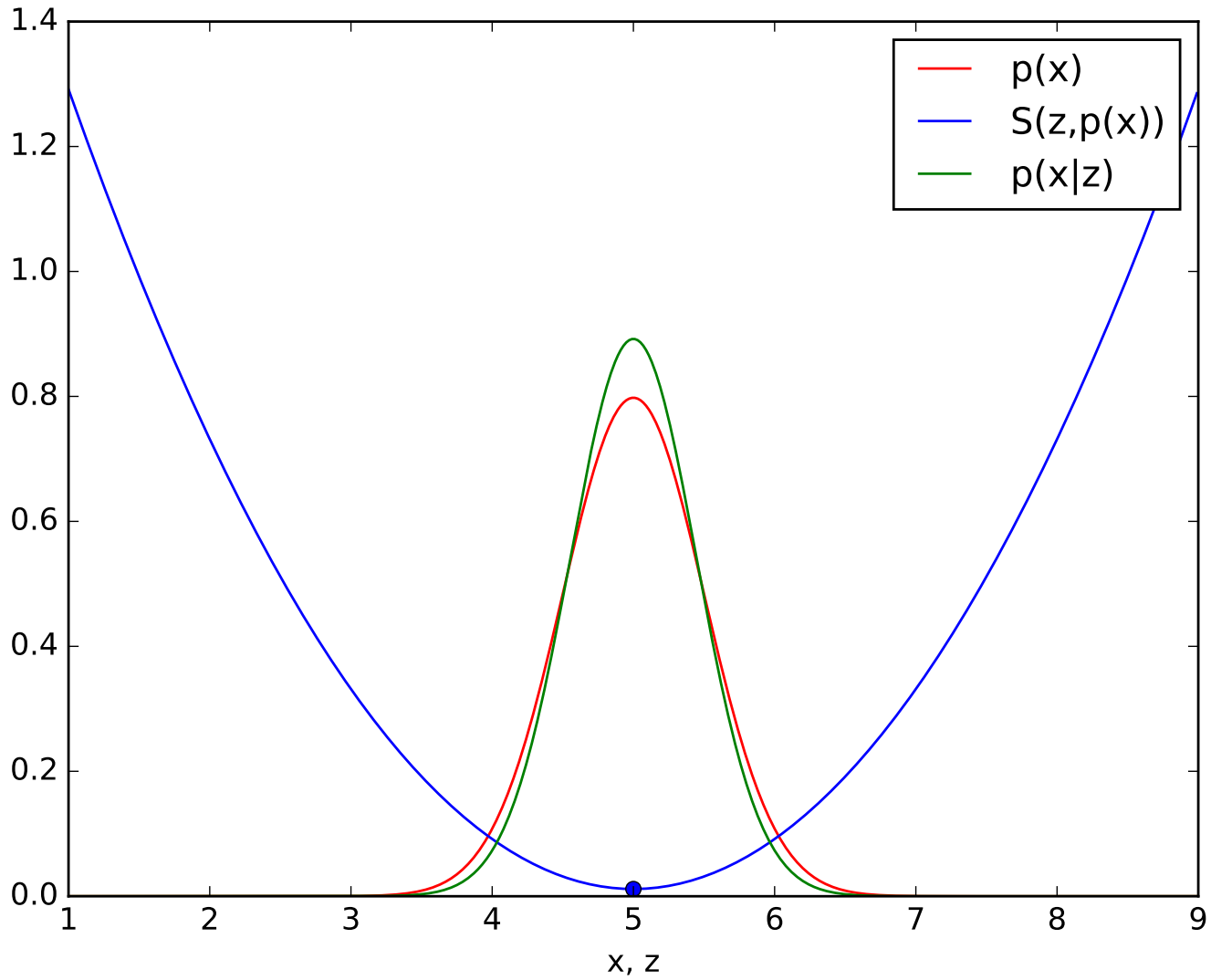
Quantifying Model Inconsistency

- For regression example:

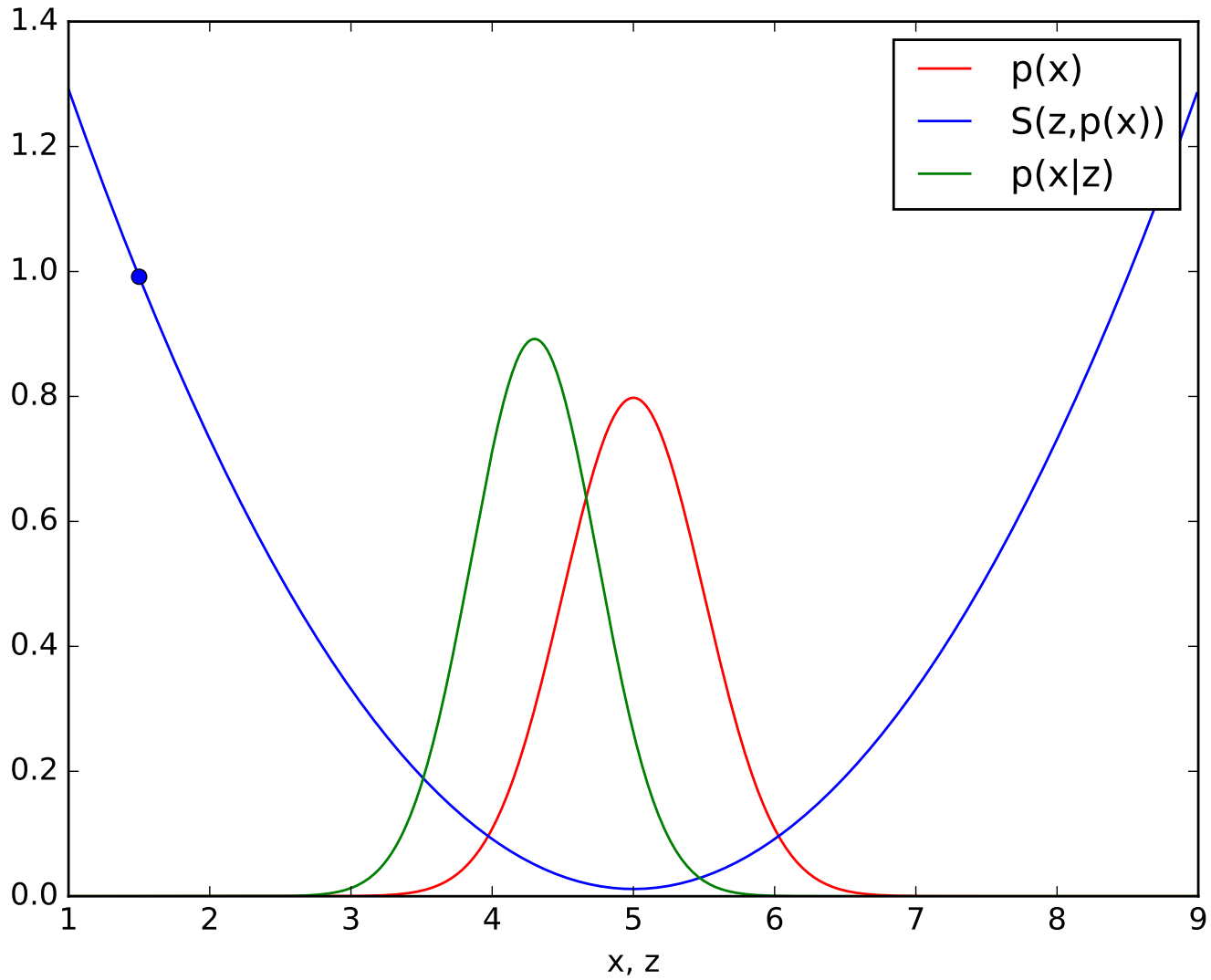
$$S_1 \triangleq S\left(z_1, \delta_{0\to i}(x)\right) = D_{KL}\left(\beta(x)||\delta_{0\to 1}(x)\right)$$



Quantifying Model Inconsistency



Quantifying Model Inconsistency



Quantifying Model Inconsistency

- Conflict between whole model and evidence:

$$C = \sum_i S_i$$

- What level of conflict is acceptable?
  - Calculate $\mathcal{E}[C]$ and $var[C]$ using model
  - Assume Gaussian distribution
  - Calculate threshold C_{th} such that $P[C \leq C_{th}] = x\%$

Reducing Model Inconsistency

- Factor relaxation
  - Introduce adjustable parameters $\boldsymbol{\theta}_i$
  - If $\boldsymbol{\theta}_i = \mathbf{0}$, factor unchanged
  - As $\boldsymbol{\theta}_i$ increase, surprise decreases
- Relaxation for regression example:

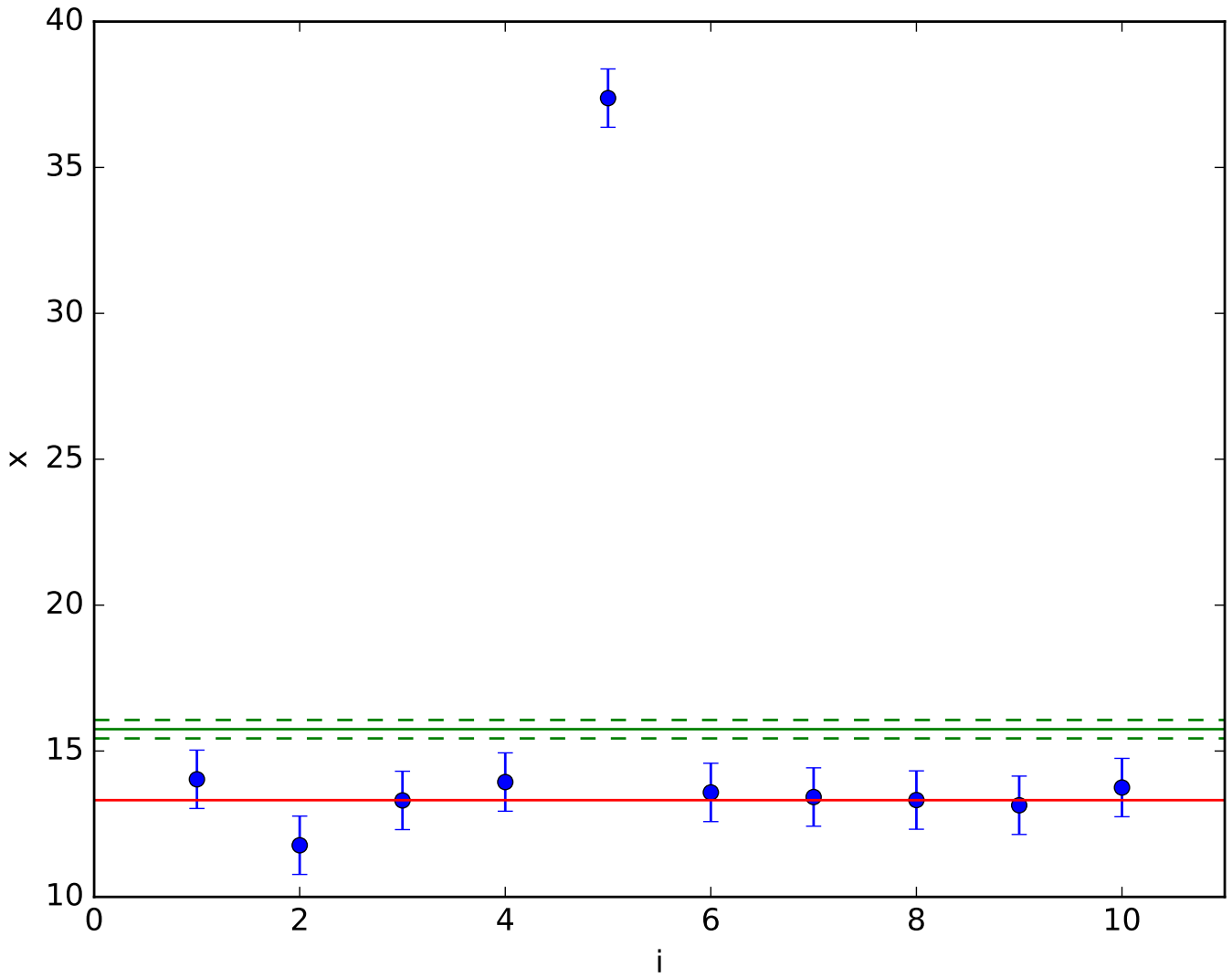
$$z_i = x + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}\left(0, (\sigma_z + \theta_i)^2\right)$$

