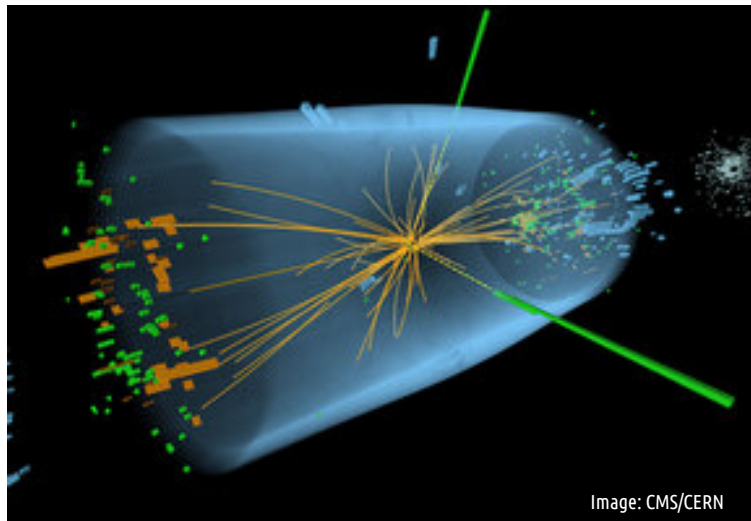


RECONSTRUCTING DISTRIBUTIONS FROM MONTE-CARLO SIMULATIONS USING ORTHOGONAL BASIS FUNCTIONS

Kirill Melnikov, Ivan Novikov, Ivan Pedron

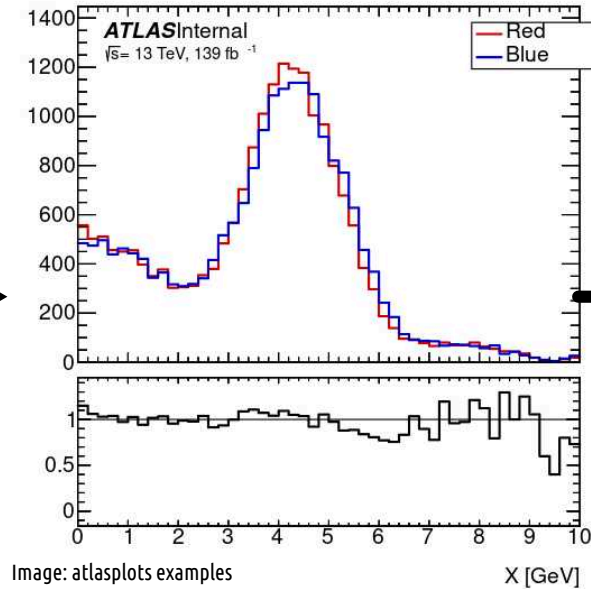
based on [arXiv:2601.10420](https://arxiv.org/abs/2601.10420)

Events



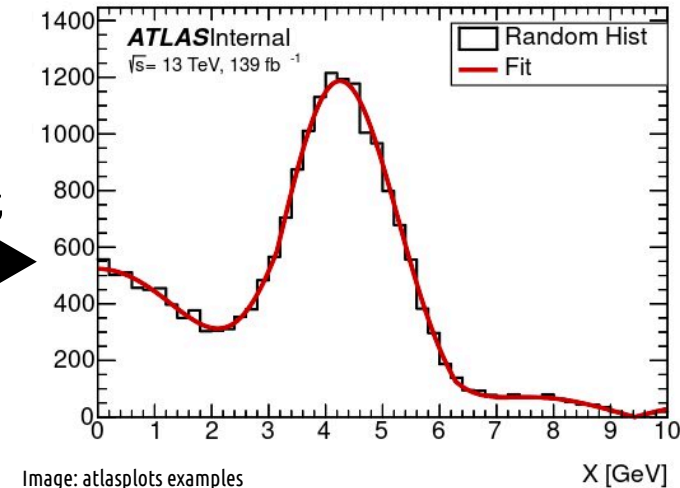
bin
→

Histogram



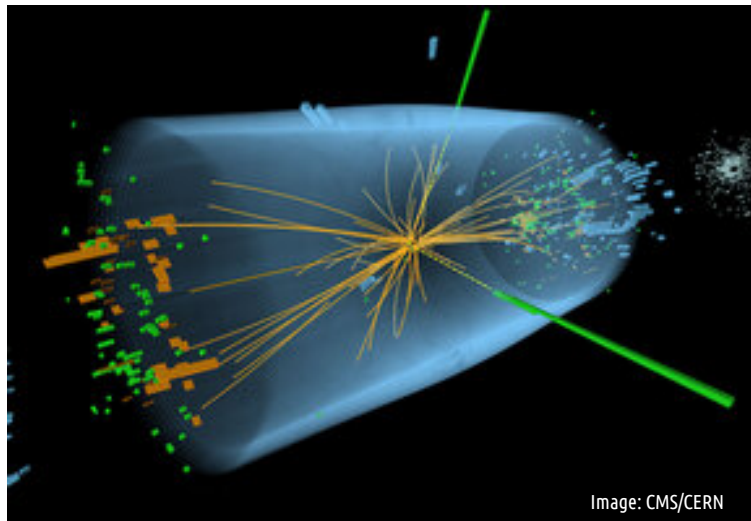
fit
→

Density function



To extract a distribution from events, we collect them in a histogram and fit it with a curve

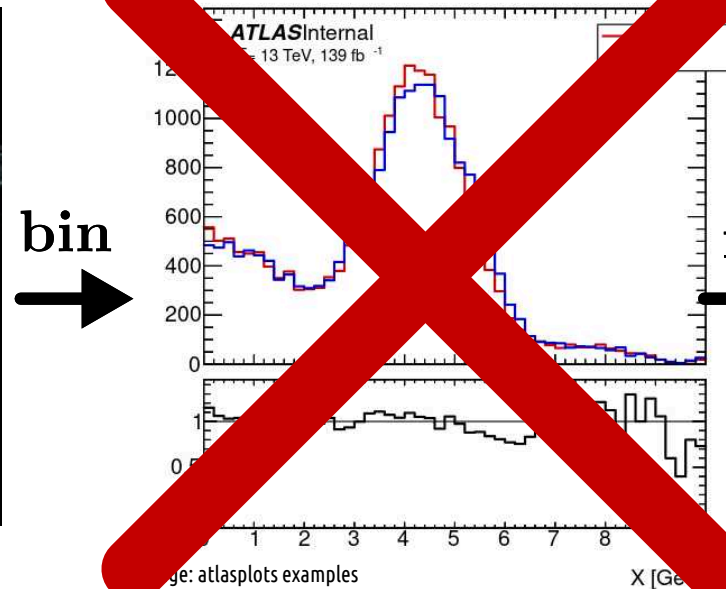
Events



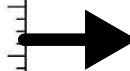
bin



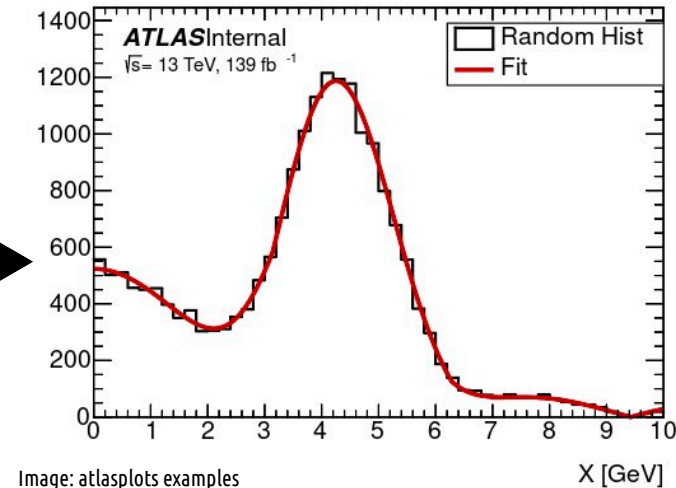
Histogram



fit



Density function



There's a way to bypass the histogram and recover the density function directly from events

$$\langle e_i, e_j \rangle = \delta_{ij}, \quad \sum_{k=0}^{\infty} e_k(x) e_k(y) = \delta(x - y), \implies f(x) = \sum_k \langle f, e_k \rangle e_k(x)$$

