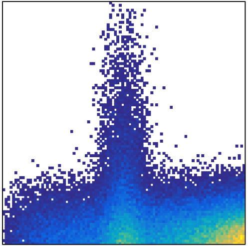
Using sWeights to disentangle signal and background

Michael Schmelling / MPI for Nuclear Physics

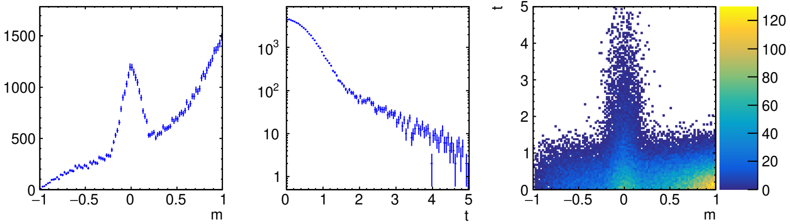
Outline

- sWeights classic
- sWeights as orthogonal functions
- Summary and outlook



1 sWeights classic

❖ toy example: extracting the lifetime distribution of an unstable particle



- data model in discriminant variable m and control variable t

$$f(m,t) = \underbrace{N_0\, d_0(m)\, s(t)}_{\text{signal}} + \underbrace{N_1\, d_1(m)\, b(t)}_{\text{background}}$$

  - known: signal PDF $d_0(m)$ and background PDF $d_1(m)$
  - wanted:signal PDF $s(t)$ and background PDF $b(t)$

Warm-up: sideband subtraction

- applicable if signal and background densities factorise in m and t

Warm-up: sideband subtraction

- applicable if signal and background densities factorise in m and t
- start from histograms of the t-distribution for different regions in m
  - m in signal window: $S(t) = N_0\, s(t) + f_b\, N_1\, b(t)$
  - m in sideband: $B(t) = (1 - f_b)\, N_1\, b(t)$
  - determine f_b from the data model in m
  - solve for $N_0 s(t) = S(t) - f_b/(1 - f_b) B(t)$

Warm-up: sideband subtraction

- applicable if signal and background densities factorise in m and t
- start from histograms of the t-distribution for different regions in m
  - m in signal window: $S(t) = N_0\, s(t) + f_b\, N_1\, b(t)$
  - m in sideband: $B(t) = (1 - f_b)\, N_1\, b(t)$
  - determine f_b from the data model in m
  - solve for $N_0 s(t) = S(t) - f_b/(1 - f_b) B(t)$
- equivalent: use weights when filling the t-distribution
  - m in signal window: $w_s = 1$
  - m in sideband: $w_s = -f_b/(1 - f_b)$

Warm-up: sideband subtraction

- applicable if signal and background densities factorise in m and t
- start from histograms of the t-distribution for different regions in m
  - m in signal window: $S(t) = N_0\, s(t) + f_b\, N_1\, b(t)$
  - m in sideband: $B(t) = (1 - f_b)\, N_1\, b(t)$
  - determine f_b from the data model in m
  - solve for $N_0 s(t) = S(t) - f_b/(1 - f_b) B(t)$
- equivalent: use weights when filling the t-distribution
  - m in signal window: $w_s = 1$
  - m in sideband: $w_s = -f_b/(1 - f_b)$
- optimisation: continuous “sWeights $w_s(m)$”

➜ M. Pivk & F. Le Diberder, arXiv:physics/0402083, NIMA 555 (2005) 356

The sWeights recipe

- fit N_0 and N_1 of the data model in m: $f(m) = N_0\, d_0(m) + N_1\, d_1(m)$

The sWeights recipe

- fit N_0 and N_1 of the data model in m: $f(m) = N_0\, d_0(m) + N_1\, d_1(m)$
- with N_0 and N_1 and their covariance matrix elements C_{00} and C_{01} determine:

$$w_s(m) = \frac{C_{00}\, d_0(m) + C_{01}\, d_1(m)}{N_0\, d_0(m) + N_1\, d_1(m)}$$

The sWeights recipe

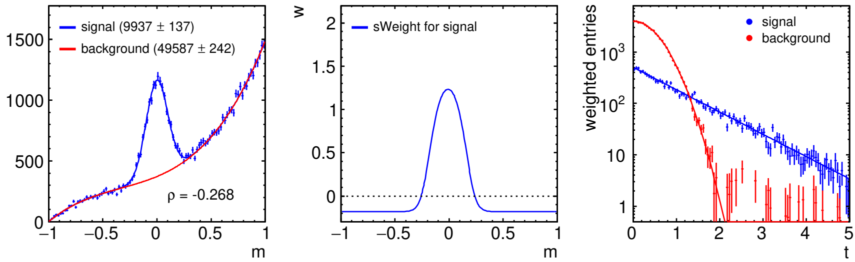
- fit N_0 and N_1 of the data model in m: $f(m) = N_0\, d_0(m) + N_1\, d_1(m)$
- with N_0 and N_1 and their covariance matrix elements C_{00} and C_{01} determine:

$$w_s(m) = \frac{C_{00}\, d_0(m) + C_{01}\, d_1(m)}{N_0\, d_0(m) + N_1\, d_1(m)}$$

  - $w_s(m)$ is a function of only m
  - use $w_s(m)$ as weight when filling t-histograms to obtain $N_0\, s(t)$
  - background distribution can be obtained by $w_b(m) = 1 - w_s(m)$
  - bin contents in bin Δt are

