The Maximum Entropy Principle

DI Danilo R. Neuber
neuber@itp.tu-graz.ac.at

Agenda

- Assigning Probabilities
  - Taking Prior Knowledge into Account
  - Entropy as a Measure of Uncertainty
  - The Principle of Maximum Entropy
- Data Analysis
  - Form-free Solutions
  - Tackling Ill-posed Problems
  - Example: Abel-Inversion

Assigning Probabilities

- "Principle of Insufficient Reason" (Bernoulli 1713)

  If given an enumeration of mutually exclusive possibilities X_j *and no further information*, then assign the same probability to all.

- Example: Ordinary die:

$$X_j \equiv \text{the face on top has } j \text{ dots}$$

$$\Rightarrow \quad p(X_j|I) = \frac{1}{6}$$

Testable Information

- Example: A die with known mean value:

$$\sum_{j=1}^{6} j\, p(X_j|I) = 4.5$$

- One possible solution:

X_j	1	2	3	4	5	6
$p(X_j)$	0	0	1/2	0	0	1/2

- How to find $p(X_j|I)$ with maximal "uncertainty"?

Measuring Uncertainty

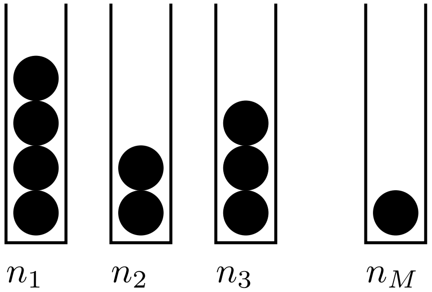
Entropy of a probability distribution (Shannon 1948)

$$S = -\sum_j p_j \log p_j$$

- Invariant with respect to permutations of indices
- Continuous with respect to p_j
- $S \geq 0$ and $S = 0$ only if $p_j = \delta_{ij}$
- Impossible events $p_j = 0$ do not contribute

The Monkey Argument I

- Monkeys throw N balls at M boxes (all boxes equally probable)
- Candidate pdf: $p_j = n_j/N$
- Reject $\{p_j\}$ unless satisfying contraints
- Repeat experiment many times
- Choose most frequent pdf $\{p_j\}$ as solution



The Monkey Argument II

- Frequency of $\{p_j\}$:

$$F(\{p_j\}) = \frac{N!}{n_1!n_2!\dots n_M!}\Big/ M^N$$

- Using $\log n! \approx n \log n - n$ we get:

$$\log F(\{p_j\}) = -N \sum_j p_j \log p_j - N \log M$$

- The most frequent pdf $\{p_j\}$ maximizes the entropy!

Principle of Maximum Entropy

- The least informative probability distribution maximizes the entropy S subject to known constraints. (Jaynes 1957)

- Example: Total ignorance

$$\mathcal{L} = -\sum_{j=1}^{N} p_j \log p_j + \lambda \left(\sum_{j=1}^{N} p_j - 1 \right)$$

$$\frac{\partial \mathcal{L}}{\partial p_j} = -\log p_j - 1 + \lambda \overset{!}{=} 0$$

$$p_j = e^{\lambda - 1} \overset{!}{=} \text{const}$$

Testable Information: Mean Value

- Maximize:

$$\mathcal{L} = -\sum_{j=1}^{N} p_j \log p_j + \lambda_1 \left(\sum_{j=1}^{N} j\, p_j - \mu \right) + \lambda_0 \left(\sum_{j=1}^{N} p_j - 1 \right)$$

- Solving for $\partial \mathcal{L} / \partial p_j = 0$ yields:

$$p_j = e^{\lambda_0 - 1} e^{\lambda_1 j} \equiv c \cdot q^j$$

Solution to the Die-Problem

- Expectation is known:

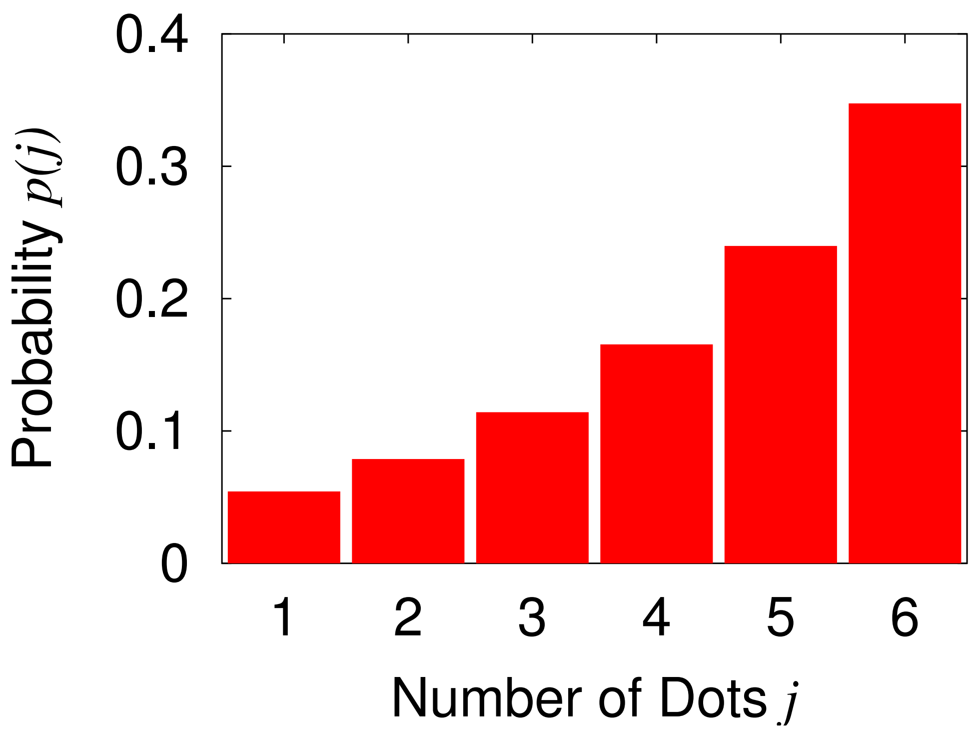
$$\sum_{j=1}^{6} j \cdot p_j = 4.5$$

- MaxEnt-Solution:

$$p_j = c \cdot q^j$$

$$c \approx 0.04$$

$$q \approx 1.45$$



Testable Information: Mean and Variance

- Maximize:

$$\mathcal{L} = -\sum_{j=1}^{N} p_j \log p_j + \lambda_1 \left(\sigma^2 - \sum_{j=1}^{N} (x_j - \mu)^2 p_j \right) + \lambda_0 \left(\sum_{j=1}^{N} p_j - 1 \right)$$

- Solution is a Gaussian distribution:

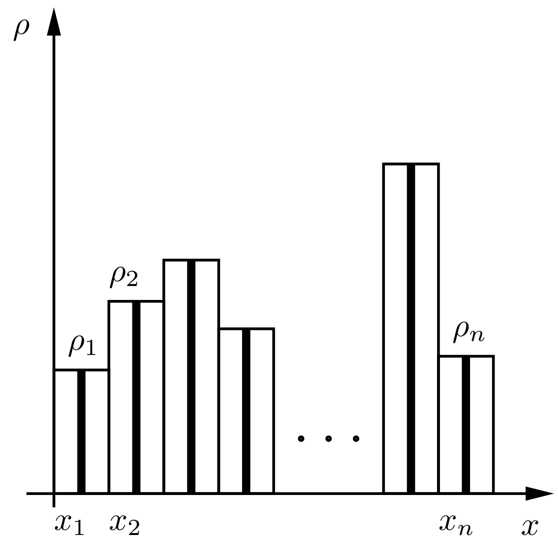
$$p(x_j|I) \sim \exp \lambda_1 (x_j - \mu)^2$$

Data Analysis

- Problem: Infer a density $\rho(x)$ from data D:
  - Positivity: $\rho(x) \geq 0$
  - No functional model with few parameters is known
- Form-free solution:
  - Discretization:

$$x \longrightarrow x_j$$

$$\rho(x) \longrightarrow \rho_j = \rho(x_j)\,\Delta x_j$$



Quantified MaxEnt I

- Quantization:

$$\rho_j \to n_j \approx \frac{\rho_j}{\Delta\rho}, \qquad n_j = 0, 1, \ldots$$

- Prior: Poission distribution for pixel n_j with expectation μ_j:

$$p(n|\mu, I) = \frac{\mu^n}{n!} e^{-n}$$

- Apply Stirling formula for $n!$ and substitute:

$$n_j = \rho_j/\Delta\rho, \qquad \mu_j = m_j/\Delta\rho, \qquad \alpha = 1/\Delta\rho$$