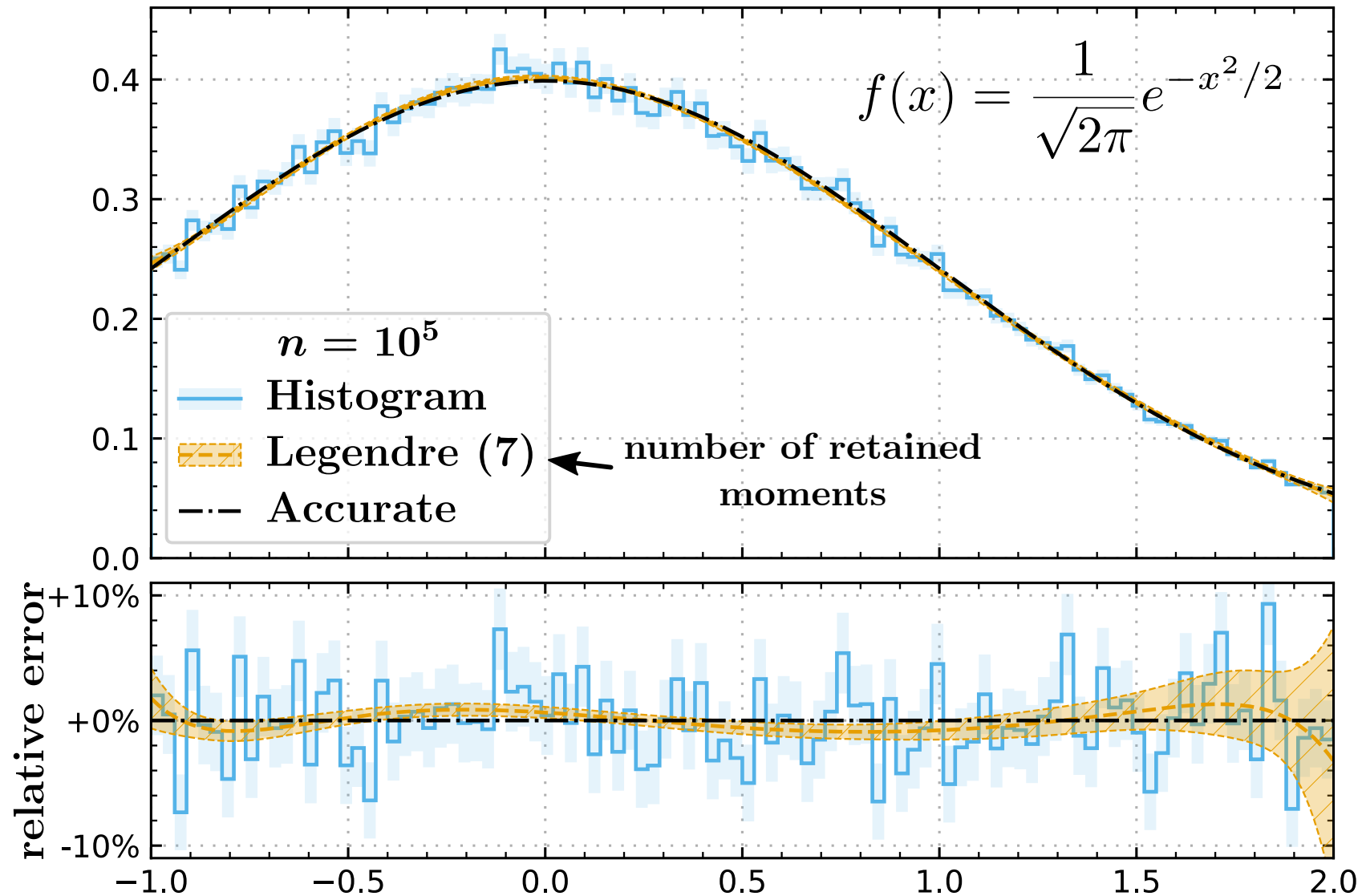
$\uparrow$
 orthonormal basis functions
 (e.g., Legendre polynomials)

$\downarrow$
 observable
 (e.g., transverse momentum)

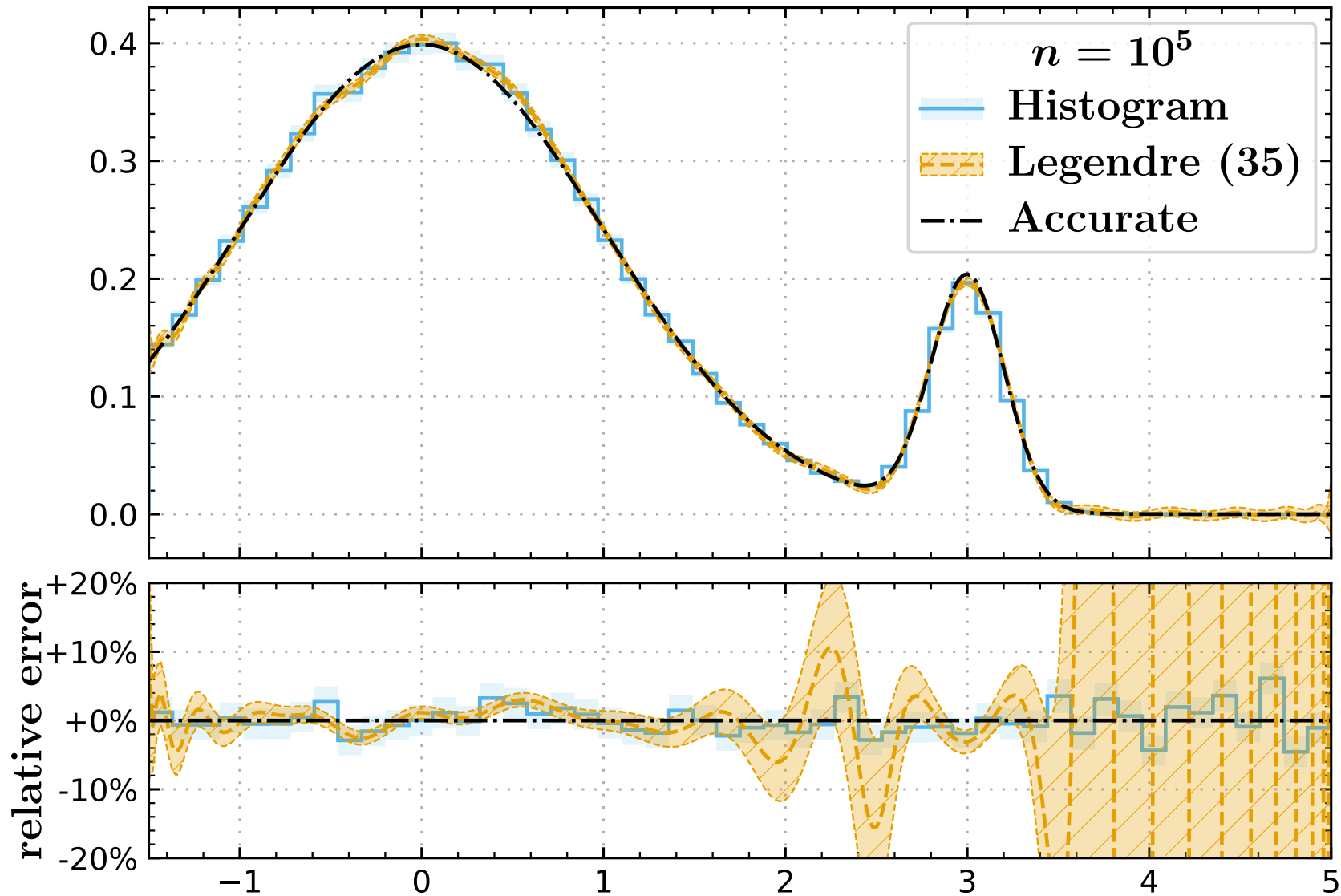
$$\langle f, e_k \rangle = \int \underbrace{\frac{d\sigma}{dx}(x)}_{=f} e_k(x) dx = \underbrace{\int_{\Phi} \frac{d\sigma}{d\phi}(\phi) e_k(x(\phi)) d\phi}_{\text{integral over phasespace}} = \mathbb{E} \left[\underbrace{\frac{1}{p(\phi)} \frac{d\sigma}{d\phi}(\phi) e_k(x(\phi))}_{\text{sampling distribution}} \right]_{\phi \sim p}$$

$$\mathbb{E}[x] \approx \frac{1}{n} \sum_i x_i \pm \sqrt{\frac{1}{n-1} \left[\frac{1}{n} \sum_i x_i^2 - \left(\frac{1}{n} \sum_i x_i \right)^2 \right]}$$