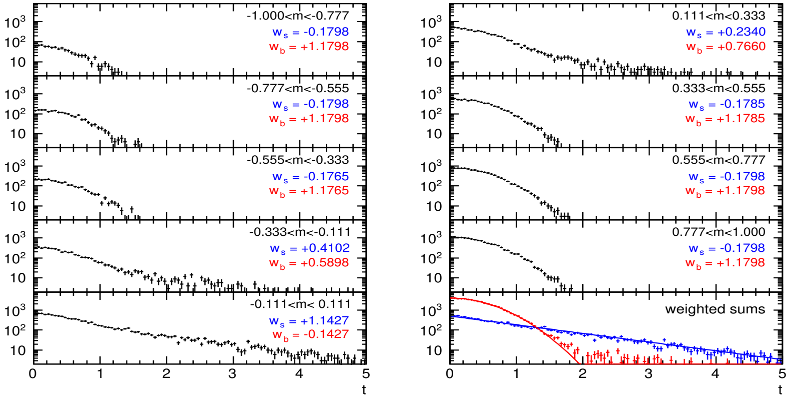
$$s(\Delta t) = \sum_{\text{entries in}\Delta t} w_s(m)$$

➤ toy model



- signal and background distributions in t are correctly disentangled
- requires only a single normalisation fit with two free parameters

illustration how sWeights assemble signal and background density ➜



- this plot: weighted sums of t-distributions in 9 slices of m
- true sWeights: continuum limit with infinitesimally thin slices in m

Impact of uncertainties in sWeights extracted from the data

Impact of uncertainties in sWeights extracted from the data

consider a bin $s(\Delta t)$ obtained by a summing contributions $a_k(\Delta t)$ from k bins in m

$$s(\Delta t) = \sum_{\text{bins } k} w_k \, a_k(\Delta t)$$

Impact of uncertainties in sWeights extracted from the data

consider a bin $s(\Delta t)$ obtained by a summing contributions $a_k(\Delta t)$ from k bins in m

$$s(\Delta t) = \sum_{\text{bins } k} w_k \, a_k(\Delta t)$$

error propagation to get the variance of s

$$\sigma^2(s) = \sum_{\text{bins } k} \left(\frac{\partial s}{\partial a_k}\right)^2 \sigma^2(a_k) + \left(\frac{\partial s}{\partial w_k}\right)^2 \sigma^2(w_k)$$

Impact of uncertainties in sWeights extracted from the data

consider a bin $s(\Delta t)$ obtained by a summing contributions $a_k(\Delta t)$ from k bins in m

$$s(\Delta t) = \sum_{\text{bins } k} w_k \, a_k(\Delta t)$$

error propagation to get the variance of s

$$\sigma^2(s) = \sum_{\text{bins } k} \left(\frac{\partial s}{\partial a_k}\right)^2 \sigma^2(a_k) + \left(\frac{\partial s}{\partial w_k}\right)^2 \sigma^2(w_k) = \sum_{\text{bins } k} w_k^2 \, a_k + a_k^2 \sigma^2(w_k)$$

Impact of uncertainties in sWeights extracted from the data

consider a bin $s(\Delta t)$ obtained by a summing contributions $a_k(\Delta t)$ from k bins in m

$$s(\Delta t) = \sum_{\text{bins } k} w_k \, a_k(\Delta t)$$

error propagation to get the variance of s

$$\sigma^2(s) = \sum_{\text{bins } k} \left(\frac{\partial s}{\partial a_k}\right)^2 \sigma^2(a_k) + \left(\frac{\partial s}{\partial w_k}\right)^2 \sigma^2(w_k) = \sum_{\text{bins } k} w_k^2 a_k + a_k^2 \sigma^2(w_k)$$

for infinitesimal m bins one has $a_k \in \{0, 1\}$, i.e. $a_k^2 = a_k$,

Impact of uncertainties in sWeights extracted from the data

consider a bin $s(\Delta t)$ obtained by a summing contributions $a_k(\Delta t)$ from k bins in m

$$s(\Delta t) = \sum_{\text{bins } k} w_k \, a_k(\Delta t)$$

error propagation to get the variance of s

$$\sigma^2(s) = \sum_{\text{bins } k} \left(\frac{\partial s}{\partial a_k}\right)^2 \sigma^2(a_k) + \left(\frac{\partial s}{\partial w_k}\right)^2 \sigma^2(w_k) = \sum_{\text{bins } k} w_k^2 a_k + a_k^2 \sigma^2(w_k)$$

for infinitesimal m bins one has $a_k \in \{0, 1\}$, i.e. $a_k^2 = a_k$, and thus

$$\sigma^2(s) = \sum_{\text{bins } k} \left[w_k^2 + \sigma^2(w_k)\right] a_k = \sum_{\text{events } k} \left[w_k^2 + O\left(\frac{1}{N}\right)\right]$$

