

7 COWs and sPlot

over 39 years of history in particle physics . . .

❖ goal: disentangle signal and background in order to measure signal properties

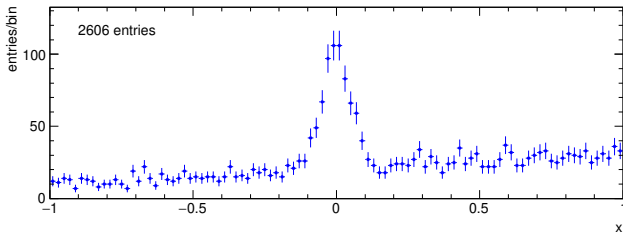
- first described: R. Barlow - J.Comp.Phys. 72 (1987) 202
- rediscovered as sWeights and sPlot: M. Pivk, F. LeDiberder - NIM A555 (2005) 356
- COWs: H. Dembinski, M. Kenzie, C. Langenbruch, MS - NIM A1040 (2022) 167270

➤ story line, illustrated by a toy model, . . .

- ▶ fitting of normalization parameters
- ▶ inclusion of shape parameters
- ▶ the sPlot method for background subtraction
- ▶ the sFit method

Fitting normalization parameters

❖ toy data: gaussian signal with $\mu = 0$ and $\sigma = 1/20$ on linear background ...



data model for IIDP measurements

$$f(x) = \langle N_0 \rangle g_0(x) + \langle N_1 \rangle g_1(x) + \langle N_2 \rangle g_2(x) \quad \text{with}$$

$$g_0(x) = \frac{20}{\sqrt{2\pi}} e^{-200x^2} \quad g_1(x) = \frac{1-x}{2} \quad g_2(x) = \frac{1+x}{2}$$

❖ fit model: $F(x) = N_0 g_0(x) + N_1 g_1(x) + N_2 g_2(x)$

■ generic Least Squares fit with arbitrary “variance function” $I(x) > 0$

$$S^2 = \sum_{\text{bins}} \frac{1}{I_{\text{bin}}} \left(n_{\text{bin}} - \int_{\text{bin}} dx F(x) \right)^2 \quad \text{with} \quad I_{\text{bin}} = \int_{\text{bin}} dx I(x)$$

■ stationary point (minimum) of S^2

$$\frac{\partial S^2}{\partial N_m} = \sum_{\text{bins}} \frac{2}{I_{\text{bin}}} \left(n_{\text{bin}} - \int_{\text{bin}} dx F(x) \right) \left(- \int_{\text{bin}} dx g_m(x) \right) = 0$$

■ infinitesimal bins sizes $\Delta x \rightarrow 0$

$$\sum_{\text{bins}} n_{\text{bin}} \frac{g_m(x_{\text{bin}})}{I(x_{\text{bin}})} = \sum_k N_k \sum_{\text{bins}} \frac{g_k(x_{\text{bin}}) g_m(x_{\text{bin}}) \Delta x}{I(x_{\text{bin}})}$$

■ using the fact that $n_i \in \{0, 1\}$ and replacing the Riemann sum by its integral

$$\sum_{i=1}^N \frac{g_m(x_i)}{I(x_i)} = \sum_k N_k \int dx \frac{g_k(x) g_m(x)}{I(x)} = \sum_k N_k A_{km}^{-1}$$

- solving the linear equation for N_k

$$N_k = \sum_{i=1}^N w_k(x_i) \quad \text{with} \quad w_k(x) = \sum_m A_{km} \frac{g_m(x)}{I(x)} \quad \text{and} \quad A_{kl}^{-1} = \int dx \frac{g_k(x) g_l(x)}{I(x)}$$

■ fit parameters are obtained summing weight functions $w_k(x_i)$ for the measurements x_i

→ **COWs**: Custom Orthogonal Weight functions

- key property: orthonormality between COWS $w_k(x)$ and basis function $g_l(x)$

$$\begin{aligned} \int dx w_k(x) g_l(x) &= \int dx \left(\sum_m A_{km} \frac{g_m(x)}{I(x)} \right) g_l(x) \\ &= \sum_m A_{km} \int dx \frac{g_m(x) g_l(x)}{I(x)} = \sum_k A_{km} A_{ml}^{-1} = \delta_{kl} \end{aligned}$$

