1 On PIDCalib Systematics

This should be a (hopefully) brief aside about what I view to be the incompleteness of the data-driven method of assessing the systematic uncertainty of PIDCalib, specifically detailing how an aspect of the systematic changes between modes and even different binning structures in these modes. I'll use the decay $\Lambda_c^+ \to p^+ K^- K^+$ to illustrate this.

The method of using reference events from the same mode as used in making the PerfHists in PIDCalib to assess the systematic is a very good way of assessing the part of the systematic arising from the statistical effects of lower stats in each bin, if the number of reference events used in the check is the number of signal events which are used in the actual calibration. There is, however, an aspect of the systematic which arises from how suitable the binning structure is to an accurate "reweighting".

There is a systematic arising from the fact that we assume within each bin in the PerfHist there is no variation in response to a given PID cut. Naturally this is not the case and there is some loss of information.

The problem with using reference tracks from the same mode used to generate the Perfhist is that in any finite region of the kinematic phase space the tracks will, on average, have the same response to any PID cut. In this case the binning structure is entirely independent of the loss of information, given that with any arbitrarily coarse binning, and even in the one bin case, the identical kinematic distributions of the tracks between the reference and calibration samples results in the same average response to any PID cut.

To put this another way imagine that we have a binning in some quantities entirely independent of the PID variables, say IP χ^2 vs. PV x-position. Because the tracks falling within that bin have the same kinematic distributions in both the reference and calibration sample, specifically the same distributions in track kinematics which *are* dependent on PID performance, we would still expect compatible PID cut responses between the two when averaged over the whole bin. This would be entirely unrepresentative of the accuracy of the assigned PID response when tracks with different daughter kinematics are used.

1.1 Improper Bin Efficiencies

To illustrate this I have a real life example. For protons the kinematic distributions in the calibration sample, $\Lambda^0 \to p^+\pi^-$, are very different from those of the protons in the $\Lambda_c \to p^+h^-h^+$ modes which I'm working on. Namely the P and P_T tends to be higher in my signal modes, and specifically at higher P we find that lower values of η are populated than in the calibration mode.

Figure 1 shows the kinematic distributions of the protons in the $\Lambda_c^+ \to p^+K^-K^+$ mode after our PID cuts. Figure 2 shows the PerfHist showing the corresponding efficiencies in these variable distributions for this cut. Take for instance the region around $P = 60 GeV/c$ and $\eta = 3$. We have signal events in this region, yet the PIDCalib output from the perfhists tells me that there are negative events in the bin covering this region (sWeight artefacts resulting from low calibration stats in this region after cuts). Correspondingly if I do not veto these events from my signal sample then they will be assigned underestimated PIDCalib efficiencies, and the average efficiency for my mode will be underestimated. This is a phenomenon that will not be emergent when using a reference sample with the same kinematics as the calibration sample.

1.2 Biases arising from PID response variation across a bin

There is a separate issue with biases arising even when we have a valid efficiency for a bin. Take the hypothetical case of a single bin in 2 PID dependent variables in Figure 3. The left graphics are before PID cuts, right after. The top shows the calibration sample tracks in that bin and the bottom shows tracks from some arbitrary signal mode with different track kinematics falling in that bin. Imagine that within this bin the PID variable response is higher in the top left of the bin, and lower in the bottom right. Correspondingly, a given PID cut will have less of an effect on the calibration sample than it will on our signal sample. We of course have no access to the bottom left plot if our signal mode's stripping line contains the PID cut, and will assign the leftover tracks on the bottom right the efficiency from the calibration sample, which is higher than the true efficiency value for the signal mode.

Were we to use the reference sample from the same mode as the calibration sample then the kinematic distributions of events in this bin would be the same. Therefore the PID efficiency of the bin would be a representative average within statistical fluctuations, as would be the case for all bins in all binning schemas. The only part of the systematic the data-driven method has access to is the statistical part of this uncertainty, it has no means to assess true *biases* which arise from the combination of finite bin widths and differing decay kinematics between calibration and signal modes.

Naturally these effects can be mitigated with finer binning and kinematic vetoes on daughter tracks, but the user requires a way to assess the biases emerging in this way, so that they can adjust their binning schema, variables used and kinematic vetoes to minimise the full and proper systematic. As a caveat these features are much more pronounced for final states involving protons, where we have both a calibration mode with a very low momentum spectrum and a much more rapidly varying PID efficiency across kinematic phase space. For K/π the variations in efficiency across the phase space are much lower and this problem becomes less pronounced, but the biases are still there.

It therefore seems to me that the only suitable way to evaluate the systematic uncertainty of PIDCalib, especially when protons are used, is to do so with a sample which has the same kinematics as your signal mode, and with which we have access to the total dataset without PID cuts. The only way of doing this sadly would be using MC.