- Relaxation cost: $J(\boldsymbol{\theta})$
  - E.g. $J(\boldsymbol{\theta}) = \|\boldsymbol{\theta}\|_1$

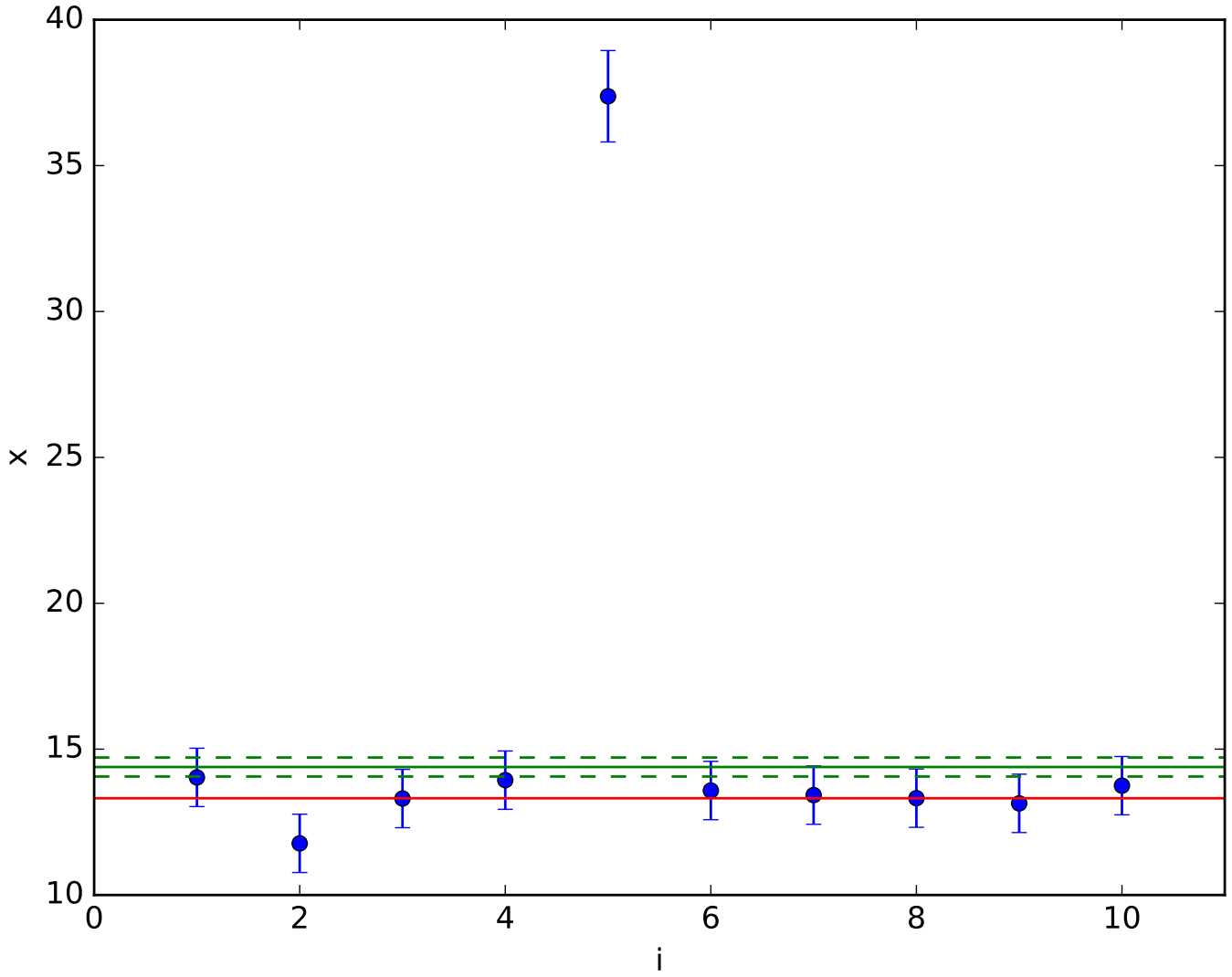
Reducing Model Inconsistency

- Problem formulation:
  - Minimise $J(\boldsymbol{\theta})$ s.t. $C \leq C_{th}$
  - Or: *find the cheapest consistent model*
- Minimisation process:
  - Calculate $\frac{\partial C}{\partial \boldsymbol{\theta}}$ using automatic differentiation
  - Use modified Newton's method