Impact of uncertainties in sWeights extracted from the data

consider a bin $s(\Delta t)$ obtained by a summing contributions $a_k(\Delta t)$ from k bins in m

$$s(\Delta t) = \sum_{\text{bins } k} w_k \, a_k(\Delta t)$$

error propagation to get the variance of s

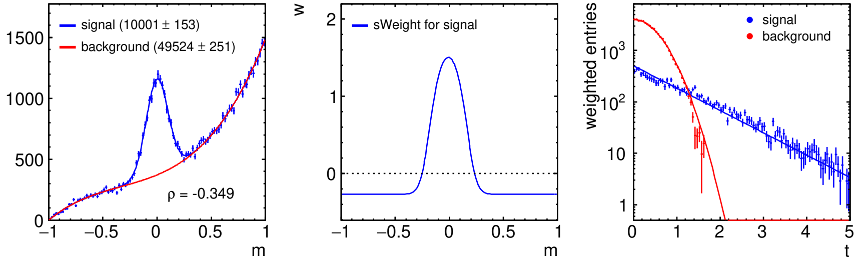
$$\sigma^2(s) = \sum_{\text{bins } k} \left(\frac{\partial s}{\partial a_k}\right)^2 \sigma^2(a_k) + \left(\frac{\partial s}{\partial w_k}\right)^2 \sigma^2(w_k) = \sum_{\text{bins } k} w_k^2 a_k + a_k^2 \sigma^2(w_k)$$

for infinitesimal m bins one has $a_k \in \{0, 1\}$, i.e. $a_k^2 = a_k$, and thus

$$\sigma^2(s) = \sum_{\text{bins } k} \left[w_k^2 + \sigma^2(w_k)\right] a_k = \sum_{\text{events } k} \left[w_k^2 + O\left(\frac{1}{N}\right)\right]$$

- asymptotically the variance in a Δt-bin is $\sum w^2$
- for finite statistics there are corrections of $O(1/N)$
- $O(1/N)$ terms occur also as covariances between bins

Fitting shape parameters in addition to the normalisations



- modified covariance matrix gives wrong sWeights and biased results
- need to adapt the sWeights method when also shape parameters are fitted, e.g.
  - first fit with all parameters floating
  - then fix all shape parameters and get the sWeights from a pure normalisation fit

2 sWeights as orthogonal functions

❖ properties of sWeights classic

- convenient, elegant and widely used method to disentangle signal and background
- fixable problem: fitting of shape parameters in addition to normalisations
- not fixable: fitting with constraints on yield ratios
  - covariance matrix elements do not give correct sWeights
  - lifting the constraint would fundamentally change the analysis
- impossible: sWeights for a smaller range in the discriminant variable than used in the fit
  - enlarged ranges may be needed to constrain a complicated background model
  - application of sWeights would maybe profit from focus on signal region
- also wanted: understanding the significance of the normalisation fit

reformulation of the sWeights method ➜

❖ starting from the data model . . .

$$f(m,t) = N_0\, d_0(m)\, s(t) + N_1\, d_1(m)\, b(t)$$

❖ starting from the data model . . .

$$f(m,t) = N_0\, d_0(m)\, s(t) + N_1\, d_1(m)\, b(t)$$

consider a bin Δt of the sWeighted signal histogram

$$s(\Delta t)$$

❖ starting from the data model . . .

$$f(m,t) = N_0\, d_0(m)\, s(t) + N_1\, d_1(m)\, b(t)$$

relation to the signal density $s(t)$

$$s(\Delta t) = \int_{\Delta t} dt\, N_0\, s(t)$$

❖ starting from the data model . . .

$$f(m,t) = N_0\, d_0(m)\, s(t) + N_1\, d_1(m)\, b(t)$$

approximation by the sum over sWeights

$$s(\Delta t) = \int_{\Delta t} dt\, N_0\, s(t) = \sum_{\text{entries in} \Delta t} w_s(m)$$

❖ starting from the data model . . .

$$f(m,t) = N_0\, d_0(m)\, s(t) + N_1\, d_1(m)\, b(t)$$

replace the sum by the integral over the density

$$s(\Delta t) = \int_{\Delta t} dt\, N_0\, s(t) = \sum_{\text{entries in}\Delta t} w_s(m) = \int_{\Delta t} dt \int dm\, f(m,t)\, w_s(m)$$

❖ starting from the data model . . .

$$f(m,t) = N_0\, d_0(m)\, s(t) + N_1\, d_1(m)\, b(t)$$

consider on a bin Δt of the sWeighted signal histogram

$$s(\Delta t) = \int_{\Delta t} dt\, N_0\, s(t) = \sum_{\text{entries in}\Delta t} w_s(m) = \int_{\Delta t} dt \int dm\, f(m,t)\, w_s(m)$$

relation between signal density, data model and sWeights

$$N_0\, s(t) = \int dm\, f(m,t)\, w_s(m)$$

❖ starting from the data model . . .

$$f(m,t) = N_0\, d_0(m)\, s(t) + N_1\, d_1(m)\, b(t)$$

consider on a bin Δt of the sWeighted signal histogram

$$s(\Delta t) = \int_{\Delta t} dt\, N_0\, s(t) = \sum_{\text{entries in}\Delta t} w_s(m) = \int_{\Delta t} dt \int dm\, f(m,t)\, w_s(m)$$

substitute the explicit form for $f(m,t)$

$$N_0\, s(t) = \int dm\, f(m,t)\, w_s(m) = \int dm \left[N_0\, d_0(m)\, s(t) + N_1\, d_1(m)\, b(t) \right] w_s(m)$$

❖ starting from the data model . . .

$$f(m,t) = N_0\, d_0(m)\, s(t) + N_1\, d_1(m)\, b(t)$$

consider on a bin Δt of the sWeighted signal histogram

$$s(\Delta t) = \int_{\Delta t} dt\, N_0\, s(t) = \sum_{\text{entries in}\Delta t} w_s(m) = \int_{\Delta t} dt \int dm\, f(m,t)\, w_s(m)$$

separate into a sum of two integrals

$$N_0\, s(t) = \int dm\, f(m,t)\, w_s(m) = \int dm \left[N_0\, d_0(m)\, s(t) + N_1\, d_1(m)\, b(t) \right] w_s(m)$$

$$= N_0\, s(t) \int dm\, d_0(m)\, w_s(m) + N_1\, b(t) \int dm\, d_1(m)\, w_s(m)$$