Decompose the target distribution function in an orthonormal basis of functions e_k and compute the coefficients = moments $\langle f, e_k \rangle$ in Monte-Carlo



This works well for distributions without any high-frequency features



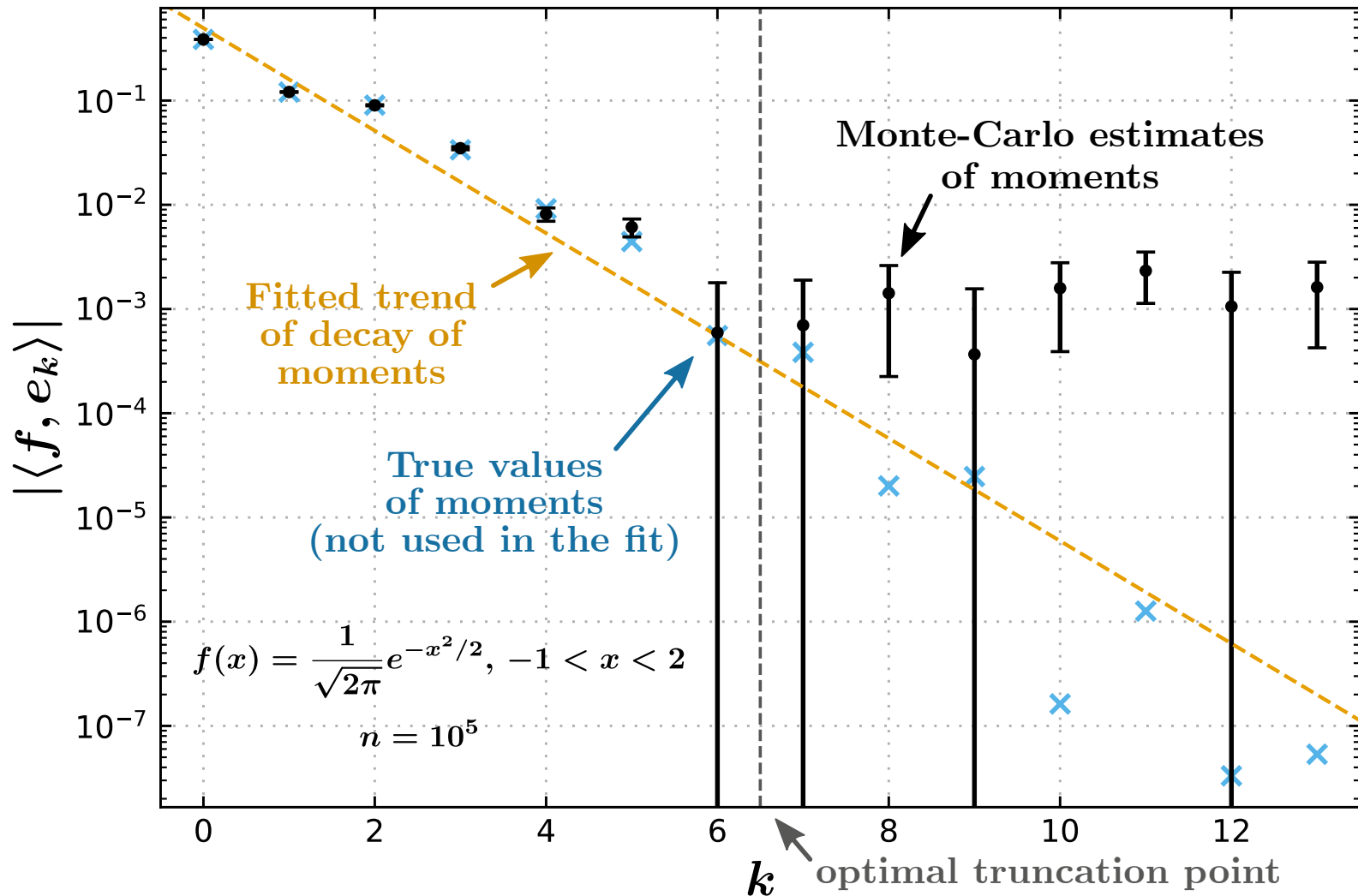
The convergence is slower for distributions with sharper peaks

Monte-Carlo error
in coefficient k

$$f_{\text{reconstructed}}(x) - f(x) = \sum_{k=0}^m (\langle f, e_k \rangle + \delta c_k) e_k(x) - \sum_{k=0}^{\infty} \langle f, e_k \rangle e_k(x)$$

$$= \underbrace{\sum_{k=0}^m \delta c_k e_k(x)}_{\text{Monte-Carlo error}} - \underbrace{\sum_{k=m+1}^{\infty} \langle f, e_k \rangle e_k(x)}_{\text{Truncation error}} \sim \sqrt{\sum_{k=0}^m \delta c_k^2 + \sum_{k=m+1}^{\infty} \langle f, e_k \rangle^2}$$

Including more moments decreases the truncation error but increases the Monte-Carlo error.
Choose the truncation point m to minimize the total uncertainty.

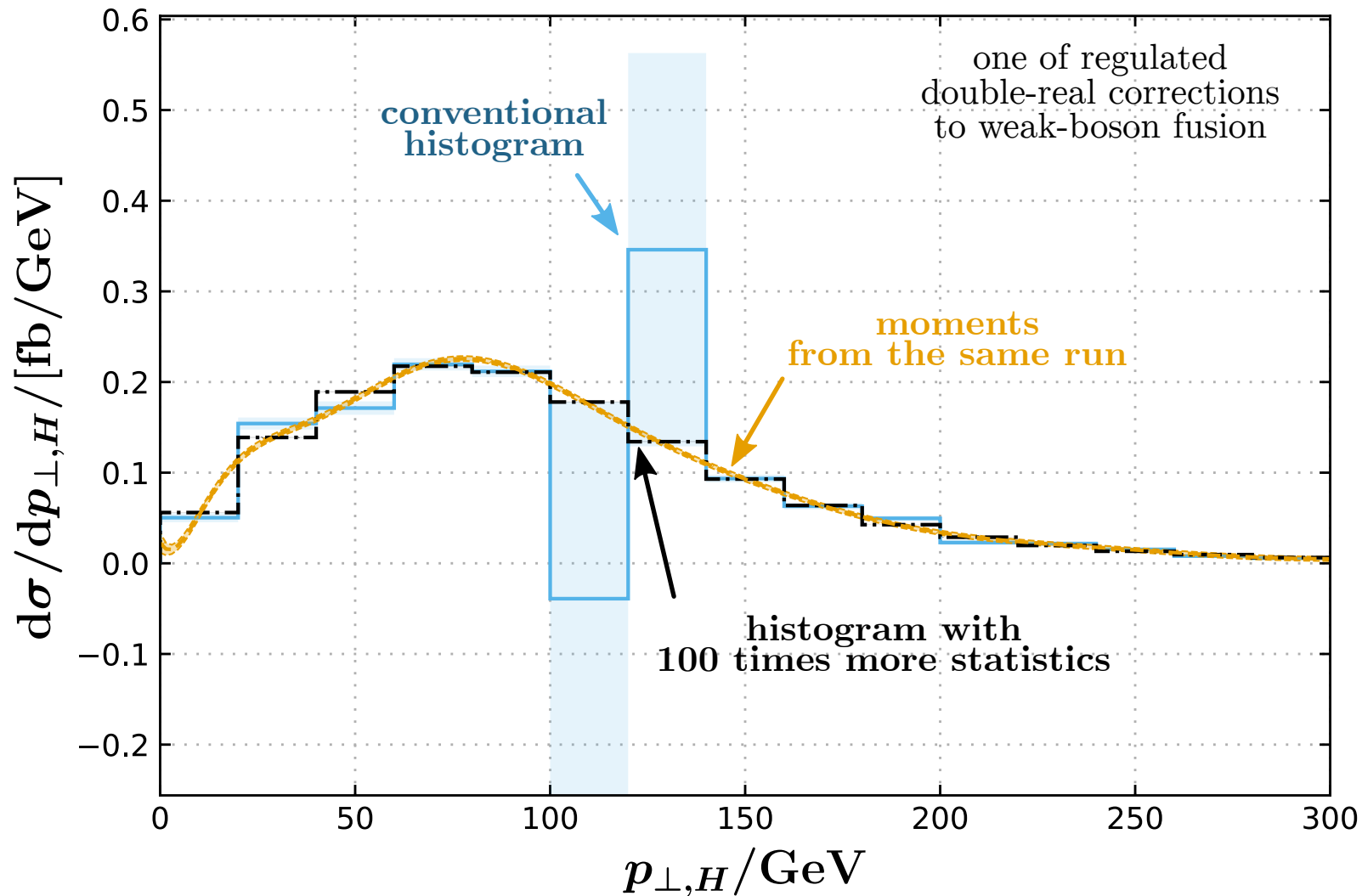


Moments of well-behaved functions decay exponentially.
 To estimate the truncation error, extrapolate this trend with a line.