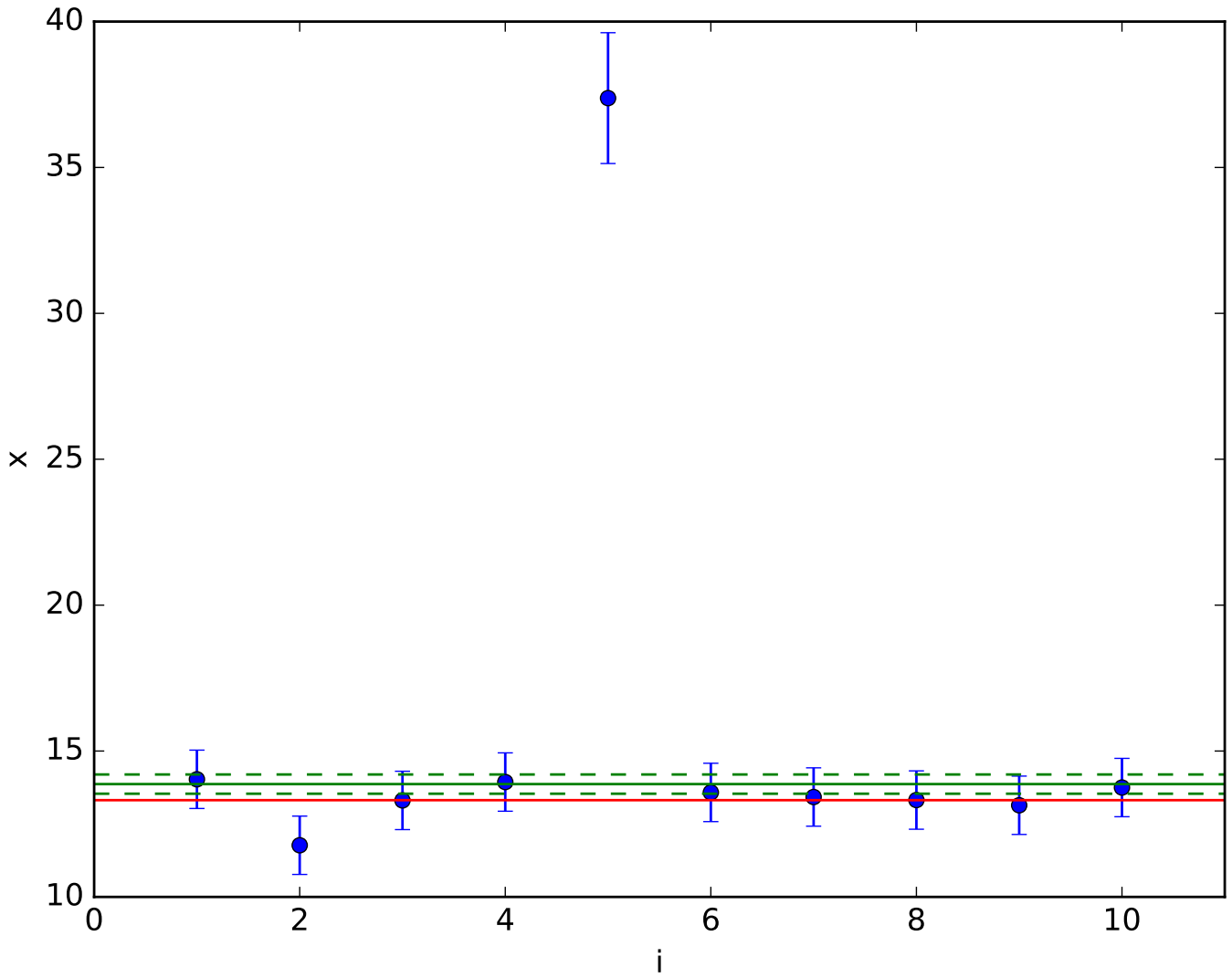
Regression Example



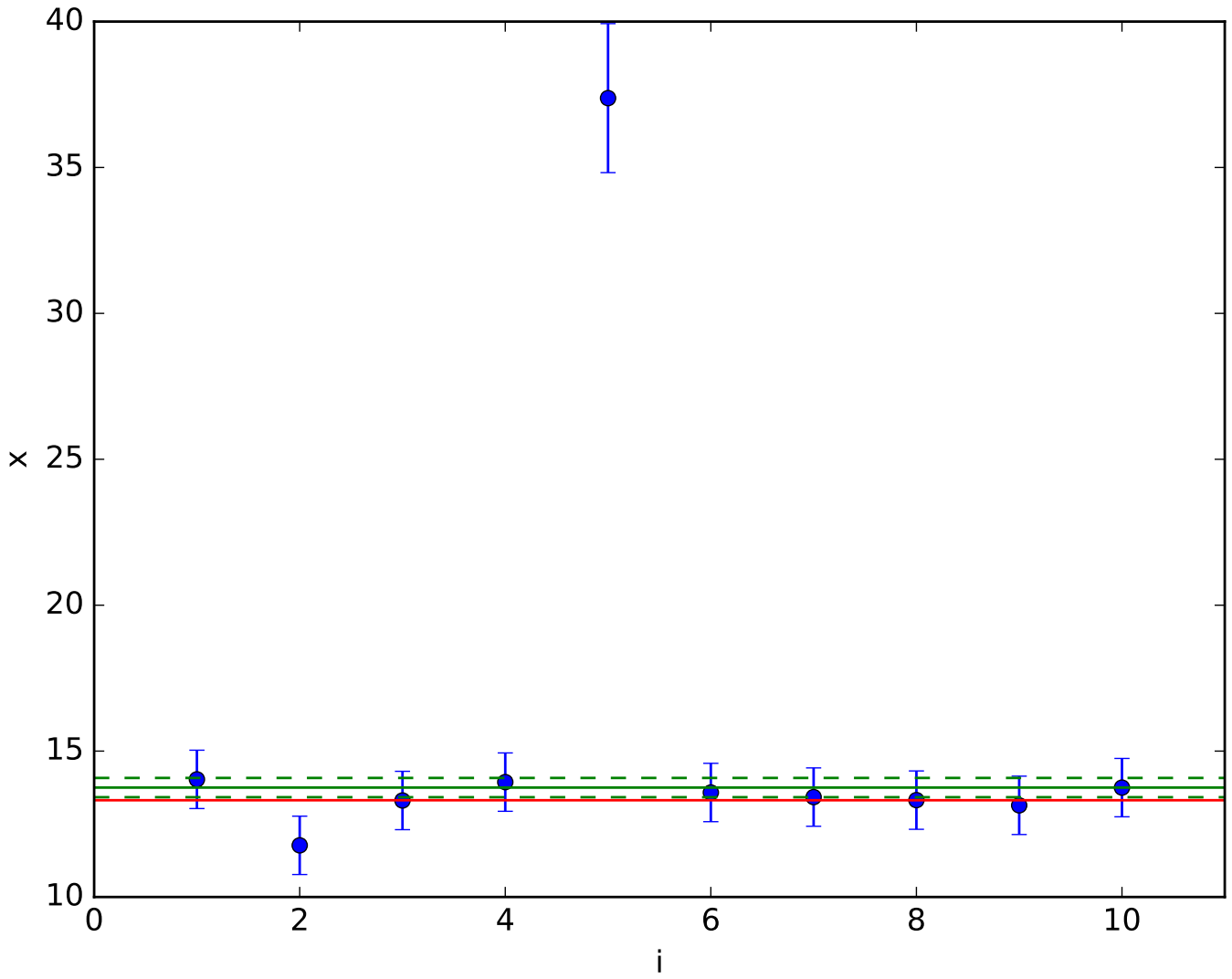