❖ starting from the data model . . .

$$f(m,t) = N_0\, d_0(m)\, s(t) + N_1\, d_1(m)\, b(t)$$

consider on a bin Δt of the sWeighted signal histogram

$$s(\Delta t) = \int_{\Delta t} dt\, N_0\, s(t) = \sum_{\text{entries in}\Delta t} w_s(m) = \int_{\Delta t} dt \int dm\, f(m,t)\, w_s(m)$$

read off the condition for sWeights to pick out the signal density

$$N_0\, s(t) = \int dm\, f(m,t)\, w_s(m) = \int dm \left[N_0\, d_0(m)\, s(t) + N_1\, d_1(m)\, b(t) \right] w_s(m)$$

$$= N_0\, s(t) \underbrace{\int dm\, d_0(m)\, w_s(m)}_{1} + N_1\, b(t) \underbrace{\int dm\, d_1(m)\, w_s(m)}_{0}$$

❖ starting from the data model ...

$$f(m,t) = N_0\, d_0(m)\, s(t) + N_1\, d_1(m)\, b(t)$$

consider on a bin Δt of the sWeighted signal histogram

$$s(\Delta t) = \int_{\Delta t} dt\, N_0\, s(t) = \sum_{\text{entries in}\Delta t} w_s(m) = \int_{\Delta t} dt \int dm\, f(m,t)\, w_s(m)$$

read off the condition for sWeights to pick out the signal density

$$N_0\, s(t) = \int dm\, f(m,t)\, w_s(m) = \int dm \left[N_0\, d_0(m)\, s(t) + N_1\, d_1(m)\, b(t) \right] w_s(m)$$

$$= N_0\, s(t) \underbrace{\int dm\, d_0(m)\, w_s(m)}_{1} + N_1\, b(t) \underbrace{\int dm\, d_1(m)\, w_s(m)}_{0}$$

two necessary conditions ➜

❖ key observation regarding sWeights

$$\int dm\, d_0(m) w_s(m) = 1 \qquad \text{and} \qquad \int dm\, d_1(m) w_s(m) = 0$$

- $w_s(m)$ is orthogonal to the background PDF $d_1(m)$
- $w_s(m)$ is normalised with respect to the signal PDF $d_0(m)$

❖ key observation regarding sWeights

$$\int dm\, d_0(m) w_s(m) = 1 \qquad \text{and} \qquad \int dm\, d_1(m) w_s(m) = 0$$

- $w_s(m)$ is orthogonal to the background PDF $d_1(m)$
- $w_s(m)$ is normalised with respect to the signal PDF $d_0(m)$
- infinitely many functions satisfy these constraints
  - ► select the one that minimises the variance of the total signal yield

$$N_0 = \sum_{\text{events}} w_s(m)$$

❖ key observation regarding sWeights

$$\int dm\, d_0(m) w_s(m) = 1 \qquad \text{and} \qquad \int dm\, d_1(m) w_s(m) = 0$$

- $w_s(m)$ is orthogonal to the background PDF $d_1(m)$
- $w_s(m)$ is normalised with respect to the signal PDF $d_0(m)$
- infinitely many functions satisfy these constraints
  - ► select the one that minimises the variance of the total signal yield

$$N_0 = \sum_{\text{events}} w_s(m)$$

$$\sigma^2(N_0) = \sum_{\text{events}} w_s^2(m)$$

❖ key observation regarding sWeights

$$\int dm\, d_0(m) w_s(m) = 1 \qquad \text{and} \qquad \int dm\, d_1(m) w_s(m) = 0$$

- $w_s(m)$ is orthogonal to the background PDF $d_1(m)$
- $w_s(m)$ is normalised with respect to the signal PDF $d_0(m)$
- infinitely many functions satisfy these constraints
  - select the one that minimises the variance of the total signal yield

$$N_0 = \sum_{\text{events}} w_s(m)$$

$$\sigma^2(N_0) = \sum_{\text{events}} w_s^2(m) = \int dt\, dm\, f(m,t)\, w_s^2(m)$$

❖ key observation regarding sWeights

$$\int dm\, d_0(m) w_s(m) = 1 \qquad \text{and} \qquad \int dm\, d_1(m) w_s(m) = 0$$

- $w_s(m)$ is orthogonal to the background PDF $d_1(m)$
- $w_s(m)$ is normalised with respect to the signal PDF $d_0(m)$
- infinitely many functions satisfy these constraints
  - select the one that minimises the variance of the total signal yield

$$N_0 = \sum_{\text{events}} w_s(m)$$

$$\sigma^2(N_0) = \sum_{\text{events}} w_s^2(m) = \int dt\, dm\, f(m,t)\, w_s^2(m) = \text{min}$$

constrained minimisation problem ➜

❖ result

$$w_s(m) = \frac{\alpha_0\, d_0(m) + \alpha_1\, d_1(m)}{N_0\, d_0(m) + N_1\, d_1(m)}$$

with α_0 and α_1 given by the solution of a matrix equation

$$\begin{pmatrix} W_{00} & W_{01} \\ W_{01} & W_{11} \end{pmatrix} \cdot \begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad W_{kl} = \int dm\, \frac{d_k(m)\, d_l(m)}{N_0\, d_0(m) + N_1\, d_1(m)}$$