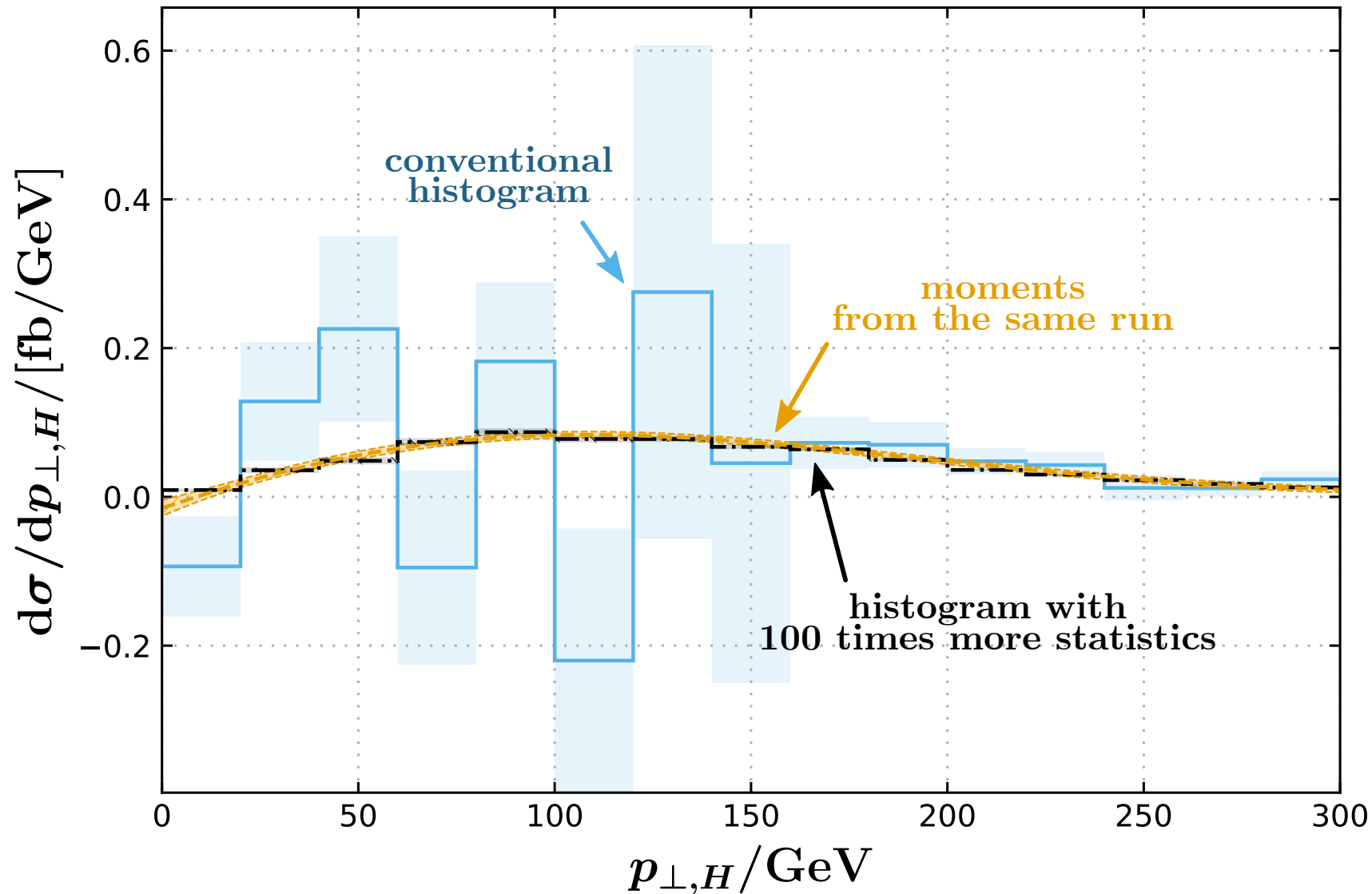
$$\begin{aligned}
 & \int \left| \text{Diagram} \right|^2 d\phi_{\text{Born}} d\phi_m \quad \leftarrow \text{Divergent for } d=4 \\
 & \quad \downarrow \text{Integrate analytically in } d = 4 - 2\epsilon \\
 & = \int \left| \text{Diagram} \right|^2 \otimes \left[\int \left| \text{Diagram} \right|^2 d\phi_m \right] d\phi_{\text{Born}} \quad \text{Finite for } d=4 \\
 & \quad \downarrow \\
 & + \int \left[\underbrace{\left| \text{Diagram} \right|^2}_{\text{Event}} - \underbrace{\left| \text{Diagram} \right|^2 \otimes \left| \text{Diagram} \right|^2}_{\text{Counterevent}} \right] d\phi_{\text{Born}} d\phi_m
 \end{aligned}$$

In calculations with local subtractions a real-emission event must be summed together with counterevents that cancel soft and collinear singularities.

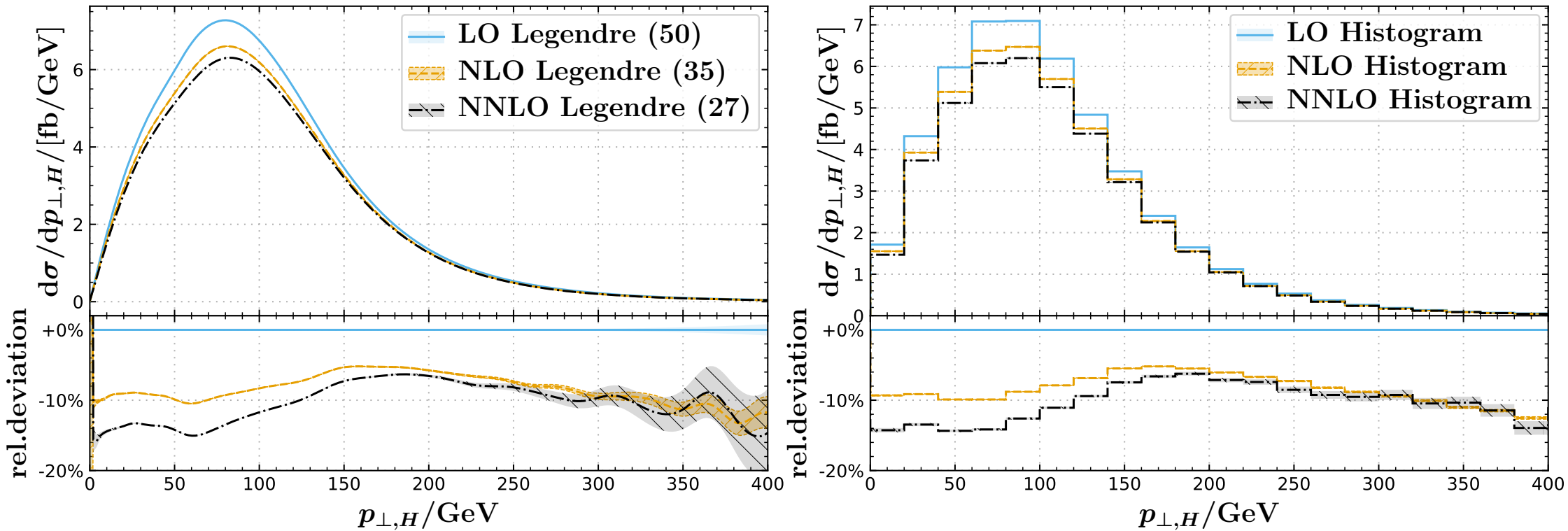
However, the value of the observable can be different between the event and the counterevent.