❖ covariance matrix of sums with Poisson distributed sample size N

$$C_{kl} = \langle N_k N_l \rangle - \langle N_k \rangle \langle N_l \rangle \quad \text{for sums} \quad N_k = \sum_{i=1}^N w_k(x_i)$$

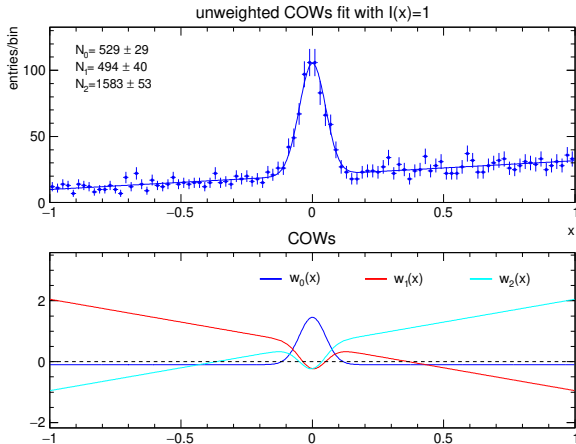
With the probability P_N for a sample size N one finds

$$\langle N_k \rangle = \sum_{N=0}^{\infty} P_N \left\langle \sum_{i=1}^N w_k(x_i) \right\rangle = \sum_{N=0}^{\infty} P_N N \langle w_k \rangle$$

$$\langle N_k N_l \rangle = \sum_{N=0}^{\infty} P_N \left\langle \sum_{i,j=1}^N w_k(x_i) w_l(x_j) \right\rangle = \sum_{N=0}^{\infty} P_N \left(N(N-1) \langle w_k \rangle \langle w_l \rangle + N \langle w_k w_l \rangle \right)$$

and thus for Poisson probabilities with $\sum P_N N = \langle N \rangle$ and $\sum P_N N(N-1) = \langle N \rangle^2$:

$$C_{kl} = \langle N_k N_l \rangle - \langle N_k \rangle \langle N_l \rangle = \langle N \rangle \langle w_k w_l \rangle \quad \text{plugin estimate} \quad \hat{C}_{kl} = \sum_{i=1}^N w_k(x_i) w_l(x_i)$$



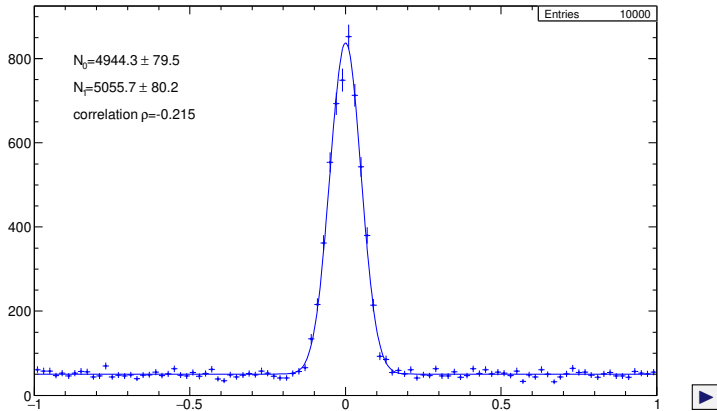
problems and exercises

Assume a gaussian peak with standard deviation $\sigma = 0.05$ on uniform background. The signal fraction is $s \in [0, 1]$, the total range considered is $x \in [-1, +1]$. With a variance function $I(x) = 1$ the COWs are

$$w_0 = \frac{1}{1 - \sqrt{\pi}\sigma} \left(\sqrt{2} e^{-x^2/2\sigma^2} - \sqrt{\pi}\sigma \right) \quad \text{and}$$
$$w_1 = \frac{1}{1 - \sqrt{\pi}\sigma} \left(1 - \sqrt{2} e^{-x^2/2\sigma^2} \right)$$

Select a signal fraction s , generate N events distributed according to the data model described above, and use the COWs to estimate the actual yields for the two components and their covariance matrix.

