"The k2-filter"

Presented by:

Niels Jørgen Kjær, CERN, Switzerland

Work done together with Jorgen D'Hondt, IIHE (ULB-VUB), Brussels, Belgium

Outline of talk

- The Monte Carlo paradigm
- The Kalman filter
- What is optimal?
- The Minimum Variance Bounds
- The "actions" and constraints
- The optimization
  - Analysis corrections
  - Analysis parameter estimation
  - Combining analyses
- Optimizing globally
- Correlated measurements
- Robustness
- Systematic errors
- Conclusions and outlook

The Monte Carlo

- An integrator for average estimations.
- Reweights event i: $w_i = \frac{P_i(x^{\mathrm{new}})}{P_i(x^{\mathrm{gen}})}$
- Differentiates: $\frac{\partial w_i}{\partial x^{\mathrm{gen}}}$
- Analysis on generator level information:

$$\ln P(x) \simeq \sum_{i=1}^{n} 1 + \frac{\partial \ln(P_i(x))}{\partial x}\Big|_{x^{\mathrm{gen}}}(x - x^{\mathrm{gen}}) + \frac{1}{2}\frac{\partial^2 \ln(P_i(x))}{\partial x^2}\Big|_{x^{\mathrm{gen}}}(x - x^{\mathrm{gen}})^2$$

If $g_i = \frac{\partial \ln(P_i(x))}{\partial x}\big|_{x^{\mathrm{gen}}}$ and $g_{i2} = -\frac{\partial^2 \ln(P_i(x))}{\partial x^2}\big|_{x^{\mathrm{gen}}}$

$$x - x^{\mathrm{gen}} \simeq \sum_i g_i / \sum_i g_{i2}$$

g_i are linear quantities. Bckg $g_i = 0$

$g_i \propto$ how much an event changes the result

Similar on the analysis level: a_i, a_{i2}

$$x - x^{\mathrm{ref}} \simeq \sum_i a_i / \sum_i a_{i2}$$

a_i and g_i are denoted “actions”

The Kalman filter

- The Kalman filter works on a
  - A state with estimated parameters
  - A covariance matrix of the parameters
  - A model for stochastic linear noise
  - One new linear piece of information
- Linearly propagates information to update
  - the state with estimated parameters
  - the covariance matrix of the parameters
  - the model for linear noise
- Only needs to solve set of linear equations
- Very powerful optimal Bayesian analysis tool
- Used very many places

The Kalman filter process controls processes

HEP analyses are processes...

What is optimal?

- Likelihood analyses need to be calibrated.
- In linear approximation:
  Bias(=0), slope(=1), and rms. of pull(=1).
  - Bias=0 $\Leftrightarrow$ $\sum_i a_i = 0$
  - Slope=1 $\Leftrightarrow$ $\sum_i a_i g_i = \sum_i a_{i2}$
  - Pull=1 $\Leftrightarrow$ $\sum_i a_{i2} = \sum_i a_i^2$
- a_{i2} and g_{i2} are not relevant!
- Error becomes $\sigma^2 = 1/(\sum_i a_i^2)$
- Constraints:
  - $\sum_i a_i = 0$
  - $\sum_i a_i(a_i - g_i) = 0$
- Optimal is to minimize the error while fulfilling the constraints.

The Minimum Variance Bounds

Constraints can be ignored if:

$1/\sigma^2 = \sum_i (g_i^2 - (a_i - g_i)^2)$

And this error is minimized.

Related to Fisher information and MVBs:

$$\mathcal{I}_g = \sum_i g_i^2$$

$$\mathcal{I}_a = \sum_i g_i^2 - (a_i - g_i)^2$$

$\sum_i (a_i - g_i)^2$ is information loss $\Rightarrow$ Minimize

The "actions" and constraints

- The "actions", a_i and g_i, are linear
- $\sum_i (a_i - g_i)^2 \Rightarrow$ linear optimization
- $\sum_i a_i(a_i - g_i) = 0$ is not linear

Solve non-linear equations by either

Iterative constrained fit with Lagrange multipliers

$a_i' = \alpha a_i + \beta$, minimize $\sum_i (a_i' - g_i)^2$.