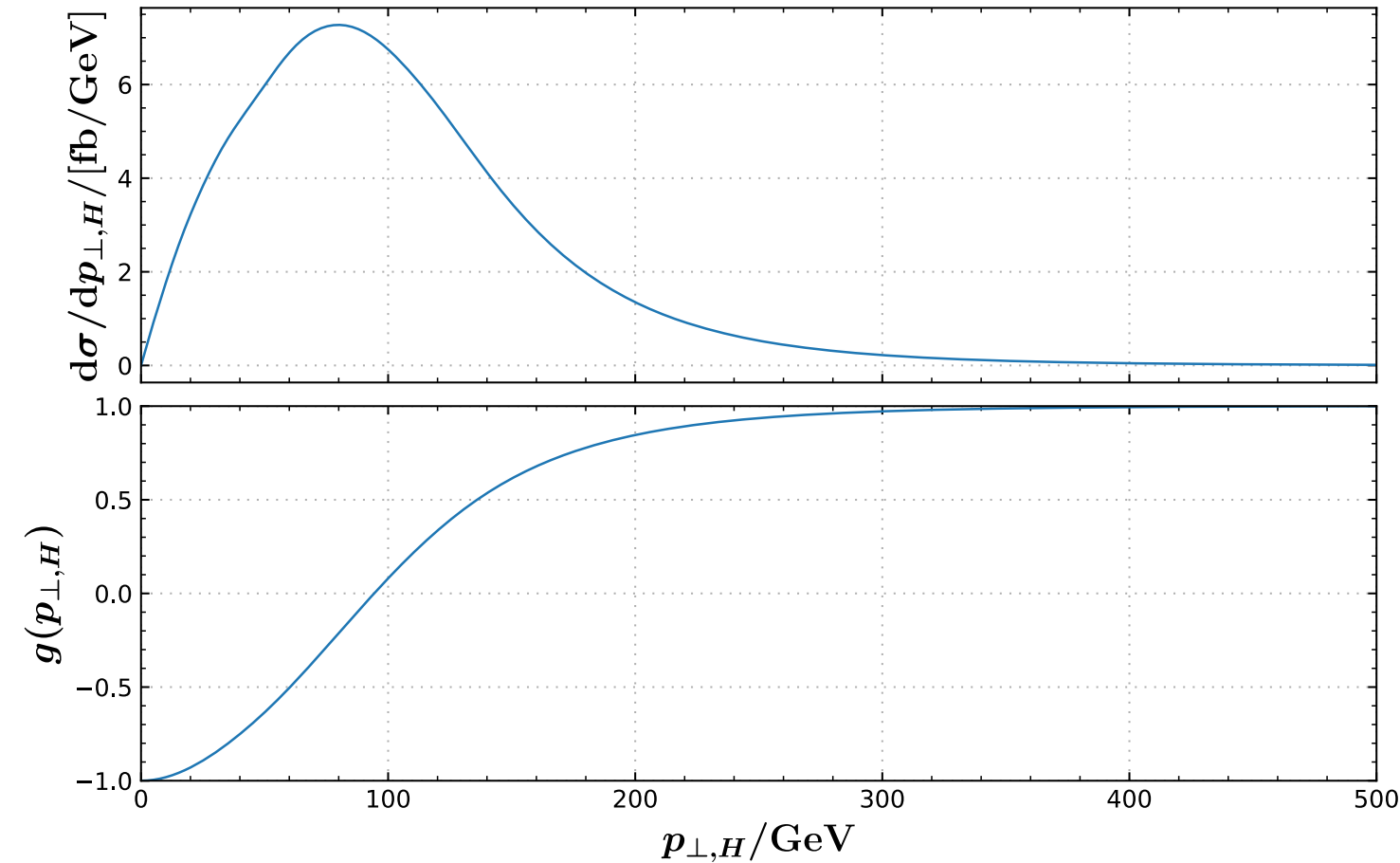
Cancellation between a real-emission event and a counter-event does not happen when they land in different bins, leading to a large fluctuation. Moments are immune to this issue.



Some real-emission corrections are particularly sensitive to such fluctuations



The absolute approximation error of moments-based approximation is uniform across the interval.
 As a result, the relative approximation error is large where the distribution function is small.