► solution



Maximum Likelihood

covariance matrix of fit parameters N_k for IIDP data with sample size N

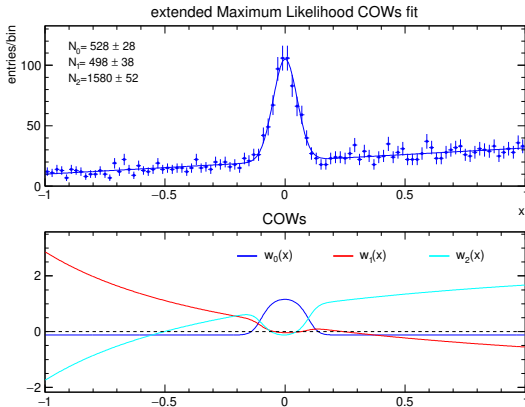
$$\begin{aligned} C_{kl} &= \langle N \rangle \langle w_k w_l \rangle = \langle N \rangle \int dx f(x) \left(\sum_i A_{ki} \frac{g_i(x)}{I(x)} \right) \left(\sum_j A_{lj} \frac{g_j(x)}{I(x)} \right) \\ &= \langle N \rangle \sum_{i,j} A_{ki} A_{lj} \int dx f(x) \frac{g_i(x) g_j(x)}{I^2(x)} \quad \text{with} \quad A_{ij}^{-1} = \int dx \frac{g_i(x) g_j(x)}{I(x)} \end{aligned}$$

Cramer-Rao bound: per event variance $\geq$ inverse of the Fisher information

$$F_{kl} = \int dx \frac{g_k(x) g_l(x)}{f(x)} \stackrel{I(x)=f(x)}{=} A_{kl}^{-1}$$

- ▶ $I(x) = f(x)$ yields $C_{kl} = \langle N \rangle A_{kl} = \langle N \rangle F_{kl}^{-1}$, the minimum possible variance
- ▶ parameter estimates are the EML estimates

❖ implementation: iteration of the variance functions $I(x) = f(x)$



$$I^{(0)}(x) = 1$$

$$N_k^{(n)} = \sum_{i=1}^N w_k^{(n)}(x_i)$$

$$I^{(n+1)}(x) = \sum_k N_k^{(n)} g_k(x)$$

❖ constructing optimal weight functions to fit $f(x) = \sum N_k g_k(x)$

$$\text{COWs: } w_k(x) = \sum_l A_{kl} \frac{g_l(x)}{f(x)} \quad \text{with} \quad A_{kl}^{-1} = \int dx \frac{g_k(x) g_l(x)}{f(x)}$$

$$\text{sWeights: } s_k(x) = \sum_l B_{kl} \frac{g_l(x)}{f(x)} \quad \text{with} \quad B_{kl}^{-1} = \sum_{i=1}^N \frac{g_k(x_i) g_l(x_i)}{f^2(x_i)}$$

■ by the LLN B^{-1} is a sample estimate for A^{-1}

■ asymptotically B^{-1} and A^{-1} are identical

■ at finite statistics $s_k(x)$ is only approximately orthonormal to $g_l(x)$

consequences →

❖ fit parameters

$$\sum_{i=1}^N s_k(x_i) = \sum_{i=1}^N w_k(x_i) = N_k$$

- ▶ identical best-fit parameters

❖ covariance matrices

$$\hat{C}^{(B)} = \sum_{i=1}^N s(x_i) s(x_i)^T = B \quad \text{and} \quad \hat{C}^{(A)} = \sum_{i=1}^N w(x_i) w(x_i)^T = A B^{-1} A$$

- ▶ different estimates for the same true covariance matrix

❖ application COWs/sWeights in the sPlot method

- ▶ COWs have uncorrelated bins in the control variable
- ▶ for sWeights correlations are present (see NIM A1040 (2022) 167270)

❖ extended toy model with signal mean and leading order expansion

$$\begin{aligned} f(x) &= N_0 g_0(x, \mu) + N_1 g_1(x) + N_2 g_2(x) \\ &\approx N_0 \left(g_0(x, \hat{\mu}) + \frac{\partial g_0(x, \hat{\mu})}{\partial \hat{\mu}} (\mu - \hat{\mu}) \right) + N_1 g_1(x) + N_2 g_2(x) \\ &= N_0 g_0(x, \hat{\mu}) + N_1 g_1(x) + N_2 g_2(x) + N_3 \frac{\partial g_0(x, \hat{\mu})}{\partial \hat{\mu}} \quad \text{with} \quad N_3 = N_0 (\mu - \hat{\mu}) \end{aligned}$$