❖ result

$$w_s(m) = \frac{\alpha_0\, d_0(m) + \alpha_1\, d_1(m)}{N_0\, d_0(m) + N_1\, d_1(m)}$$

with α_0 and α_1 given by the solution of a matrix equation

$$\begin{pmatrix} W_{00} & W_{01} \\ W_{01} & W_{11} \end{pmatrix} \cdot \begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad W_{kl} = \int dm\, \frac{d_k(m)\, d_l(m)}{N_0\, d_0(m) + N_1\, d_1(m)}$$

- the result is equivalent to the normalisation-fit recipe
- the optimal weights can be calculated directly from the measurement model
- no-refit needed after fitting shape parameters or when changing m-ranges
- any uncertainties on N_0, N_1 or shapes can be propagated into $w_s(m)$

sWeights implementations

- via Extended Maximum Likelihood fit of signal and background yields

sWeights implementations

- via Extended Maximum Likelihood fit of signal and background yields
- numerical/analytical calculation of W_{kl}

sWeights implementations

- via Extended Maximum Likelihood fit of signal and background yields
- numerical/analytical calculation of W_{kl}
- plug-in estimate for W_{kl} based on data
  - express the W_{kl} integral as an expectation value over the m-density
  - estimate this expectation value from the data

$$W_{kl} = \int dm \, \frac{d_k(m) \, d_l(m)}{N_0 \, d_0(m) + N_1 \, d_1(m)}$$

sWeights implementations

- via Extended Maximum Likelihood fit of signal and background yields
- numerical/analytical calculation of W_{kl}
- plug-in estimate for W_{kl} based on data
  - express the W_{kl} integral as an expectation value over the m-density
  - estimate this expectation value from the data

$$
\begin{aligned}
W_{kl} &= \int dm \, \frac{d_k(m)\, d_l(m)}{N_0\, d_0(m) + N_1\, d_1(m)} \\
&= N \int dm \, \frac{d_k(m)\, d_l(m)}{(N_0\, d_0(m) + N_1\, d_1(m))^2} \left(\frac{N_0}{N}\, d_0(m) + \frac{N_1}{N}\, d_1(m) \right)
\end{aligned}
$$

sWeights implementations

- via Extended Maximum Likelihood fit of signal and background yields
- numerical/analytical calculation of W_{kl}
- plug-in estimate for W_{kl} based on data
  - express the W_{kl} integral as an expectation value over the m-density
  - estimate this expectation value from the data

$$
\begin{aligned}
W_{kl} &= \int dm \, \frac{d_k(m)\, d_l(m)}{N_0\, d_0(m) + N_1\, d_1(m)} \\
&= N \int dm \, \frac{d_k(m)\, d_l(m)}{(N_0\, d_0(m) + N_1\, d_1(m))^2} \left(\frac{N_0}{N}\, d_0(m) + \frac{N_1}{N}\, d_1(m) \right) \\
&= N \left\langle \frac{d_k(m)\, d_l(m)}{(N_0\, d_0(m) + N_1\, d_1(m))^2} \right\rangle
\end{aligned}
$$

sWeights implementations

- via Extended Maximum Likelihood fit of signal and background yields
- numerical/analytical calculation of W_{kl}
- plug-in estimate for W_{kl} based on data
  - express the W_{kl} integral as an expectation value over the m-density
  - estimate this expectation value from the data

$$W_{kl} = \int dm \, \frac{d_k(m)\, d_l(m)}{N_0\, d_0(m) + N_1\, d_1(m)}$$

$$= N \int dm \, \frac{d_k(m)\, d_l(m)}{(N_0\, d_0(m) + N_1\, d_1(m))^2} \left(\frac{N_0}{N}\, d_0(m) + \frac{N_1}{N}\, d_1(m) \right)$$

$$= N \left\langle \frac{d_k(m)\, d_l(m)}{(N_0\, d_0(m) + N_1\, d_1(m))^2} \right\rangle \rightarrow \sum_i \frac{d_k(m_i)\, d_l(m_i)}{(\hat{N}_0\, d_0(m_i) + \hat{N}_1\, d_1(m_i))^2}$$

sWeights implementations

- via Extended Maximum Likelihood fit of signal and background yields
- numerical/analytical calculation of W_{kl}
- plug-in estimate for W_{kl} based on data
  - express the W_{kl} integral as an expectation value over the m-density
  - estimate this expectation value from the data

$$
\begin{aligned}
W_{kl} &= \int dm \, \frac{d_k(m)\, d_l(m)}{N_0\, d_0(m) + N_1\, d_1(m)} \\
&= N \int dm \, \frac{d_k(m)\, d_l(m)}{(N_0\, d_0(m) + N_1\, d_1(m))^2} \left(\frac{N_0}{N}\, d_0(m) + \frac{N_1}{N}\, d_1(m) \right) \\
&= N \left\langle \frac{d_k(m)\, d_l(m)}{(N_0\, d_0(m) + N_1\, d_1(m))^2} \right\rangle \;\rightarrow\; \sum_i \frac{d_k(m_i)\, d_l(m_i)}{(\hat{N}_0\, d_0(m_i) + \hat{N}_1\, d_1(m_i))^2}
\end{aligned}
$$