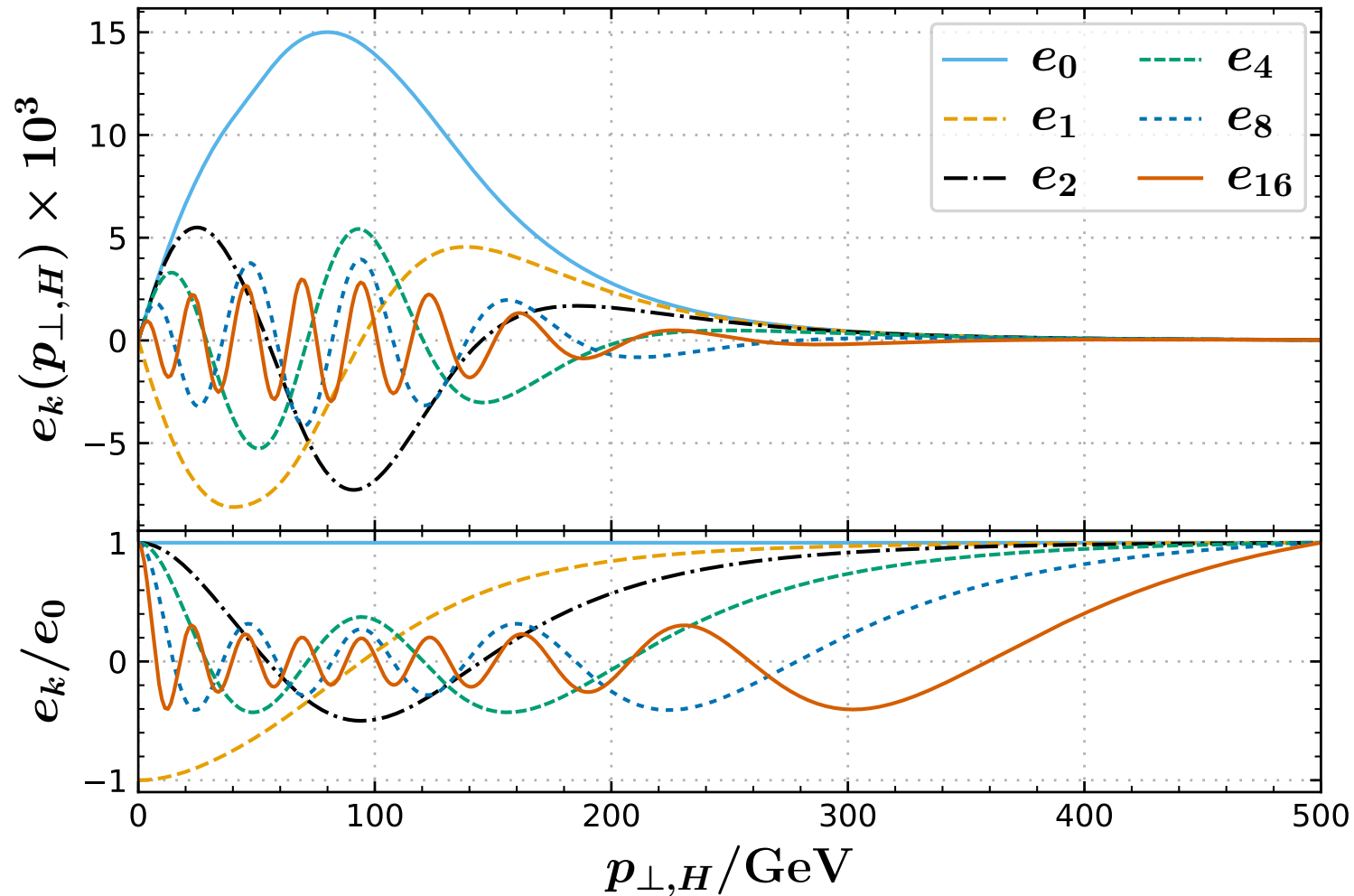


$$g(x) = -1 + 2 \frac{\int_a^x f_0(y) dy}{\int_a^b f_0(y) dy}$$

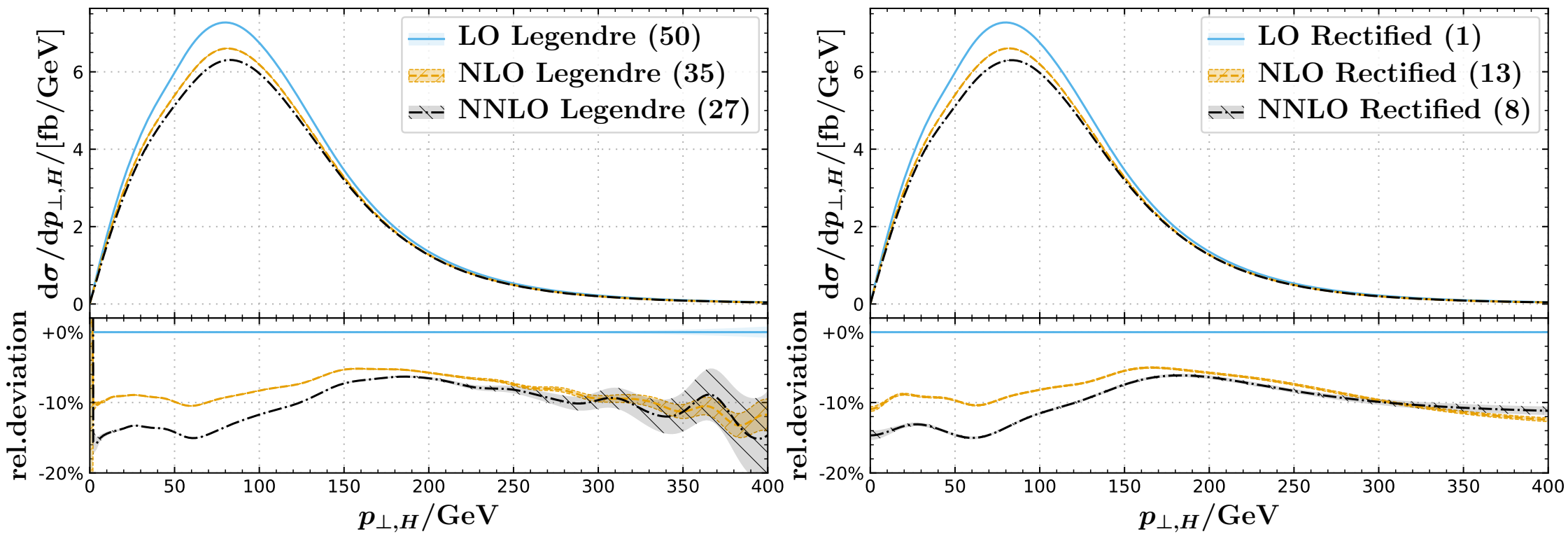
$$e_k(x) = g'(x) e_{0k}(g(x))$$

An initial approximation to the target distribution (e.g., a leading-order result), can be used to construct an orthonormal basis optimized for this particular distribution:

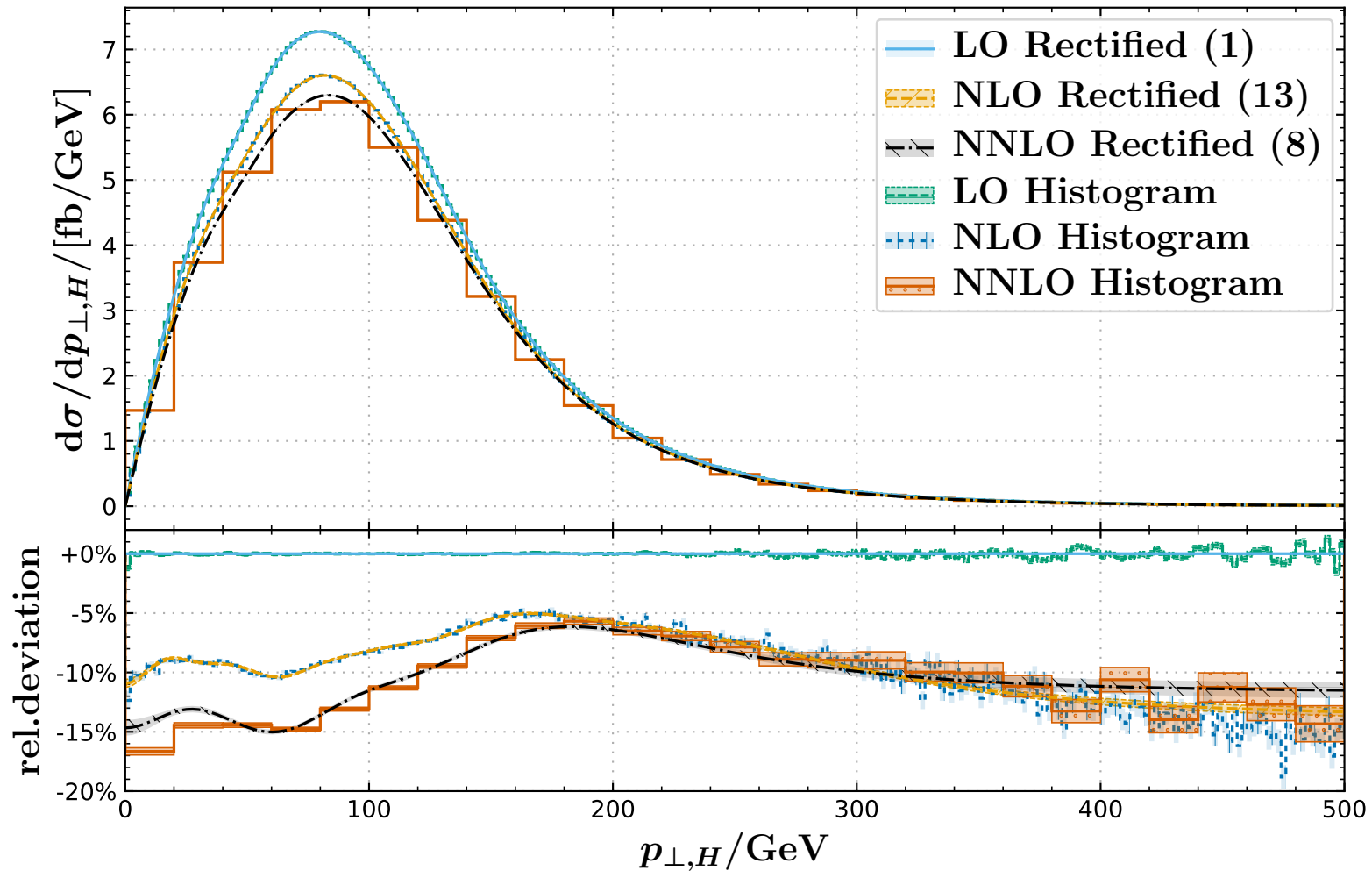
The idea is to make a change of variables that makes the initial approximation flat, then we can expect higher-order corrections to have a simple shape in those coordinates



The first function e_0 in the optimized basis is proportional to the initial approximation f_0 , and the higher functions are variations around it. Fluctuations of the higher basis functions are concentrated around the peak of the distribution, and all basis functions feature the exponentially-decaying tail.