► $\hat{\mu}$ initial estimate for the shape parameter

■ fit parameters basis functions for the COWs formalism

$$N_k \quad \text{and} \quad g_k(x) = \frac{\partial f(x)}{\partial N_k} \quad \text{with} \quad k = 0, \dots, 3$$

■ $\{N_0, N_1, N_2\}$ normalization estimates

■ N_3 correction to shape parameter: improved estimate $\hat{\mu}^{(m+1)} = \hat{\mu}^{(m)} + N_3/N_0$

The general COWs formalism

- ❖ fitting a data model with normalization- and/or shape parameters

$$f(x, a) \quad \text{with parameters} \quad a = \{a_0, a_1, a_2, \dots, a_n\}$$

- construction of COWs

$$w_k(x) = \sum_l \frac{A_{kl}}{I(x)} \frac{\partial f(x, a)}{\partial a_l} \quad \text{with} \quad A_{kl}^{-1} = \int dx \frac{1}{I(x)} \frac{\partial f(x, a)}{\partial a_k} \frac{\partial f(x, a)}{\partial a_l}$$

- ▶ $I(x) = f(x)$: A^{-1} is the Fisher information matrix and COWs do an EML fit

- estimates of normalization parameters and corrections to shape parameters

$$a_k = \sum_i w_k(x_i) \quad \text{if } a_k \text{ is the normalization of } g_k(x)$$

$$\Delta a_l = \sum_i \frac{w_l(x_i)}{a_k} \quad \text{if } a_l \text{ is a shape parameter of } g_k(x)$$

- ▶ iteration until all parameters are stable: $a_k = \text{const}$ and $\Delta a_l = 0$

❖ error estimates

$$C_{kk} = \sum_i w_k(x_i)^2 \quad \text{and} \quad C_{ll} = \sum_i \frac{w_l(x_i)^2}{a_k^2}$$

- variance of shape parameters is given by the variance of the correction term
- the variance of the correction term follows from an extended bootstrap

$$\Delta a_l = \frac{s_l}{a_k} = \frac{\sum_i w_l(x_i)}{\sum_i w_k(x_i)} \quad \text{bootstrap variations} \quad \Delta a_l = \frac{\sum_i n_i w_l(x_i)}{\sum_i n_i w_k(x_i)}$$

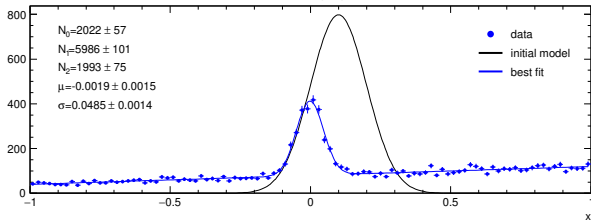
- error propagation with $\sigma^2(n_i) = 1$, independent n_i and derivatives taken at $n_i = 1$

$$\sigma^2(\Delta a_l) = \sum_i \left(\frac{\partial \Delta a_l}{\partial n_i} \right)^2 = \sum_i \left(\frac{w_l(x_i)}{a_k} - \frac{s_l}{a_k} \frac{w_k(i)}{a_k} \right)^2 = \sum_i \frac{w_l(x_i)^2}{a_k^2}$$

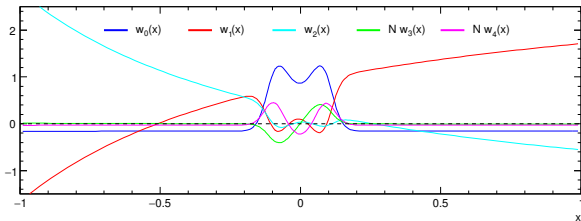
- the last step follows because convergence implies $s_l/a_k = \Delta a_l = 0$

application to toy model with 2 shape and 3 normalization parameters →

COWs fit with shape parameters



COWs



Wrap-up

- COWs are orthonormal functions with respect to the basis $\partial f(x, a)/\partial a_k$
- fit parameters are estimated by (iterated) sums over the measurements
- a variance function $I(x) \propto f(x)$ yields Maximum Likelihood estimators