Quantified MaxEnt II

- Quantified MaxEnt-prior:

$$p(\rho|\alpha, \mathbf{m}, I) = \frac{(2\pi)^{-N/2}\alpha^{N/2}}{\sqrt{\rho_1 \cdots \rho_N}} \exp\left(\alpha S\right)$$

- Generalized entropy with *default model* $\mathbf{m}$:

$$S = \sum_j \rho_j - m_j - \rho_j \log \frac{\rho_j}{m_j}$$

- Regularisation: $p(\rho|\alpha, \mathbf{m}) \rightarrow \delta(\rho - \mathbf{m})$ for $\alpha \rightarrow \infty$

Evidence Approximation I

- Posterior:

$$p(\rho|D,I) = \int d\alpha\, p(\rho|\alpha,D,I)p(\alpha|D,I)$$
$$\approx p(\rho|\alpha^*,D,I)$$
$$\sim \exp\left(-\frac{1}{2}\chi^2 + \alpha^* S\right)$$

- Optimal hyperparameter α^*:
  * Historic MaxEnt: $\chi^2 \stackrel{!}{=} N_d$
  * Classical MaxEnt: max $p(\alpha|D,I)$

Evidence Approximation II

- Probability for α:

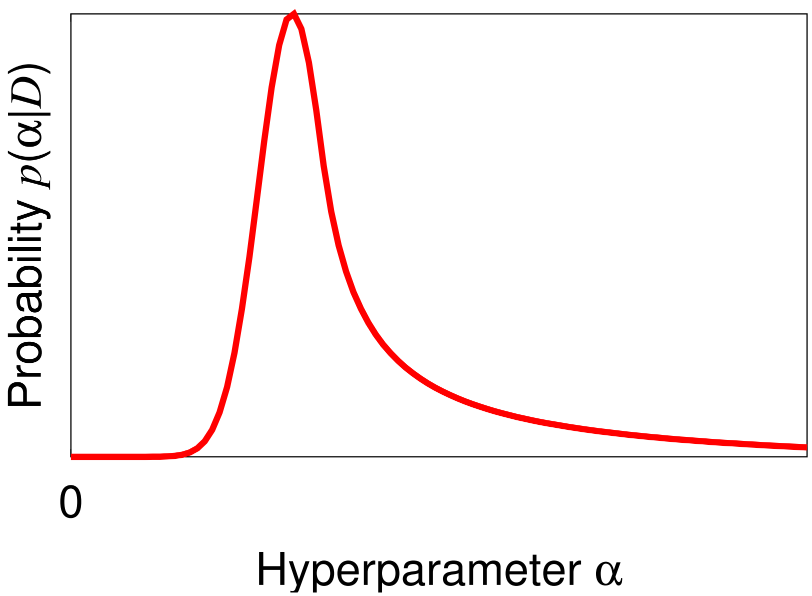
$$p(\alpha|D,I) = \int d^N\rho \underbrace{p(D|\rho,I)}_{\text{Likelihood}} \underbrace{p(\rho|\alpha,I)}_{\text{Entropy prior}} \underbrace{p(\alpha|I)}_{\text{Jeffreys' pr.}}$$

