Razor Boost Without Explicit Fitting

Sezen Sekmen[1] and Harrison B. Prosper[2]

[1] CERN, Geneva, Switzerland

[2]Department of Physics, Florida State University, Tallahassee, USA

CMS Razor Group

January 27, 2013

Razor Analyses

A key assumption of the razor analyses, which are based on the razor variables M_R and R^2, is that the background density in the razor plane (M_R, R^2) can be adequately modeled using the 4-parameter density

$$\begin{aligned} p(x, y|\theta) &= (z-1)e^{-n\,z}, \\ \text{where} \quad z &= b\,|(x-x_0)\,(y-y_0)|^{1/n}, \\ x &= M_R, \\ y &= R^2, \end{aligned} \tag{1}$$

and $\theta = x_0, y_0, b$, and n are free parameters. The razor likelihood in either the side-band region or the signal region is given by a marked Poisson process

$$L(x, y|\mu, \theta) = \text{Poisson}(N|\lambda) \prod_{j=1}^{K} \left(s\,p_s(x_j, y_j|\mu) + \sum_{l=1}^{M} b_l\,p_{bl}(x_j, y_j|\theta_l) \right), \tag{2}$$

where N is the total observed count, n_s and b_l are the expected signal and background counts, respectively, $\lambda = n_s + \sum_{l=1}^{M} b_l$ is the total expected count, and p_s and p_{bl} are the *normalized* signal and background densities, respectively. In general, the background parameters θ_l differ from one background component to another.

In the current razor analyses, a maximum likelihood fit is performed in a side-band region in which it is assumed that the signal content is negligible. The fit provides estimates $\hat{x}_0, \hat{y}_0, \hat{b}, \hat{n}$ and the associated 4×4 covariance matrix. In effect, the likelihood in the side-band region is approximated by a 4-variate Gaussian. The model in Eq. (1) is then applied to the signal region, along with the multivariate Gaussian, which serves as a prior that constrains the nuisance parameters x_0, y_0, b, n in the signal region.

The advantage of this approach is that a background model can be constructed once and for all and re-used with different signal hypotheses. The disadvantages are the need to assume negligible signal contamination in the side-band and the neglect of possible non-Gaussian tails in the side-band likelihood.

Alternative to Fitting It is always helpful to remind ourselves of the experimental goal. Our goal is to measure the effective differential cross section $d^2\sigma_{\text{eff}}/dxdy$ across the razor plane, where $\sigma_{\text{eff}} = \epsilon\,\sigma\text{BR}$ in which ϵ is the signal acceptance and σBR is the cross section times branching ratio. Given this experimental information, together with a clear, precise, and complete description of the event selection and a readily accessible approximation of the CMS detector response, any theoretical model that yields predictions for the differential cross section in the razor plane can be tested.

A straightforward way to achieve the goal is to discretize the razor plane and measure the signal content of each bin. This reduces the analysis problem to the well-understood multi-count experiment for which the likelihood is generally taken to be

$$L(x, y|\sigma, b, \mathcal{L}) = \prod_{j=1}^{K} \text{Poisson}(N_j|\sigma_j\,\mathcal{L} + \sum_{l=1}^{M} b_{lj}), \tag{3}$$

where, now, K is the number of bins in the razor plane, N_j is the observed count in bin j, σ denotes the effective cross sections $\sigma_1, \ldots, \sigma_K$, and

$$\begin{aligned} b(\theta) &= b_{\text{tot}} \int_{x_{\min}}^{x_{\max}} dx \int_{x_{\min}}^{x_{\max}} dy\, p(x, y|\theta), \\ &\approx b_{\text{tot}}\,(x_{\max} - x_{\min})(y_{\max} - y_{\min})\, p(x, y|\theta), \end{aligned} \tag{4}$$

gives the expected background count in a given bin for a given background component, where b_{tot} is the total expected background in the razor plane. The likelihood can then be reduced to a function of the effective cross sections only, either by profiling over the background parameters θ_l and the integrated luminosity $\mathcal{L}$, constrained by the luminosity prior $\pi(\mathcal{L})$, or by marginalization. This model is one that could, and probably should, be implemented using `HistFactory`.