  - for finite N this scheme satisfies $\sum w_s = \hat{N}_0$

Factorisation breaking and global weights

❖ data model

$$f(m,t) = N_0\, d_0(m)\, s(t) + N_1\, d_1(m)\, b(m)$$

- two terms that each factorise in m and t; the total is does not factorise

Factorisation breaking and global weights

❖ data model

$$f(m,t) = N_0\, d_0(m)\, s(t) + N_1\, d_1(m)\, b(m)$$

- two terms that each factorise in m and t; the total is does not factorise
- generalisation

$$f(m,t) = \sum_{k=0}^{n} N_k\, d_k(m) g_k(t) = \underbrace{\sum_{k=0}^{m-1} N_k\, d_k(m) g_k(t)}_{\text{signal}} + \underbrace{\sum_{k=m}^{n} N_k\, d_k(m) g_k(t)}_{\text{background}}$$

Factorisation breaking and global weights

❖ data model

$$f(m,t) = N_0\, d_0(m)\, s(t) + N_1\, d_1(m)\, b(m)$$

- two terms that each factorise in m and t; the total is does not factorise
- generalisation

$$f(m,t) = \sum_{k=0}^{n} N_k\, d_k(m) g_k(t) = \underbrace{\sum_{k=0}^{m-1} N_k\, d_k(m) g_k(t)}_{\texttt{signal}} + \underbrace{\sum_{k=m}^{n} N_k\, d_k(m) g_k(t)}_{\texttt{background}}$$

- for $n \to \infty$ any 2-dim function can be represented

Factorisation breaking and global weights

❖ data model

$$f(m,t) = N_0\, d_0(m)\, s(t) + N_1\, d_1(m)\, b(m)$$

- two terms that each factorise in m and t; the total is does not factorise
- generalisation

$$f(m,t) = \sum_{k=0}^{n} N_k\, d_k(m) g_k(t) = \underbrace{\sum_{k=0}^{m-1} N_k\, d_k(m) g_k(t)}_{\text{signal}} + \underbrace{\sum_{k=m}^{n} N_k\, d_k(m) g_k(t)}_{\text{background}}$$

- for $n \to \infty$ any 2-dim function can be represented
- sWeights $w_k(m)$ allows one to project out each component $g_k(t)$

$$w_s(m) = \sum_{k=0}^{m-1} w_k(m) \quad \text{and} \quad w_b(m) = \sum_{k=m}^{n} w_k(m)$$

Factorisation breaking and global weights

❖ data model

$$f(m,t) = N_0\, d_0(m)\, s(t) + N_1\, d_1(m)\, b(m)$$

- two terms that each factorise in m and t; the total is does not factorise
- generalisation

$$f(m,t) = \sum_{k=0}^{n} N_k d_k(m) g_k(t) = \underbrace{\sum_{k=0}^{m-1} N_k d_k(m) g_k(t)}_{\texttt{signal}} + \underbrace{\sum_{k=m}^{n} N_k d_k(m) g_k(t)}_{\texttt{background}}$$

- for $n \to \infty$ any 2-dim function can be represented
- sWeights $w_k(m)$ allows one to project out each component $g_k(t)$

$$w_s(m) = \sum_{k=0}^{m-1} w_k(m) \quad \text{and} \quad w_b(m) = \sum_{k=m}^{n} w_k(m)$$

consider one signal and multiple background components ➜

❖ data model model discussed so far

$$f(m,t) = N_0\, d_0(m)\, s(t) + N_1\, d_1(m)\, b(m)$$

❖ include a global weight function (e.g. efficiency)

$$f(m,t) = \varepsilon(m,t)\left[N_0\, d_0(m)\, s(t) + N_1\, d_1(m)\, b(m)\right]$$

❖ allow for non-factorising background

$$f(m,t) = \varepsilon(m,t)\left[N_0\, d_0(m)\, s(t) + \sum_{k=1}^{n} N_k\, d_k(m)\, b_k(t)\right]$$

❖ data model with weight function & non-factorising background

$$f(m,t) = \varepsilon(m,t)\left[N_0\, d_0(m)\, s(t) + \sum_{k=1}^{n} N_k\, d_k(m)\, b_k(t)\right]$$

❖ data model with weight function & non-factorising background

$$f(m,t) = \varepsilon(m,t)\left[N_0\, d_0(m)\, s(t) + \sum_{k=1}^{n} N_k\, d_k(m)\, b_k(t)\right]$$

- work with $1/\varepsilon(m,t)$ weighted events, i.e. $f(m,t)/\varepsilon(m,t)$
- construct weights $w_s(m)$ that project out $N_0\, s(t)$ from $f(m,t)/\varepsilon(m,t)$

❖ data model with weight function & non-factorising background

$$f(m,t) = \varepsilon(m,t)\left[N_0\, d_0(m)\, s(t) + \sum_{k=1}^{n} N_k\, d_k(m)\, b_k(t)\right]$$

- work with $1/\varepsilon(m,t)$ weighted events, i.e. $f(m,t)/\varepsilon(m,t)$
- construct weights $w_s(m)$ that project out $N_0\, s(t)$ from $f(m,t)/\varepsilon(m,t)$

$$\int dm\, w_s(m)\, \frac{f(m,t)}{\varepsilon(m,t)}$$

❖ data model with weight function & non-factorising background

$$f(m,t) = \varepsilon(m,t)\left[N_0\, d_0(m)\, s(t) + \sum_{k=1}^{n} N_k\, d_k(m)\, b_k(t)\right]$$

- work with $1/\varepsilon(m,t)$ weighted events, i.e. $f(m,t)/\varepsilon(m,t)$
- construct weights $w_s(m)$ that project out $N_0\, s(t)$ from $f(m,t)/\varepsilon(m,t)$

$$\int dm\, w_s(m)\, \frac{f(m,t)}{\varepsilon(m,t)} = N_0\, s(t) \int dm\, w_s(m)\, d_0(m) + \sum_{k=1}^{n} N_k\, b_k(t) \int dm\, w_s(m)\, d_k(m)$$