- Weak regularization $\alpha \to 0$:

$$p(\alpha|D,I) \sim \alpha^{N/2-1}$$

- Strong regularization $\alpha \to \infty$:

$$p(\alpha|D,I) \sim \frac{1}{\alpha}$$



Example: Abel-Inversion I

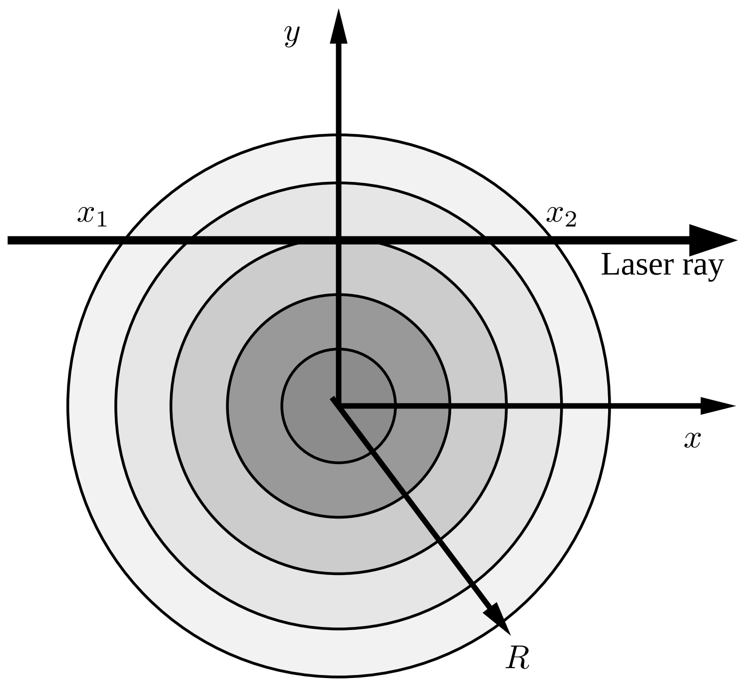
- Problem: Inference of plasma density from absorption measurements

  * Density is spherical symmetric

  * Absorbance of plasma:

$$A(y) \sim \int_{x_1}^{x_2} \rho(\sqrt{x^2+y^2})\, dx$$

$$= \int_y^R \frac{2r\rho(r)}{\sqrt{r^2-y^2}} dr$$



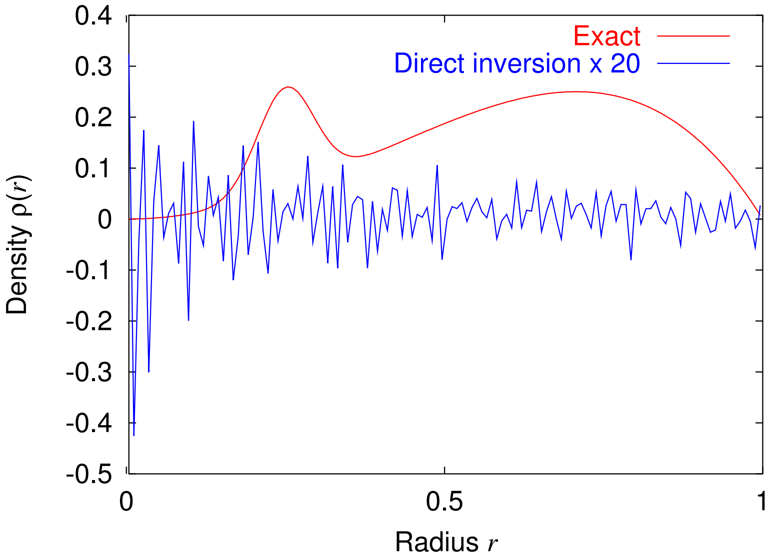
Example: Abel-Inversion II

- Discretization yields matrix equation:

$$A_i = \sum_j K_{ij} \rho_j$$

- Direct inversion of noisy data is hopeless!

- Apply Quantified MaxEnt!



Demonstration

- MatLab performing Abel-inversion
- Slide show:
  - ∗ Noise level
  - ∗ Default model
  - ∗ Smoothness of spline interpolation