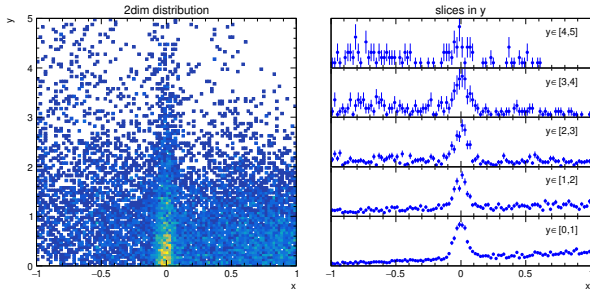
❖ correspondence between fitting and differential geometry

fitting	differential geometry
$f(x, a)$	Riemannian manifold $\mathcal{R}$
$\partial f(x, a)/\partial a_k$	covariant basis of tangent space
A_{kl}	metric on $\mathcal{R}$
COWs	contravariant basis

$\mathcal{R}$ is flat if $f(x, a)$ has no shape parameters and $I(x)$ does not depend on a

The sPlot method

❖ extension of the toy model to 2 dimensions



$$f(x, y) = N_0 g_0(x) h_0(y) + N_1 g_1(x) h_1(y) + N_2 g_2(x) h_2(y)$$

$$\text{with } h_0(y) = e^{-y}, \quad h_1(y) = \sqrt{\frac{2}{\pi}} e^{-y^2/2}, \quad h_2(y) = y e^{-y} \quad \text{for } y > 0$$

❖ estimate the densities in y

- x : “discriminant” variable $\rightarrow$ the PDFs $g_k(x)$ are known
- y : “control” variable $\rightarrow$ the PDFs $h_k(y)$ shall be determined

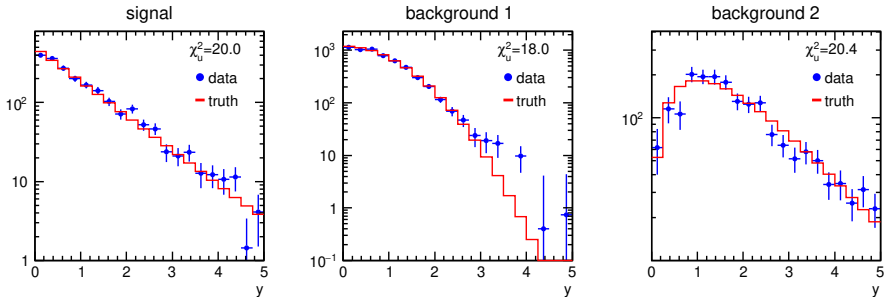
❖ the sPlot method

- the data model in each slice Δy of the 2-dim PDF is the same
- the yields in a given bin are

$$n_k(\Delta y) = N_k \int dy h_k(y)$$

- ▶ when integrating over y , the sum of the COWs estimates the total yields N_k
- ▶ when integrating over Δy , the sum of the COWs estimates the yields n_k
- ▶ given (x_i, y_i) and histogramming y_i with a weight $w_k(x_i)$ estimates $h_k(y)$

❖ result for the 2-dim toy model with 20 bins in y



- uncertainties are estimated by summing the weights in quadrature
- works for COWs and sWeights
- requires that the data model in the discriminant variable is correct

Mismodelling the PDF in the discriminant variable x

❖ assume in x a correct background model and a wrong signal model $g_0(x)$

■ actual PDF

$$f(x) = N_0(g_0(x) + \delta(x)) + N_1 g_1(x) + N_2 g_2(x)$$

■ orthonormality of the COWS implies

$$\int dx w_0(x) f(x) = N_0(1 + \Delta_0) \quad \text{with} \quad \Delta_0 = \int dx w_0(x) \delta(x)$$

$$\int dx w_1(x) f(x) = N_1 + N_0 \Delta_1 \quad \text{with} \quad \Delta_1 = \int dx w_1(x) \delta(x)$$

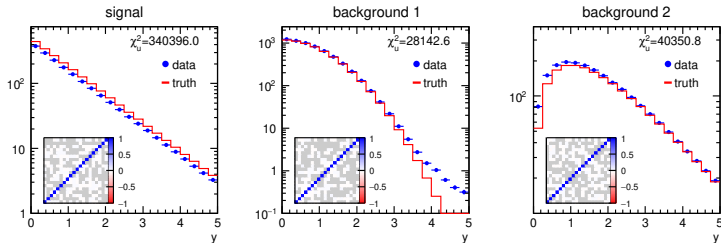
$$\int dx w_2(x) f(x) = N_2 + N_0 \Delta_2 \quad \text{with} \quad \Delta_2 = \int dx w_2(x) \delta(x)$$