Regression Example



Regression Example



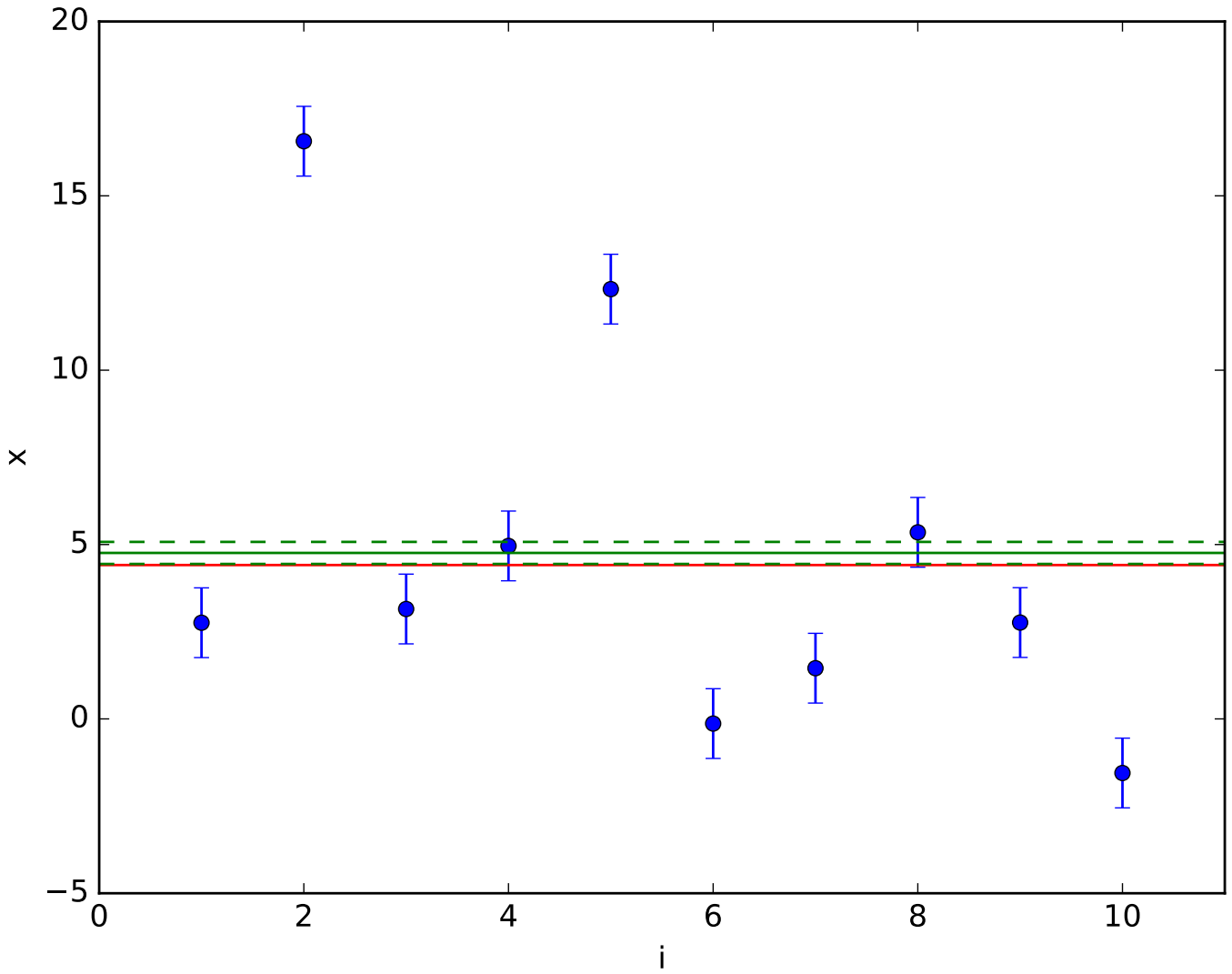
Regression Example