Conclusion

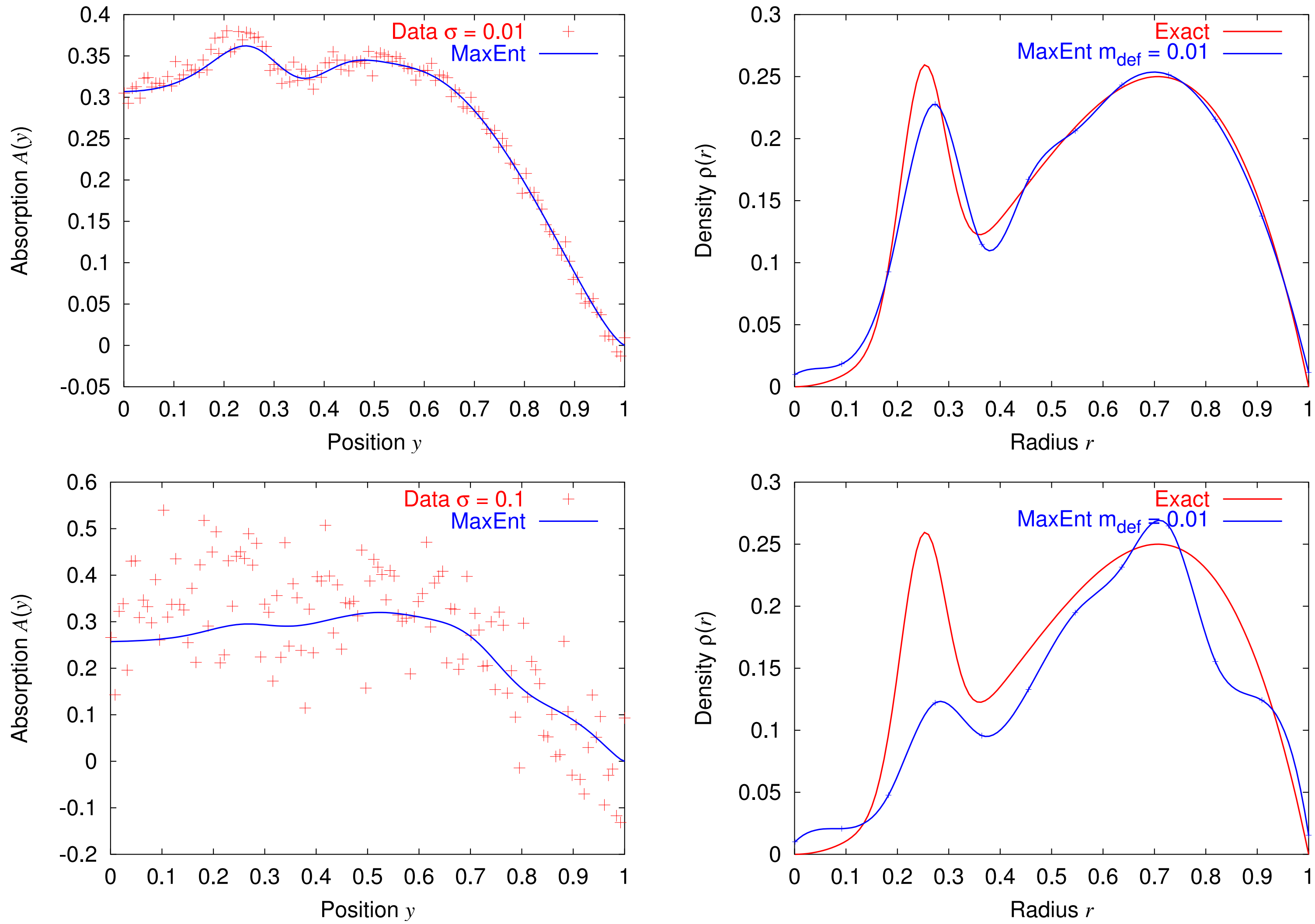
- Shannon Entropy as a Measure of Uncertainty
- The MaxEnt Principle for Assigning Probabilities
- Form-free Solutions with Quantified MaxEnt
- Abel-Inversion as an Ill-posed Problem

References

- Sivia, D.S.: *Data Analysis – A Bayesian Tutorial*, Oxford University Press 1996
- von der Linden, W. and A. Prüll: Skriptum *Wahrscheinlichkeitstheorie, Statistik und Datenanalyse* http://www.itp.tu-graz.ac.at/LV/Statistik/A_WS_pdf.pdf

