Entfaltungs-Methoden in der Datenanalyse

Matthias Bartelt
Universität Dortmund

bartelt@physik.uni-dortmund.de

Übersicht

- Motivation

- Mathematisches Problem der Entfaltung
- Regularisierte Entfaltung
- Iterative Methode: Bayes´sche Entfaltung

- Stand der Dinge
- Ausblick

Motivation

- Vergleich experimenteller Daten mit theoretischen Vorhersagen
- Abweichungen in der Verteilung der Messdaten aufgrund begrenzte Akzeptanz und Messgenauigkeit des Detektors
  - Unzulänglichkeiten des Detektors in die Vorhersage einbeziehen Messdaten „entfalten“
- Vergleich von Messdaten verschiedener Experimente

Das mathematische Problem

- Gemessene Verteilung f_{Messung}

$$f_{\mathrm{Messung}}(x) = \int R(x, y)\, f_{\mathrm{Wahr}}(y)\, dy$$

- f_{Wahr} parametrisierbar $\Rightarrow$ Standard Methoden (z.B. Maximum Likelihood)
- Ansonsten $\Rightarrow$ Histogramm mit n Bins:

$$\nu_i = \sum_{j=1}^{n} R_{ij}\, \mu_j \qquad i = 1 \ldots n$$

Response-Funktion $R(x, y)$

- R hängt ausschließlich vom der verwendeten Messapparatur ab

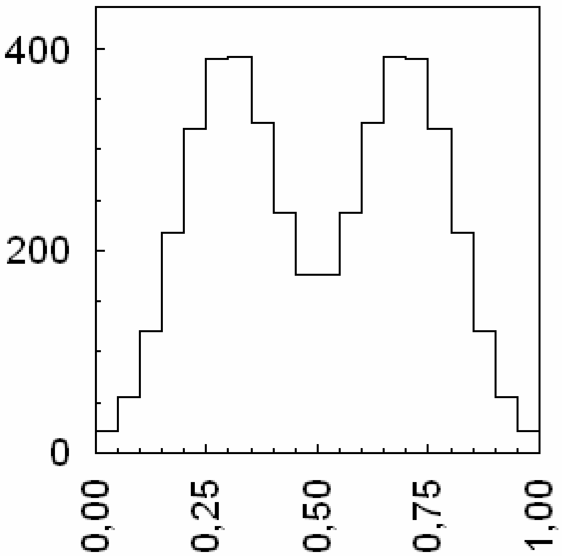
- R ist eine bedingte Wahrscheinlichkeit:

$$R_{ij} = P(\text{beobachtet in } i \mid \text{wahrer Wert in } j)$$

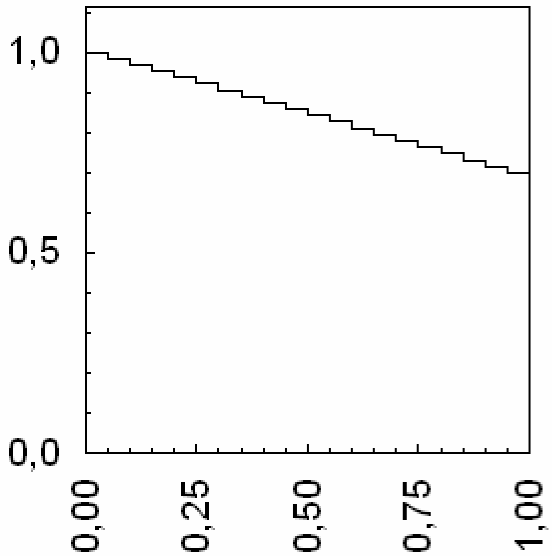
$$\sum_{i=1}^{n} R_{ij} = P(\text{beobachtet} \mid \text{wahrer Wert in } j)$$