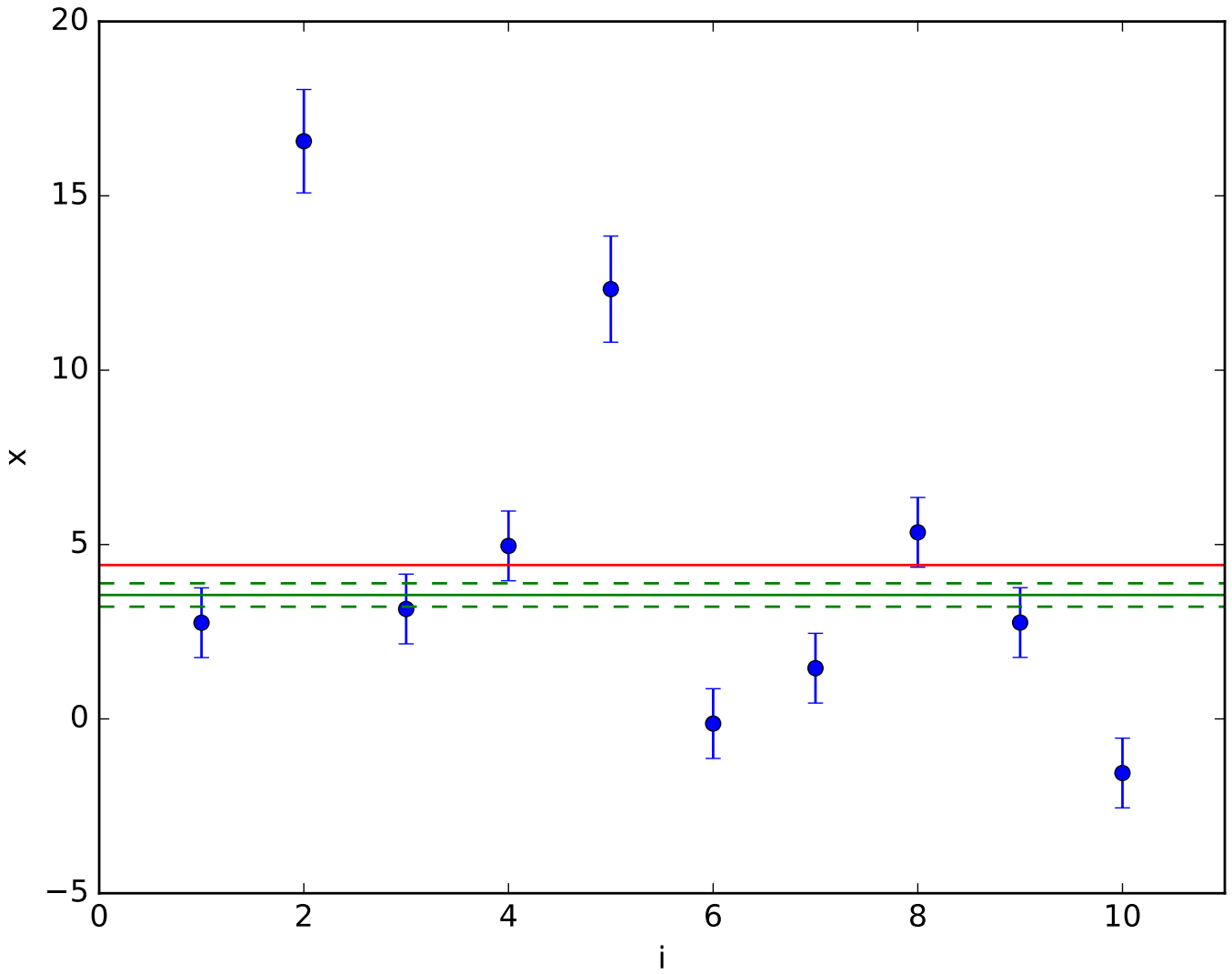
Regression Example

But why do we not just use outlier detection and removal?