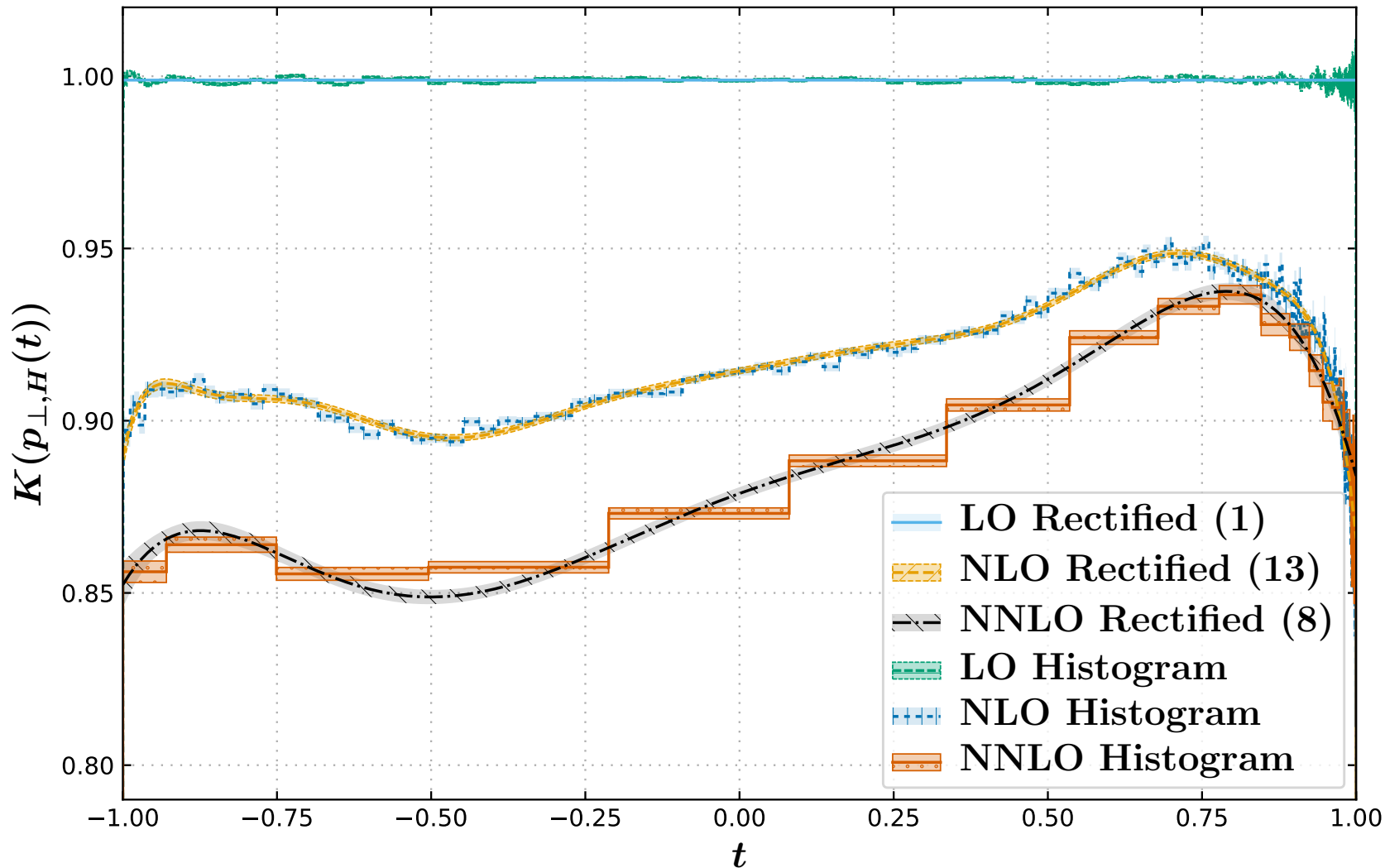


Such an optimized basis converges faster and smoothes out fluctuations in the tails.



The optimized basis can capture very fine features of the distribution around its peak while yielding a smooth approximation in the tail.

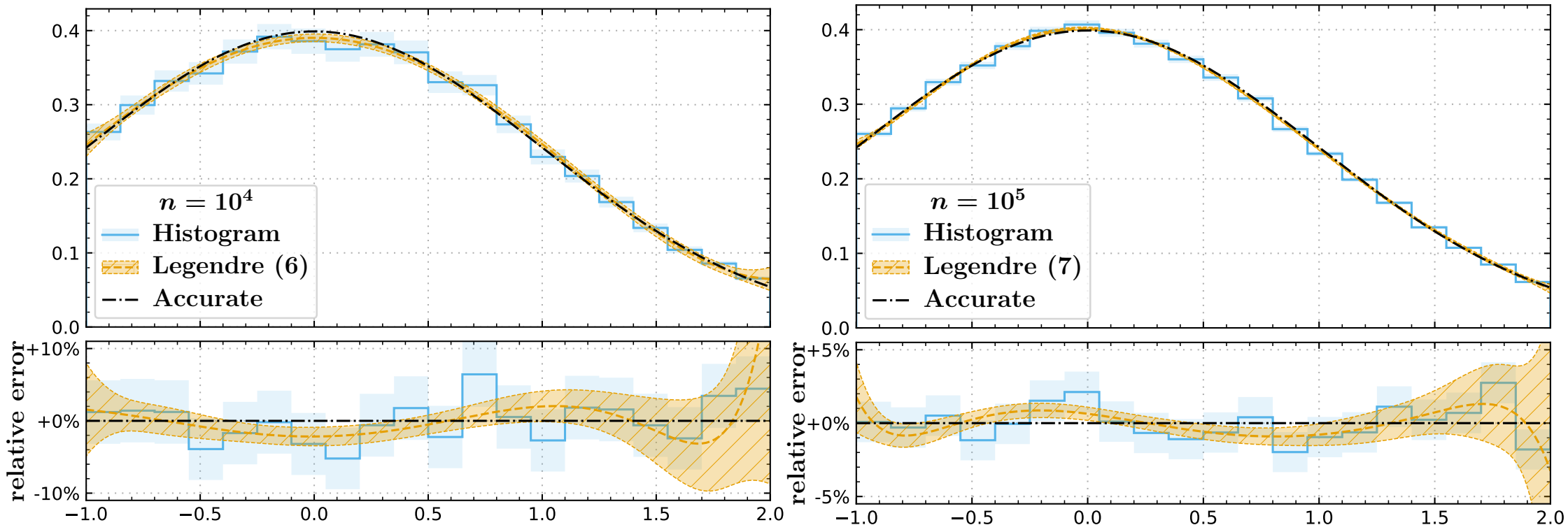
However, the smoothing in the tail region can sometimes be too aggressive.



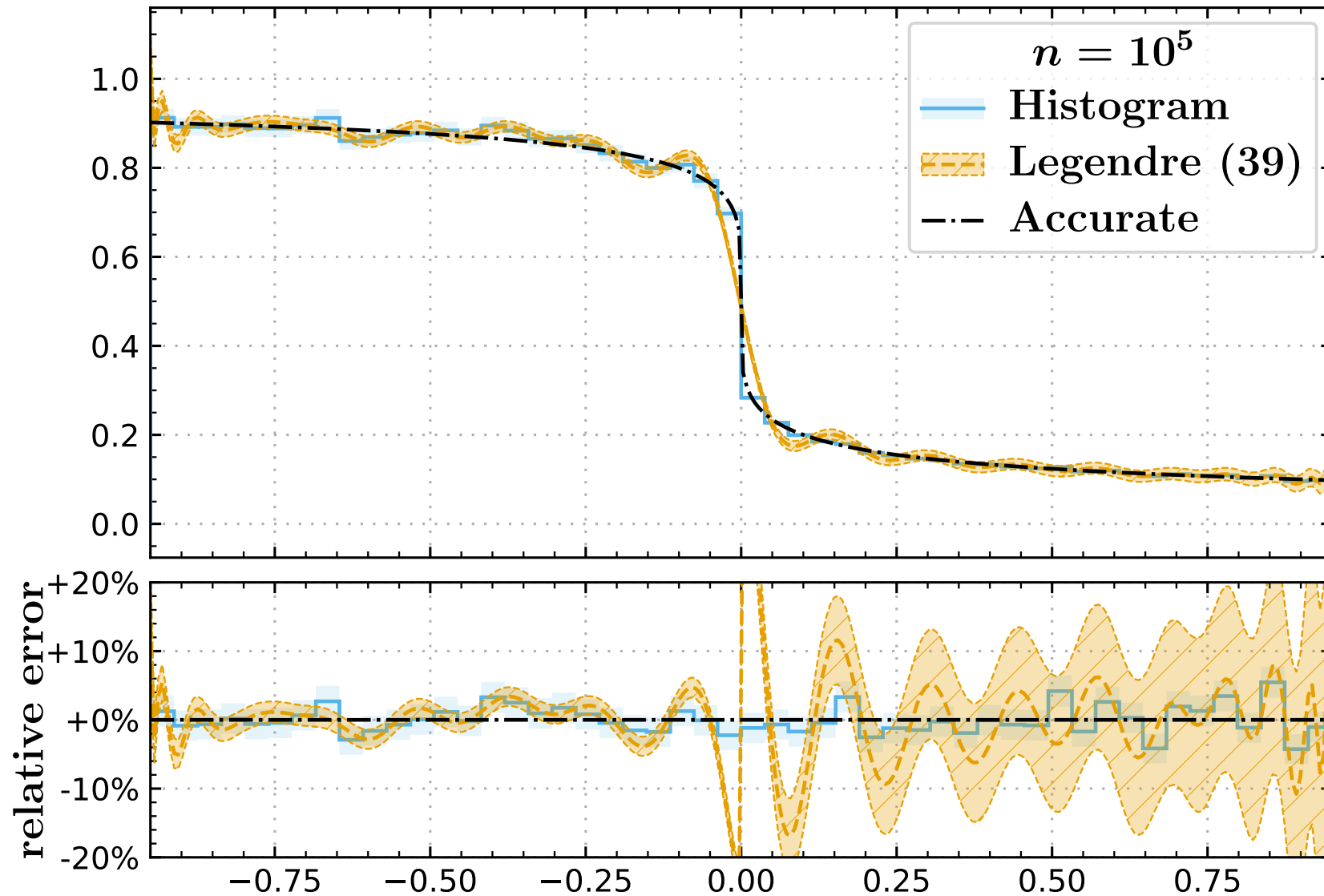
In this case the excessive smoothing in the tail region happens because the tail of the distribution $300\text{GeV} \lesssim p_T \leq 500\text{GeV}$ corresponds to a narrow range $0.95 \lesssim t \lesssim 1$ in the transformed coordinates.