$$\sum_{i=1}^{n} R_{ij} = \varepsilon_j$$

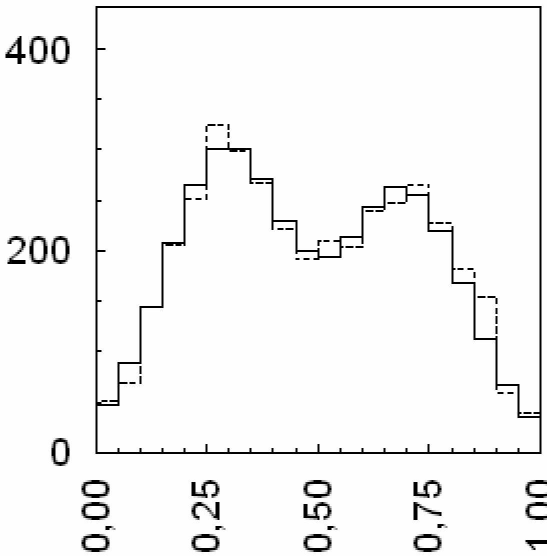
Ein Beispiel



Wahre Daten μ



Effizienz ε



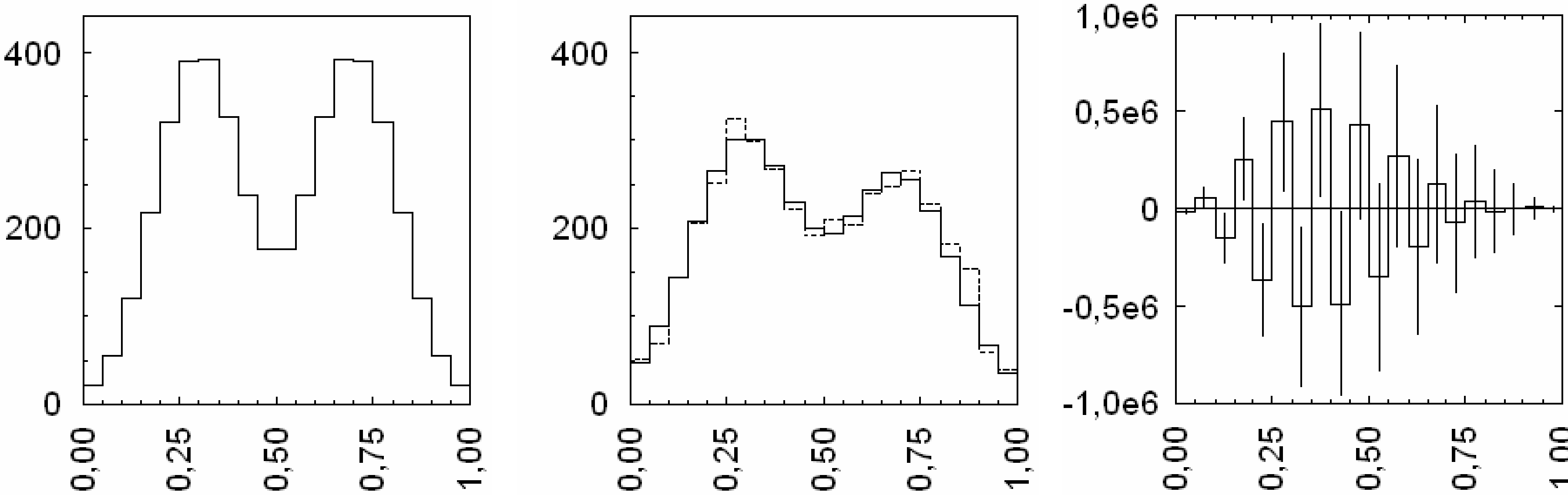
··· Messwerte x

– Erwartung ν

Entfaltung

- Entfalten durch Invertierung von R

$$\mu_i = \sum_{j=1}^{n} R^{-1}{}_{ij}\, \nu_j \Rightarrow \hat{\mu}_i = \sum_{j=1}^{n} R^{-1}{}_{ij}\, x_j$$



Wahre Daten μ

··· Messwerte x
– Erwartung ν

Erwartungswerte $\hat{\mu}$ für μ

Entfaltung

- Starke Schwankungen in den Erwartungswerten
- Einfache Korrekturfaktoren C_i (üblicherweise aus MC) liefern Erwartungswerte mit geringeren Schwankungen

$$C_i = \mu_{i,\mathrm{MC}} / \nu_{i,\mathrm{MC}}$$

$$\hat{\mu}_i = C_i \cdot x_i$$

Regularisierte Entfaltung

- Wirkungsfunktional $F(\mu)$ aufstellen, z.B.

$$F(\mu) = \sum_i \frac{\left(\nu_i - \sum_i R_{ij}\mu_j\right)^2}{\sigma(\mu)}$$

- Minimierung von F ergibt mathematisch die Lösung

- Auch hier treten starke Oszillationen auf

Regularisierte Entfaltung

- Unterdrückung der Oszillationen durch Nebenbedingungen: Regularisierungs-Funktion $S(\mu)$ und -Parameter α

- Neues Funktional Φ bilden:

$$\Phi(\mu) = \alpha \cdot F(\mu) + S(\mu)$$

- Minimierung von Φ führt dann auf die Lösung

Satz von Bayes

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

- $P(A|B)$ = Wahrscheinlichkeit für ein Ereignis B unter der Bedingung, dass A auftritt
- Für abzählbar viele Ereignisse gilt

$$P(B) = \sum_j P(B|A_j) \cdot P(A_j)$$

$$\text{d.h.}\ P(A_i|B) = \frac{P(B|A_i) \cdot P(A_i)}{\sum_j P(B|A_j) \cdot P(A_j)}$$

Bayes´sche Entfaltung

- Startbedingung $p_i^0 = 1/N$
- Erwartungswerte $\hat{\mu}_i$ bestimmen

$$\hat{\mu}_i^m = \tfrac{1}{\varepsilon_i} \sum_{j=1}^{N} P\left(\text{wahrer Wert in } i \,\middle|\, \text{beobachtet in } j\right) x_j$$

$$= \tfrac{1}{\varepsilon_i} \sum_{j=1}^{N} \frac{R_{ij}\; p_i^{m-1}}{\sum_k R_{jk}\; p_k^{m-1}} x_k$$

- Neue Wahrscheinlichkeiten berechnen

$$p_i^m = \hat{\mu}_i^m \Big/ \sum_i \hat{\mu}_i^m$$

Bisherige Entfaltung: RUN

- Regularisierte Entfaltung
- Begrenzte Anzahl (3) an Parameter
- Fortran-Code, der rechnerspezifische Funktionen verwendet

```
 PROGRAM UNFOLDMS
*******************************************************************
*                                                                 *
*  Main program for run-package of V.Blobel                       *
*   run only on alphas!!                                          *
```

- Optimierung läuft halbautomatisch

Ausblick / Zusammenfassung

- Neues Entfaltungsprogramm in C++
- Beliebige Anzahl an Parametern
- Implementierung verschiedener Methoden der Entfaltung
  - Regularisierte Entfaltung
  - Bayes´sche Entfaltung
  - Kombination aus Regularisierung und Bayes
- Automatische Optimierung