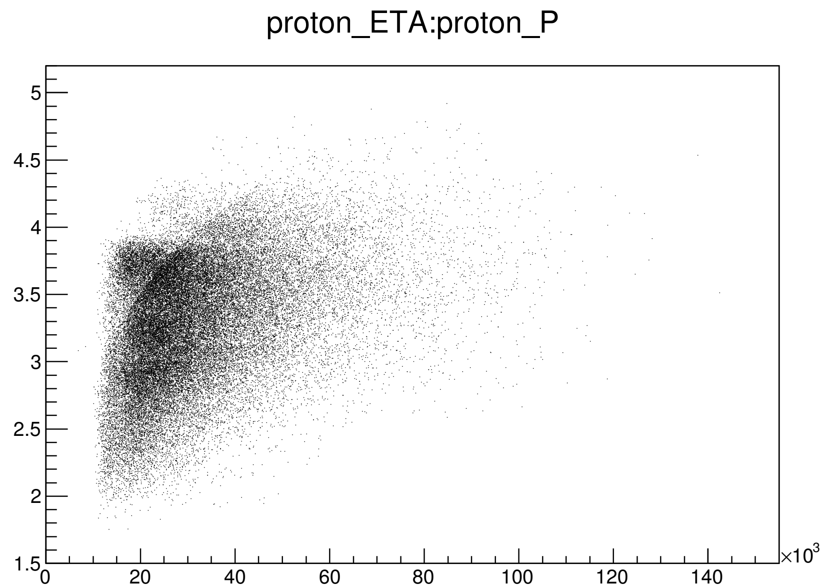

Figure 1: Proton kinematics in $\Lambda_c^+ \to p^+ K^- K^+$

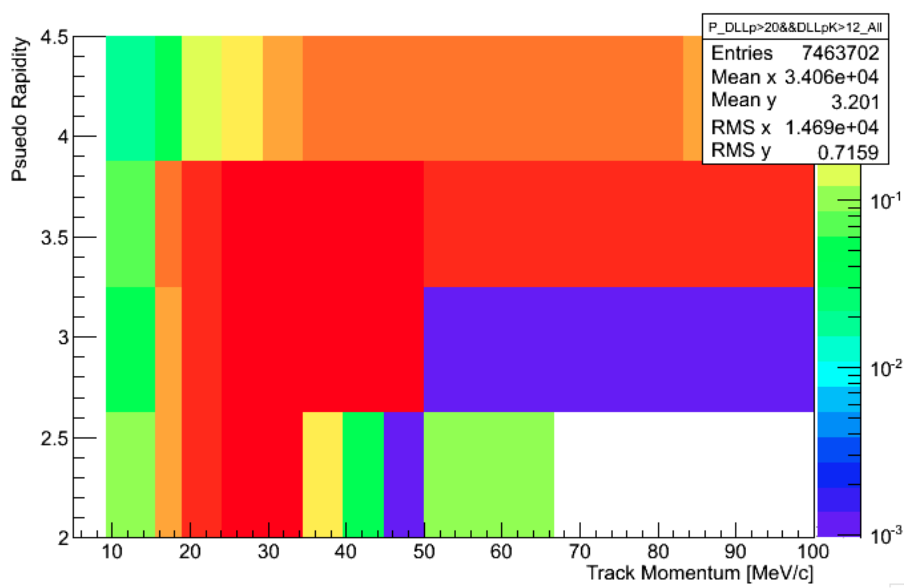

Figure 2: PerfHist for protons made with the $\Lambda^0 \to p^+ \pi^-$ calibration data.

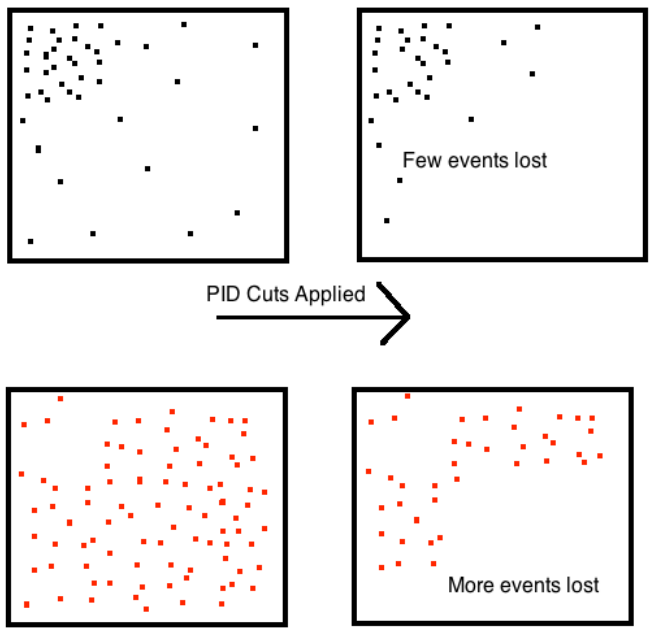

Figure 3: A rough representation of a systematic bias which arises from differing track kinematics between calibration and signal modes, which cannot be assessed by using a reference sample with the same kinematics as the calibration sample. Imagine this to be the same bin in a PerfHist, with calibration tracks top in black and signal tracks with a different kinematic distribution bottom in red. After PID cuts a higher fraction of signal tracks are removed compared to the calibration tracks. The bin efficiency of a PID cut assigned by PIDCalib will therefore be higher than the true efficiency of the same PID cut on the signal sample.