Iterative fits: More constraints possible

Make analysis independent of for example:

- Background level: $\sum_{i \in b} a_i = 0$
- MC parameter: $\partial(\sum_i a_i)/\partial p = 0$
- MC model:$\sum_{i \in 1} a_i = 0 \wedge \sum_{i \in 2} a_i = 0$

The optimization

- Any method valid to optimize $\mathcal{I}_a$
- Actions linear: Linear is simpler
- Corrections easier than total analysis
- Beware of overtraining: Control
- Optimization should be fast
- Optimization $\sim$ linear process
  $\Rightarrow$ Kalman filter to control optimization?

Analysis corrections

The analyses can be optimized as a function of any function of any observable, $f(\vec{o})$. Choose only to look on corrections to analyses

$$\Delta a_i(f(\vec{o})) =< g_i - a_i > (f(\vec{o}))$$

Iterative procedure. When to stop?

The action is in itself an observable

$$\Delta a_i(a_i, f(\vec{o})) =< g_i - a_i > (a_i, f(\vec{o}))$$

Takes higher order correlations into account.

Even Δa_i, become “observables”.

Only one optimal observable exists, but a hierarchy of directions in observable space $\vec{o}$:

Another Kalman filter?

$\Delta a_i \Rightarrow$ Information on how analyses work.

Analysis parameter estimation

All parameters in an analysis can be optimized

$$a_i(\vec{p} + \Delta\vec{p}) = a_i(\vec{p}) + \frac{\partial a_i(\vec{p})}{\partial \vec{p}} \Delta\vec{p}$$

$$\mathcal{I} = \sum_i g_i^2 - (a_i(\vec{p}) + \frac{\partial a_i(\vec{p})}{\partial \vec{p}} \Delta\vec{p} - g_i)^2$$

$$\begin{aligned} \Delta\vec{p} &= \left(\sum_i \frac{\partial a_i(\vec{p})}{\partial \vec{p}} \cdot \frac{\partial a_i(\vec{p})}{\partial \vec{p}} \right)^{-1} \\ &\times \sum_i \frac{\partial a_i(\vec{p})}{\partial \vec{p}} \cdot (g_i - a_i(\vec{p})) \end{aligned}$$

The significance of $\Delta\vec{p}$ from Jackknife

Only correct the significant changes

All correlations taken into account

Combining analyses

- k2-filter linear $\Rightarrow$ combination easy.

$a_i(1)$, $a_i(2)$ actions from analyses 1,2.

$$a_i(a_i(1), a_i(2)) = < g_i > (a_i(1), a_i(2))$$

$$\Delta a_i(1)(a_i(1), a_i(2)) =$$
$$< g_i - a_i(1) > (a_i(1), a_i(2))$$

Exact methods

More involved with systematics

Optimizing globally

Global likelihood curves $\Rightarrow$ Reweight MC

For each reweighted x^{new} optimize $a_i(x^{\text{new}})$

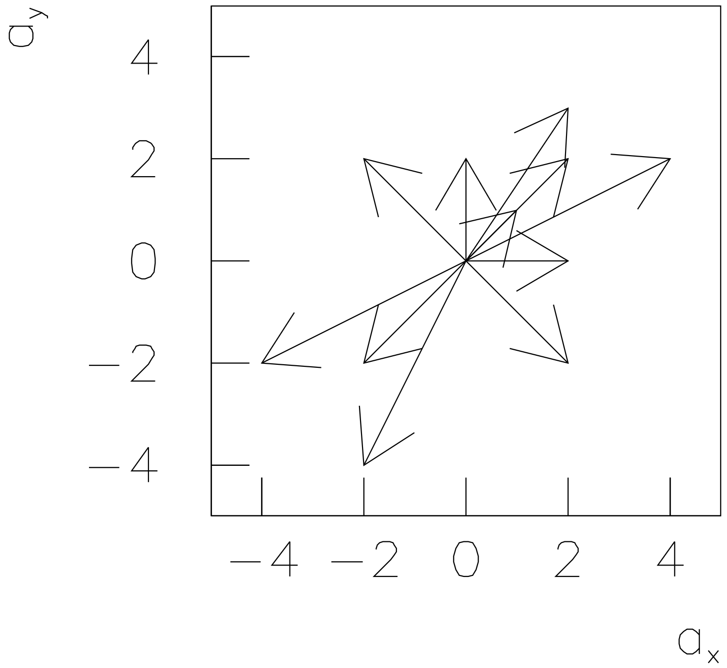
Construct global likelihood as:

$$L_i(x) = \exp\left(\int^x a_i(x^{\text{new}})dx^{\text{new}}\right)$$

- Linear interpolation $\Rightarrow$ Easy and continuous
- $a_i \sim$ linear $\Rightarrow$ Few (2) points
- More points near phase space boundary
- Global optimum $\Rightarrow$ a_{i2} unbiased
  $\Rightarrow$ Optimized event-by-event errors