► signal yield is scaled, background gets contaminated by signal

test the effect on sPlot results →

❖ fixed wrong shape parameters of the signal model

■ average of 10^4 data sets with and expected number of 10^4 events each



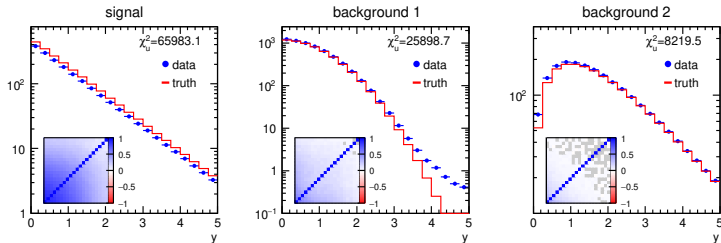
■ background successfully removed from data

■ part of the data feeds into the background distributions

- ▶ signal shape in y preserved, backshapes in y distorted
- ▶ bins in y are **biased but uncorrelated**

❖ fluctuating shape parameters of the signal model

■ average of 10^4 data sets with an expected number of 10^4 events each



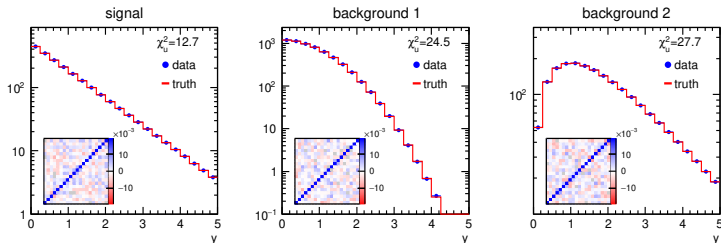
■ background successfully removed from data

■ part of the data feeds into the background distributions

- ▶ signal shape in y preserved, background shapes in y are distorted
- ▶ bins in y are **biased and correlated**

implications →

- sWeights are constructed by a normalization-only EML fit
 - ▶ shape parameters are fixed by an independent fit
 - ▶ fitted parameters are random variables that fluctuate around the true values
 - ▶ correlations in y result from the cross-feed between signal and background
- COWs that account also for shape parameters avoid these problems
 - ▶ average of 10^5 data sets with an expected number of 10^4 events each



explanation →

► comparison of sWeights and COWs

- sWeights estimate of $h_0(y)$: sPlot with weight function $w_0(x)$

$$w_0(x) \text{ orthogonal to } g_1(x) \text{ and } g_2(x)$$

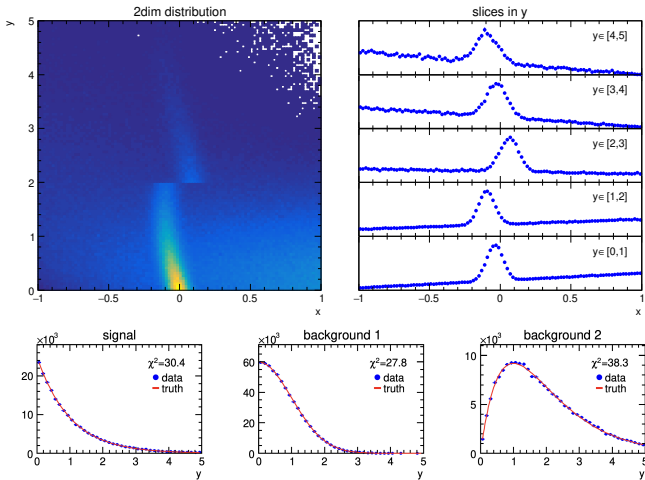
- COWs estimate of $h_0(y)$: sPlot with a weight function $w_0(x)$

$$w_0(x) \text{ orthogonal to } g_1(x) \text{ and } g_2(x) \text{ and } \frac{\partial g_0(x, \mu)}{\partial \mu} \text{ and } \dots$$

- $w_0(x)$ is also orthogonal to the difference between true and assumed signal

$$g_0(x, \mu^{\text{true}}) - g_0(x, \mu) = \Delta_\mu \frac{\partial g_0(x, \mu)}{\partial \mu} + O(\Delta_\mu^2) \quad \text{with} \quad \Delta_\mu = \mu^{\text{true}} - \mu$$

- correlations are avoided
- the difference may vary as a function of $y \rightarrow$ factorization breaking is allowed
- higher order derivatives allow to handle also strong factorization breaking ...