❖ data model with weight function & non-factorising background

$$f(m,t) = \varepsilon(m,t)\left[N_0\, d_0(m)\, s(t) + \sum_{k=1}^{n} N_k\, d_k(m)\, b_k(t)\right]$$

- work with $1/\varepsilon(m,t)$ weighted events, i.e. $f(m,t)/\varepsilon(m,t)$
- construct weights $w_s(m)$ that project out $N_0\, s(t)$ from $f(m,t)/\varepsilon(m,t)$

$$\int dm\, w_s(m)\, \frac{f(m,t)}{\varepsilon(m,t)} = N_0\, s(t) \int dm\, w_s(m)\, d_0(m) + \sum_{k=1}^{n} N_k b_k(t) \int dm\, w_s(m)\, d_k(m)$$

- orthogonality conditions to retain $N_0\, s(t)$

$$\int dm\, w_s(m)\, d_0(m) = 1 \quad \text{and} \quad \int dm\, w_s(m)\, d_k(m) = 0 \quad \text{for} \quad k \geqslant 1$$

❖ constructing optimal sWeights

- $w_s(m)$ for $1/\varepsilon(m,t)$ weighted data ➜ event-by-event weights $w_s(m)/\varepsilon(m,t)$
- optimality condition $\sum (w_s(m)/\varepsilon(m,t))^2 = \mathrm{min}$

❖ constructing optimal sWeights

- $w_s(m)$ for $1/\varepsilon(m,t)$ weighted data ➜ event-by-event weights $w_s(m)/\varepsilon(m,t)$
- optimality condition $\sum (w_s(m)/\varepsilon(m,t))^2 = \text{min}$

$$\int dm\, dt\, f(m,t) \left(\frac{w_s(m)}{\varepsilon(m,t)}\right)^2 = \text{min}$$

❖ constructing optimal sWeights

- $w_s(m)$ for $1/\varepsilon(m,t)$ weighted data ➜ event-by-event weights $w_s(m)/\varepsilon(m,t)$
- optimality condition $\sum(w_s(m)/\varepsilon(m,t))^2 = \text{min}$

$$\int dm\, dt\, f(m,t) \left(\frac{w_s(m)}{\varepsilon(m,t)}\right)^2 = \text{min} = \int dm\, q(m) w_s^2(m) \quad \text{with} \quad q(m) = \int dt\, \frac{f(m,t)}{\varepsilon^2(m,t)}$$

- $q(m)$: $1/\varepsilon^2(m,t)$-weighted m-distribution

❖ constructing optimal sWeights

- $w_s(m)$ for $1/\varepsilon(m,t)$ weighted data ➜ event-by-event weights $w_s(m)/\varepsilon(m,t)$
- optimality condition $\sum (w_s(m)/\varepsilon(m,t))^2 = \min$

$$\int dm\, dt\, f(m,t) \left(\frac{w_s(m)}{\varepsilon(m,t)}\right)^2 = \min = \int dm\, q(m) w_s^2(m) \quad \text{with} \quad q(m) = \int dt\, \frac{f(m,t)}{\varepsilon^2(m,t)}$$

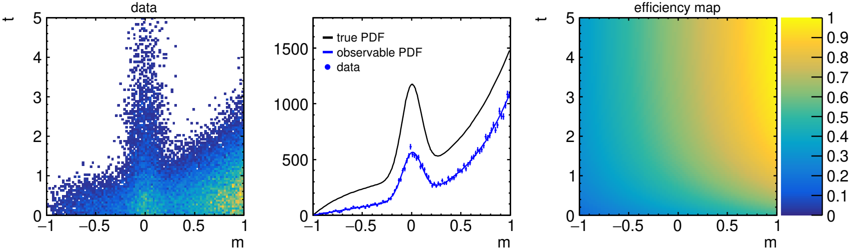
- $q(m)$: $1/\varepsilon^2(m,t)$-weighted m-distribution

■ result of constrained minimisation

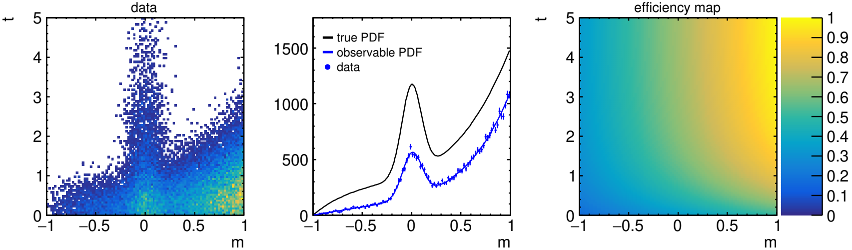
$$w_s(m) = \sum_{k=0}^{n} \frac{\alpha_k\, d_k(m)}{q(m)} \quad \text{with} \quad \sum_{l=0}^{n} W_{kl}\alpha_l = \delta_{k0} \quad \text{and} \quad W_{kl} = \int dm\, \frac{d_k(m)\, d_l(m)}{q(m)}$$

- need only $d_0(m)$ and $q(m)$ – background densities $d_k(m)$ can be chosen freely
- no assumption needed on any density in t