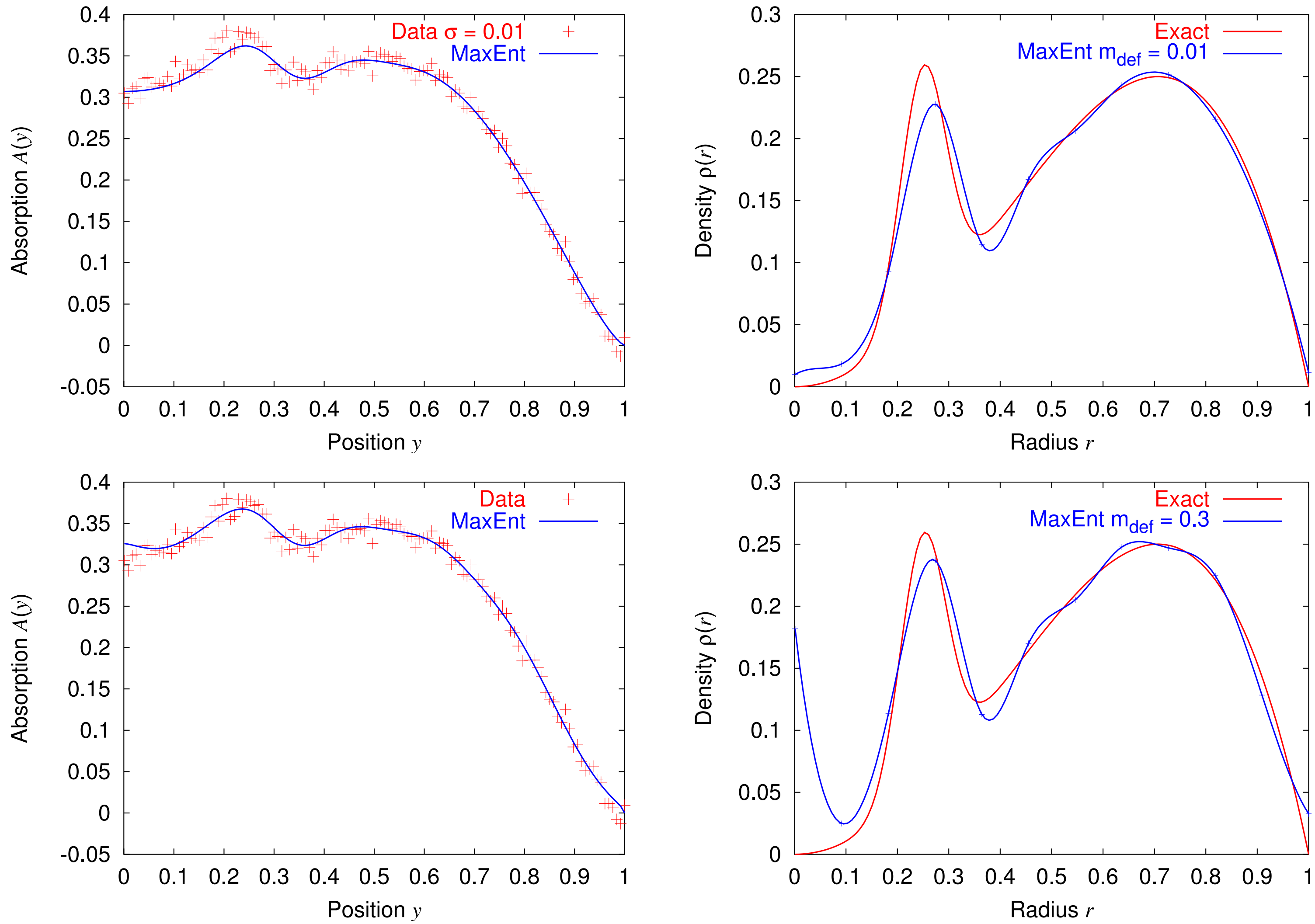
The End

Abel-Inversion — Noise level



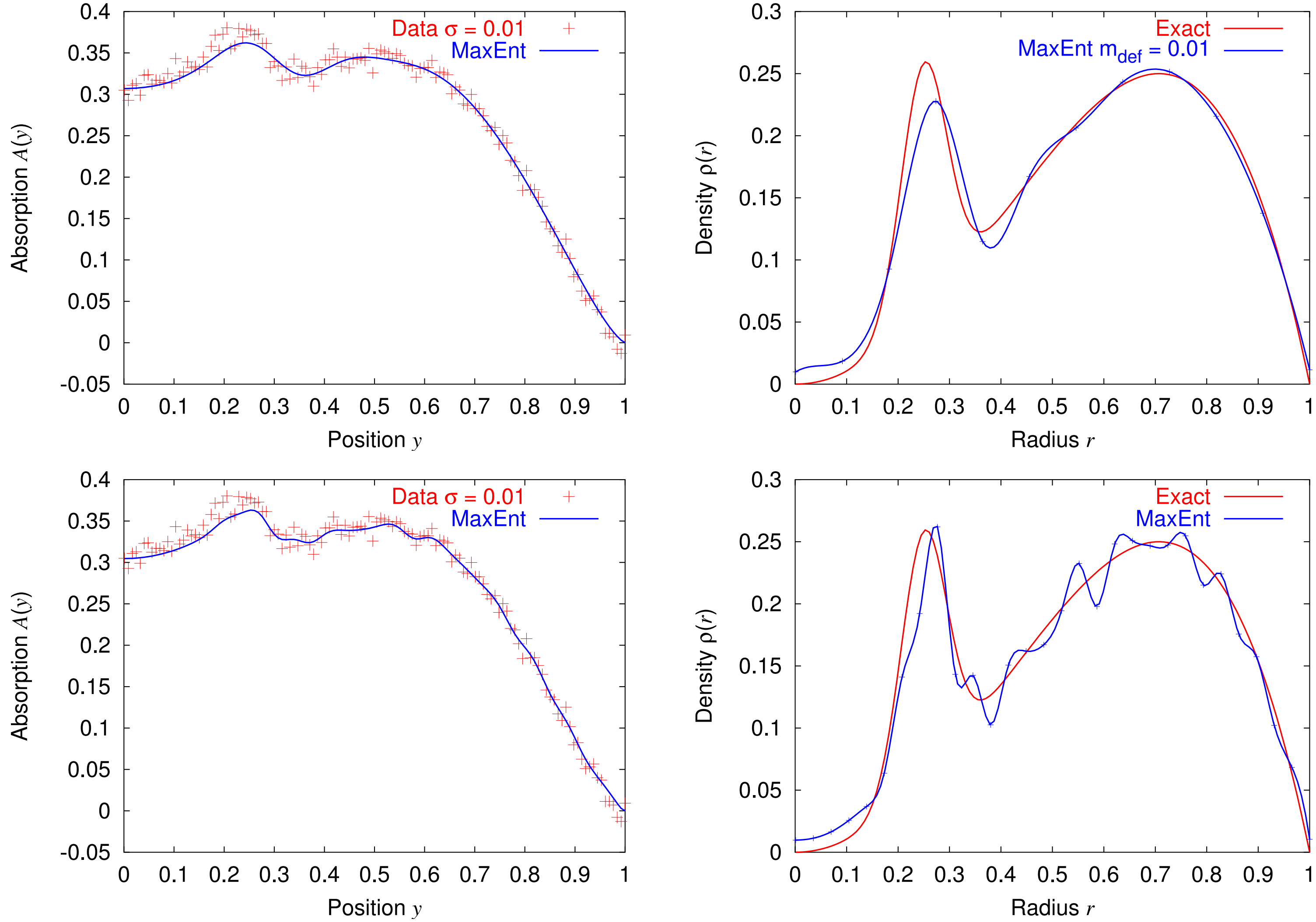
Back to the page displayed before.

Abel-Inversion — Default model



Back to the page displayed before.

Abel-Inversion — Number of Spline Knots



Back to the page displayed before.

Shannon's Theorem

- Consistency requirements:
  1. A measure of uncertainty $H_n(p_1, \ldots, p_n)$ exists
  2. Continuity of H_n with respect to all p_j
  3. Uncertainty $H_n\left(\frac{1}{n}, \ldots, \frac{1}{n}\right)$ increases with n:
  4. Addivity: $H_3(p_1, p_2, p_3) = H_2(p_1, q) + q\, H_2(\frac{p_2}{q}, \frac{p_3}{q})$

- Uncertainty and information entropy:

$$H \sim \left(- \sum_j p_j \log p_j \right)$$