❖ unbinned fitting of parameters ρ in the control variable PDF $h(y, \rho)$

■ classic:

$$\frac{\partial}{\partial \rho} \sum_i w_h(x_i) \ln h(y_i, \rho) = 0 \quad \text{with sWeights} \quad w_h(x_i)$$

- ▶ proposed by Y. Xie, [arXiv:0905.0724](#)
- ▶ unbiased parameter estimates
- ▶ naive variance estimates $(\partial^2 L / \partial \rho^2)^{-1}$ are wrong
- ▶ uncertainties can be determined by bootstrap or sandwich estimator
- ▶ proper error calculation → C. Langenbruch, [arXiv:1911.01303](#)

■ alternative: unbinned (ML) fit by COWs

example →

❖ fitting an exponential PDF to COWeighted data

$$h(y, \rho) = \frac{e^{-y/\rho}}{\rho} \quad \text{basis functions} \quad b_0(y) = h(y, \rho) \quad , \quad b_\rho(y) = \frac{\partial h(y, \rho)}{\partial \rho}$$

COWs: $w_0(y)$ and $w_\rho(y)$ with $I(y) = h(y, \rho)$

■ iteration of COWeighted averages until $\Delta\rho = 0$

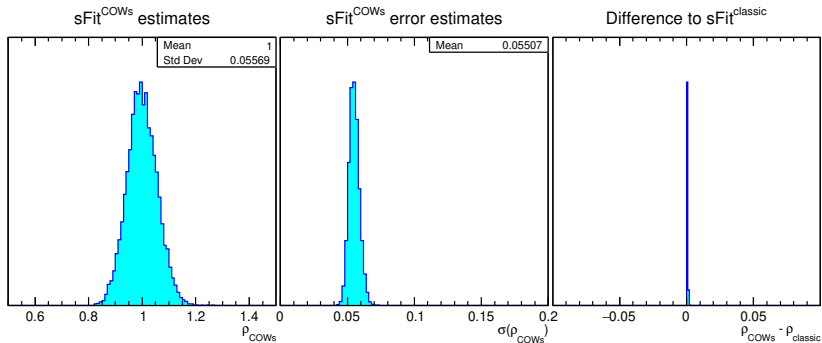
$$\rho \rightarrow \rho + \Delta\rho \quad \text{with} \quad \Delta\rho = \frac{\sum_i w_h(x_i) w_\rho(y_i)}{\sum_i w_h(x_i)}$$

- ▶ $w_h(x)$ selects the background subtracted signal in y
- ▶ $w_\rho(y)$ determines the shape parameter

■ variance of result (bootstrap estimate as for other shape parameters)

$$\sigma^2(\rho) = \frac{\sum_i (w_h(x_i) w_\rho(y_i))^2}{(\sum_i w_h(x_i))^2}$$

❖ numerical check: toy model fit $h(y, \rho) \sim e^{-y/\rho}$ with $\rho^{\text{true}} = 1$



fit results from 10 000 toy data sets

Summary

- **COWs**: orthonormal functions on the function space spanned by a data model $f(x, a)$

$$w_k(x) = \sum_l A_{kl} \frac{g_l(x)}{I(x)} \quad \text{with} \quad A_{kl}^{-1} = \int dx \frac{g_k(x) g_l(x)}{I(x)} \quad \text{and} \quad g_k(x) = \frac{\partial f(x, a)}{\partial a_k}$$

- ▶ summing $w_k(x)$ estimates normalization parameters a_k
 - ▶ summing $w_l(x)$ estimates corrections to shape parameters a_l
 - ▶ a variance function $I(x) \propto f(x, a)$ yields Maximum Likelihood estimators
- **sPlot** estimates distributions in a control variable by histogramming $w_k(x)$
 - ▶ bins are uncorrelated when the COWs account for all free parameters in x
 - **sFit** does a COWs fit on distributions in a control variable
 - ▶ equivalent to a weighted ML for variance functions proportional to the fitted density
 - for IIDP measurements variances are obtained by summing weights in quadrature