❖ toy with efficiency losses, factorising signal and non-factorising background



❖ toy with efficiency losses, factorising signal and non-factorising background

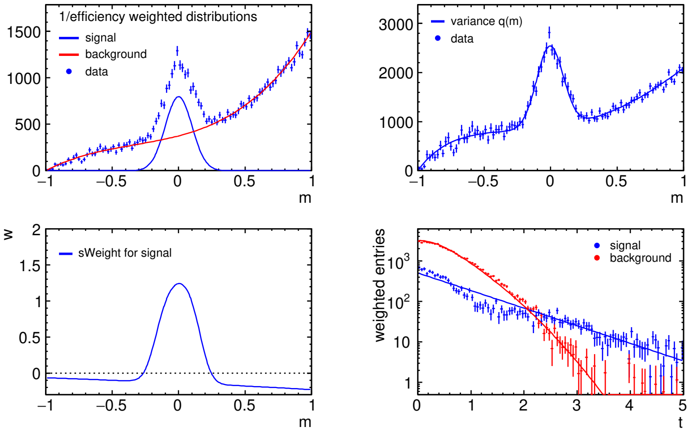


➤ numerical checks

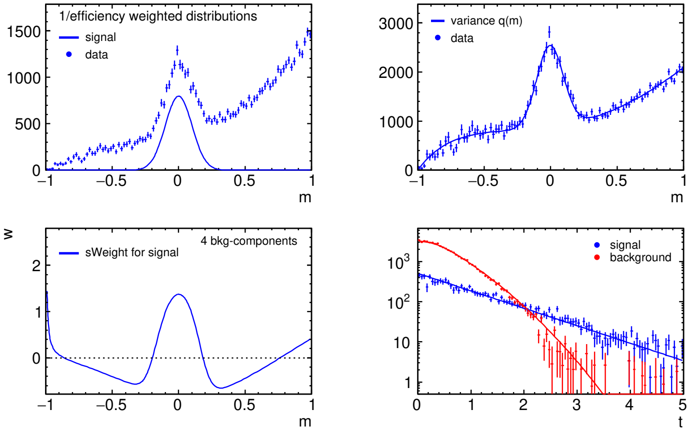
- sWeights method assuming that the background factorises in m and t
- sWeights allowing for 4 generic terms to approximate the background

$$b_1(t) + mb_2(t) + m^2 b_3(t) + m^3 b_4(t)$$

➤ failure for standard sWeights



➤ sWeights with 4 background terms



3 Summary and outlook

sWeights disentangle, with minimum total variance in the yield, signal and background in a control variable; no assumptions about the shape in the control variable are required.

3 Summary and outlook

sWeights disentangle, with minimum total variance in the yield, signal and background in a control variable; no assumptions about the shape in the control variable are required.

- sWeights classic (arXiv:physics/0402083, NIMA 555 (2005) 356)
  - requires factorisation for signal and background component
  - sWeights are a by-product of normalisation fit in the discriminant variable

3 Summary and outlook

sWeights disentangle, with minimum total variance in the yield, signal and background in a control variable; no assumptions about the shape in the control variable are required.

- sWeights classic (arXiv:physics/0402083, NIMA 555 (2005) 356)
  - requires factorisation for signal and background component
  - sWeights are a by-product of normalisation fit in the discriminant variable
- sWeights as orthogonal functions
  - direct sWeight-calculation from densities in the discriminant variable
  - allows for extension to non-factorising weights and PDFs

3 Summary and outlook

sWeights disentangle, with minimum total variance in the yield, signal and background in a control variable; no assumptions about the shape in the control variable are required.

- sWeights classic (arXiv:physics/0402083, NIMA 555 (2005) 356)
  - requires factorisation for signal and background component
  - sWeights are a by-product of normalisation fit in the discriminant variable
- sWeights as orthogonal functions
  - direct sWeight-calculation from densities in the discriminant variable
  - allows for extension to non-factorising weights and PDFs
- upcoming paper by H. Dembinski, C. Langenbruch; M. Kenzie, MS
  - full math and extended examples of the material presented here
  - discussion of use of sWeights in unbinned fits ➜ also next talk by Christoph
  - wanted: a criterion to decide when the data model contains enough terms

3 Summary and outlook

sWeights disentangle, with minimum total variance in the yield, signal and background in a control variable; no assumptions about the shape in the control variable are required.

- sWeights classic (arXiv:physics/0402083, NIMA 555 (2005) 356)
  - requires factorisation for signal and background component
  - sWeights are a by-product of normalisation fit in the discriminant variable
- sWeights as orthogonal functions
  - direct sWeight-calculation from densities in the discriminant variable
  - allows for extension to non-factorising weights and PDFs
- upcoming paper by H. Dembinski, C. Langenbruch; M. Kenzie, MS
  - full math and extended examples of the material presented here
  - discussion of use of sWeights in unbinned fits ➜ also next talk by Christoph
  - wanted: a criterion to decide when the data model contains enough terms

❖ sWeights are orthogonal functions – not probabilities!

backup

test functions used for the toy studies

- signal and background densities for $m \in [-1, 1]$ and $t \in [0, 5]$

$$s(m,t) \propto \exp\left(-\frac{m^2}{2\sigma_0^2} - t\right) \quad \text{and} \quad b(m,t) \propto (1 + m + m^2 + m^3)\exp\left(-\frac{t^2}{2\sigma_1^2}\right)$$

- parameters for factorising densities

$$\sigma_0 = \frac{1}{10} \quad \text{and} \quad \sigma_1 = \frac{1}{2}$$

- parameters for non factorising densities

$$\sigma_1 = \frac{1}{2 - m}$$

- efficiency function

$$\varepsilon(m,t) = \frac{(2+m)(2-\exp(-t))}{6}$$