THANK YOU FOR THE ATTENTION!

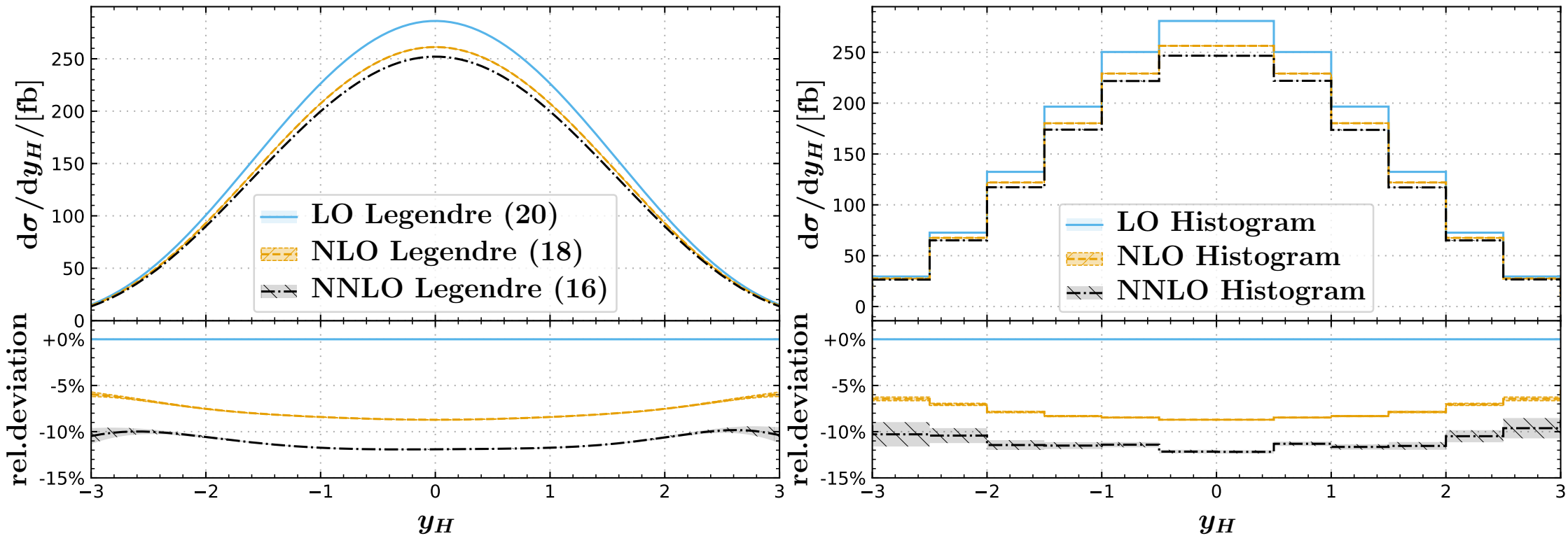
BACKUP



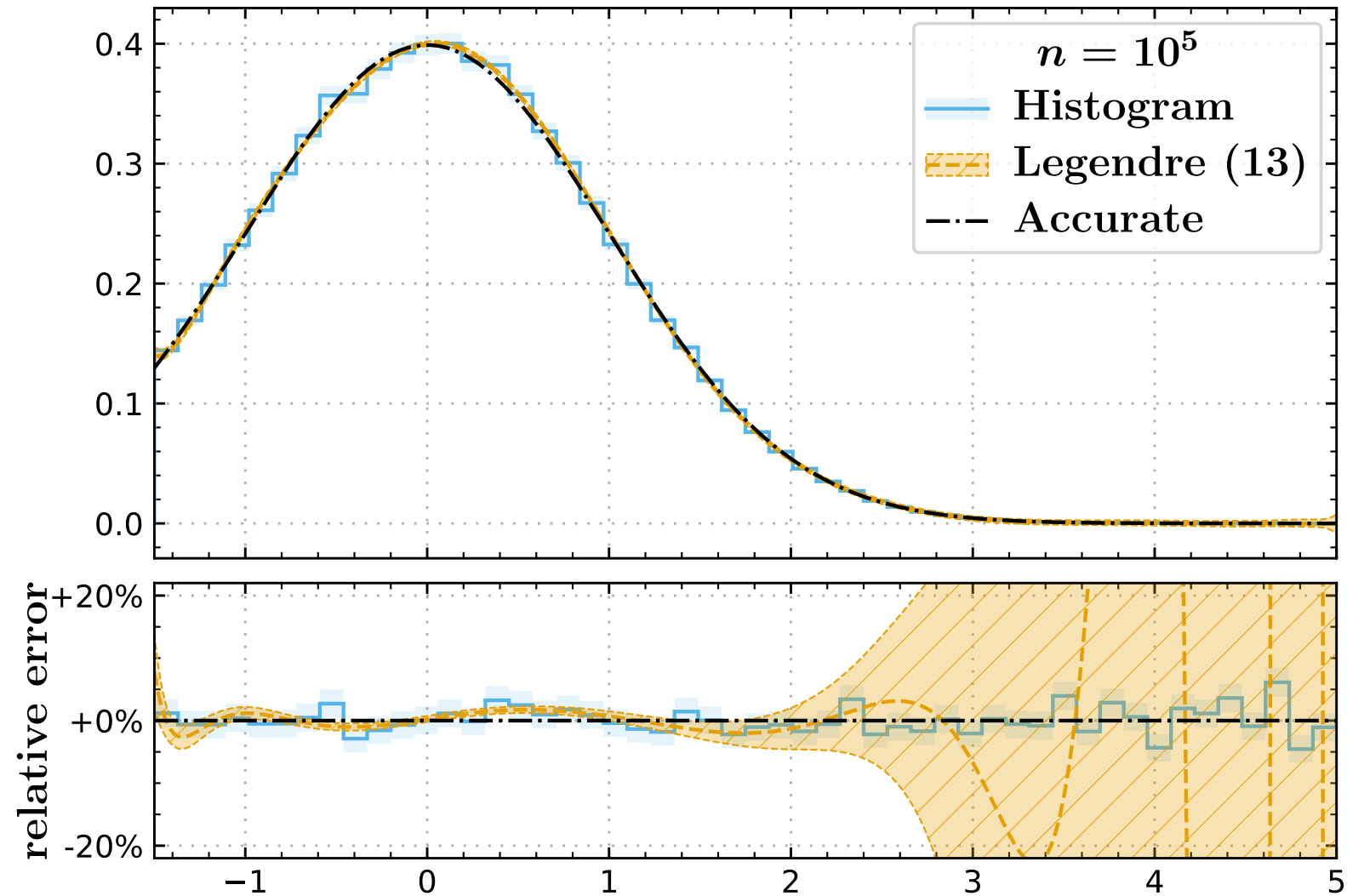
This works well for distributions without high-frequency features



Moments converge slower for functions with a sharp edge



We looked at a few different distributions in single-Higgs production in weak-boson fusion



The problem is that the absolute approximation error is uniform across the interval. As a result, the relative approximation error is large where the distribution function is small.

$$\langle f, e_k \rangle \sim \mathcal{N}(0, cr^k) \quad \delta c_k \sim \mathcal{N}(0, \delta \hat{c}_k) \quad \Rightarrow \hat{c}_k = \langle f, e_k \rangle + \delta c_k \sim \mathcal{N}\left(0, \sqrt{c^2 r^{2k} + \delta \hat{c}_k^2}\right)$$

$$\Rightarrow \text{minimize } \mathcal{L}(c, r) = \sum_{k=0}^m \left[\ln(c^2 r^{2k} + \delta \hat{c}_k^2) + \frac{\hat{c}_k^2}{c^2 r^{2k} + \delta \hat{c}_k^2} \right]$$

We fit the trend of moments decay to estimate the truncation error