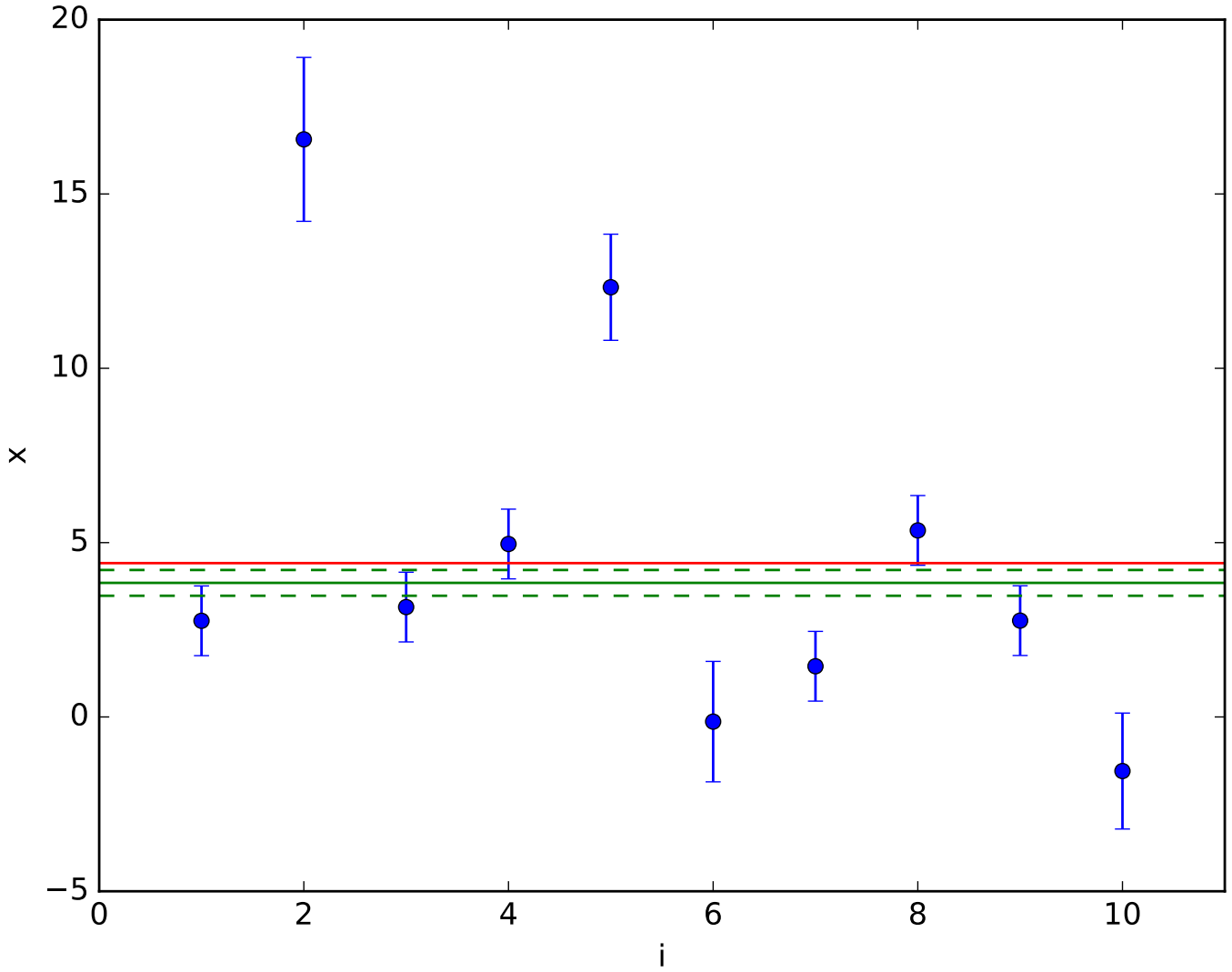
Regression Example



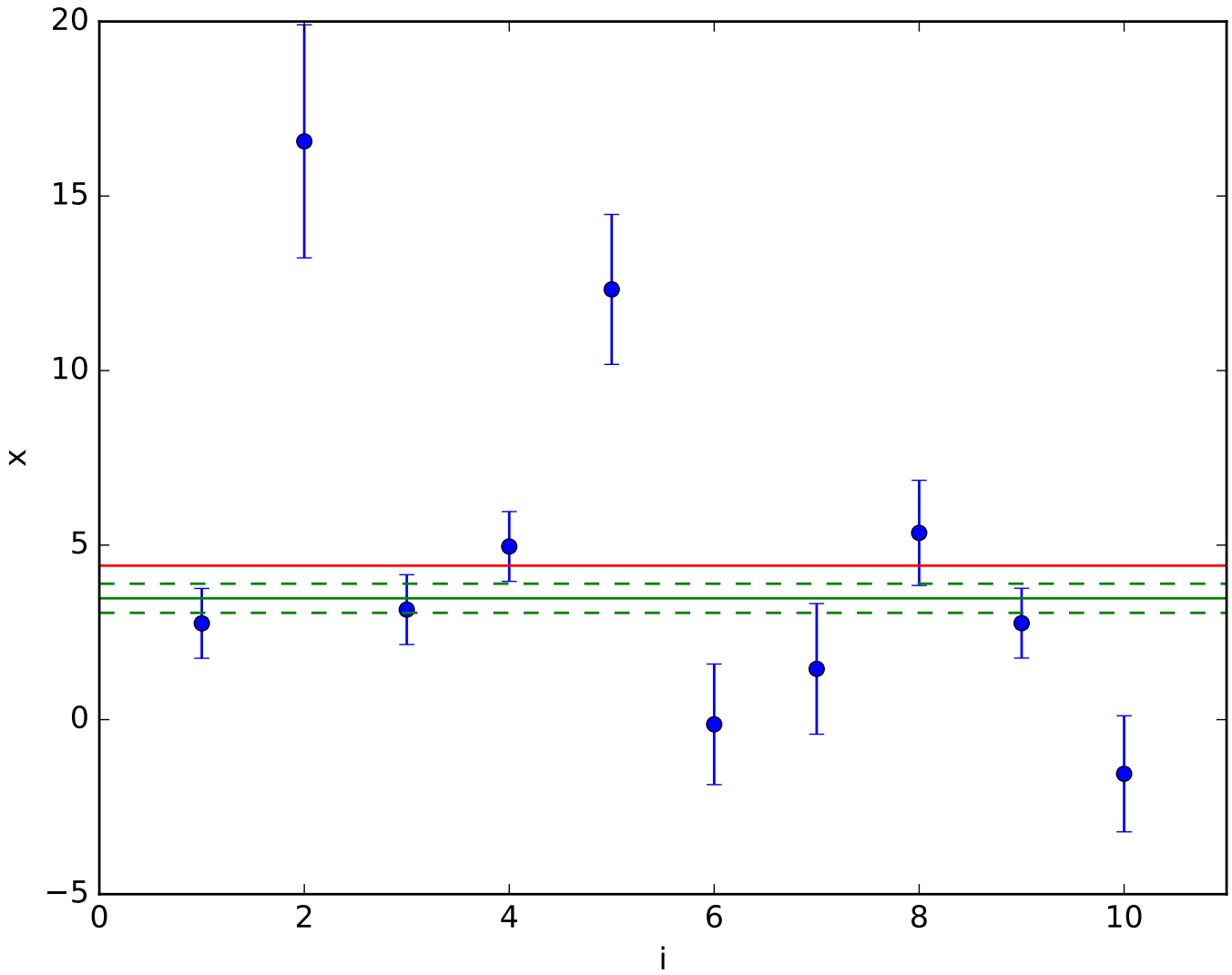
Regression Example



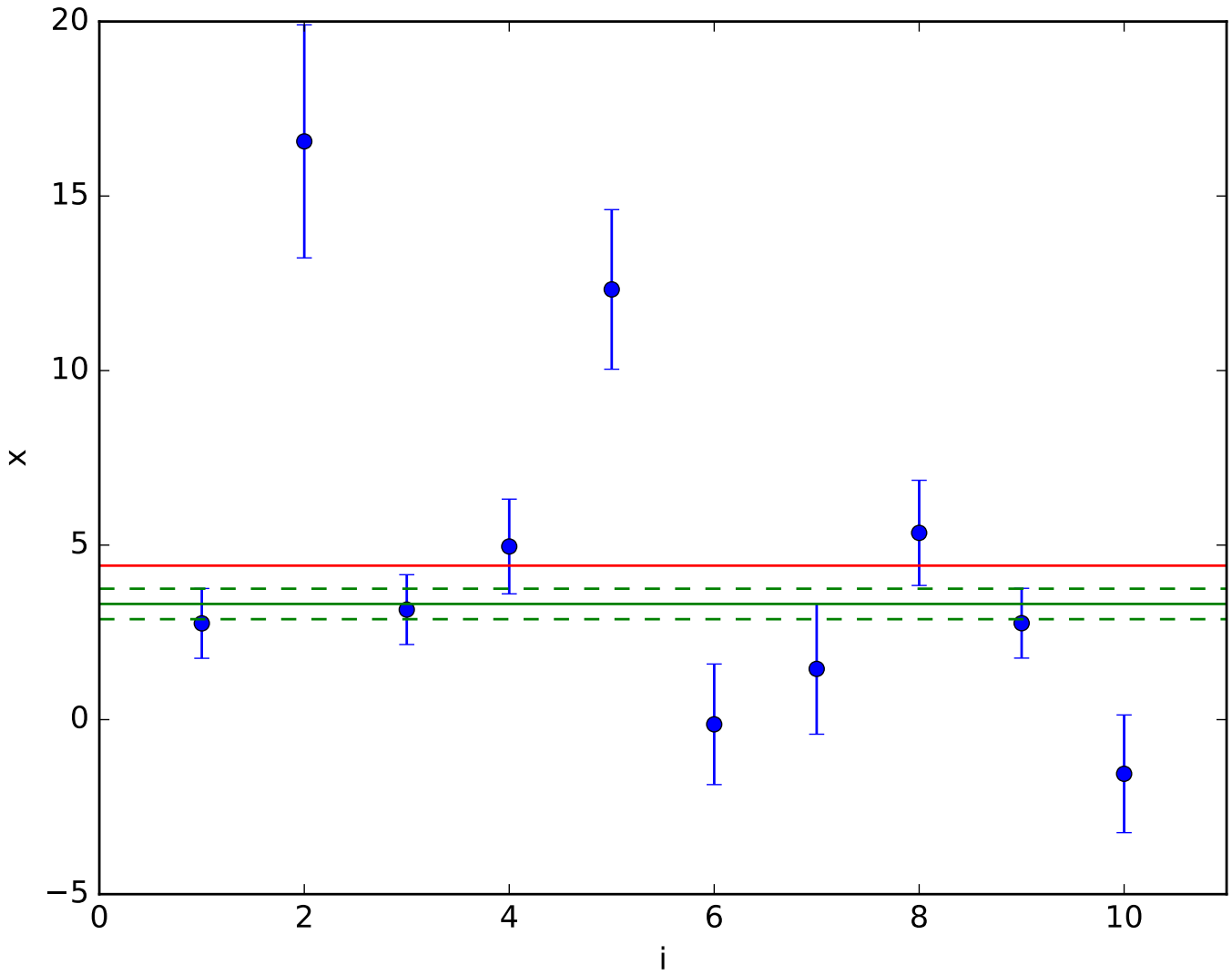
Regression Example



Regression Example



Regression Example



Outline

Conclusion

- Summary:
  - Inconsistency between model and evidence cause problems
  - We propose a method to relax a model during inference to reduce model inconsistency
  - Which does not rely on complex, accurate modelling
  - This produces better wrong answers*
- Characteristics of proposed method:
  - Fits into message passing for graphical models
  - Applicable to common continuous and discrete factors*
  - Can be extended to general factors*
  - Applicable to loopy networks*
- Further ideas and work:
  - Test for various, realistic problems
  - Extend to loopy networks
  - Default reasoning