Correlated measurements

- Meas correlated parameters simultaneously
- Linear approximation $\Rightarrow$ Independent!
  - Actions can be optimized independently
  - Event-by-event behaviour $\Rightarrow$ Correlations
- More difficult:
  - Analysis parameter optimization
  - Global optimization



Robustness

- Most important to avoid “overtraining”
- Apply only significant corrections
- Rule of thumb: 3 standard deviations.

$$\mathcal{I}' - \mathcal{I} = \sum_i (a_i - a'_i)(a_i + a'_i - 2g_i)$$

$$\mathtt{VAR}(\mathcal{I}' - \mathcal{I}) \simeq$$

$$\sum_i (a_i - a'_i)^2 (a_i + a'_i - 2g_i)^2$$

$$\mathtt{S} = \frac{\mathcal{I}' - \mathcal{I}}{\sqrt{\mathtt{VAR}(\mathcal{I}' - \mathcal{I})}}$$

Overtraining can be kept under control and even computed and corrected for!

Systematic errors

- More systematics?
  - Training on irrelevant observables
  - Outlayers
  - Data/MC disagreement
  - More information
- More systematic control?
  - Data/MC on $\boldsymbol{\Delta a_i}$
  - Check outlayers
  - More information
  - Make analysis systematic independent
- Why not “measure” the systematic?
  - $\sim$ All systematics MC parameterizable
  - Simultaneous measurement with parameter
  - Optimal treatment

Some results I

- DELPHI 4-quark m_{W} measurement at LEP2.
  - Highly involved Bayesian convolution analysis
  - $\sim$ 20 different features explored
  - Many years of work already put in
- 12 relevant event observables were checked
  - a_i, a_{i2}, y_{34}, etc.
  - 10 corrections were significant $> 3\sigma$
  - 1% improvement to error
- Analysis parameters were tuned
  - Constrained fit params $\Longrightarrow$ No improvement
  - Relative weight of 3 different jet clusterings:
    - $\ast$ Durham, Cambridge, Diclus
    - $\ast$ (1/3,1/3,1/3) $\Longrightarrow$ (0.45,0.36,0.19) $\pm$ 0.02

Some results II

Measurement of κ in the "SK1 model"

$P_{i\mathrm{CR}}(\kappa) = \exp(-\kappa\mathcal{O}_i)$

$g_{i\mathrm{CR}}(\kappa) = -\mathcal{O}_i$. $\mathcal{O}_i$ is string overlap

$g_{i\mathrm{NR}}(\kappa) = \mathcal{O}_i/(1 - \exp(-\kappa\mathcal{O}_i))$

Events generated in both scenarios: CR and NR

κ is correlated to $m_{\mathrm{W}} \Rightarrow$ syst error om m_{W}.

κ also correlates with

$a_1(m_{\mathrm{W}})$, $a_2(m_{\mathrm{W}})$, $a_1'(m_{\mathrm{W}})$, $a_2'(m_{\mathrm{W}})$.

Classical analysis look only on $m_{\mathrm{W}}' - m_{\mathrm{W}}$, which is nearly m_{W} independent.

k2-filter combines all information and applies m_{W} independence as constraint

Preliminary results show k2-filter gives factor 2 smaller error.

Systematic cross checks failed

One Δa_i ingredient is not consistent in data/MC

Under investigation.

Conclusions and outlook

- k2-filter: Consistent and complete implementation of Kalman filters in MC analyses
- k2-filter optimal likelihood analyses with MC
- MC both integrator and differentiator
- Both Bayesian and Frequentist's
- Applicable to nearly all measurements
- Slightly different implementation for searches
- Both smart and intelligent at the same time?