# Event density plots, A more dynamical alternative for histogram’s

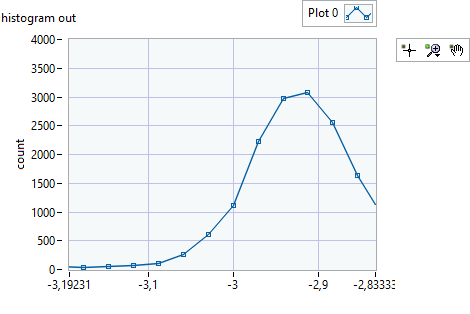
During evaluation of the results from our cytometer, we used the histograms LabVIEW library. Often, the data turns out to be located in a limited number of bins and only after tweaking the range and bin size parameters, the data is represented nicely. To overcome this problem, I devised a plot in which every point the same number of events and the bin size is different for each point. The value displayed on the y axes is than given by the number of events divided by the size of the bin. It turns out that this approach has great advantages in implementation efficiency and represents the data much better. Figure 1 shows data in the a traditional histogram where all bin sizes are equal, in figure 2, the same data is plotted as a event density plot where data point are closer to each other where the intensity is high and further apart in low intensity regions. 

Figure 1 Data plotted in a “traditional” histogram, on the vertical axis is the number of counts in each bin



Figure 2 Event density plot, each point represents the same number of events. On the vertical axes is the event density.

During my presentation I will elaborate on implementation and the advantages of this type of plot. Furthermore, I will present refinements that I implemented and I would like to receive feedback.

At some point I developed an aversion for this implementation because I needed to supply bin size and range parameters which I did not know at forehand. That is when I try to automate setting these parameters. Also, I encountered a problem that each point on the histogram carries a different information magnitude. Often, the recorded events were concentrated in a couple of bins whereas the other bins were empty. To solve this problem, I decided to make bins such that each bin is “information wise” equally important. That means that each bin contains the same number of events. This approach leads to a different histogram that has a different bin size.

The two types of figures look similar but, at the same time, are quite different. In the dynamical Histogram, the bins are wider in the areas with little events and are narrow at the areas where the event density is high. That means that with a limited number of bins, you still can see what happens at the peaks and the areas where nothing happens are skipped. I think for interpretation of the data, just looking at it but also for interpretation using automated models, is better using these histograms.

Implementation efficiency

Dynamical histograms have advantages in implementation. The algorithm is very simple, it is sorting the data then defining data segments equal sizes. Using the segments of length n, we calculate the “event density” using the data interval between first elements of subsequent segments. The event density \rho(x) is calculated using the next formula:

When I proposed this, my colleague complained that maybe there would be many observations with the same value which would explode the histogram values. I argued that observations are floating point values so the changes of having two observations with the same values are virtually zero. I was wrong! Observations in our cytometer are derived from data that are digitized values converted to floating points.

Afbeelding met tekst, lijn, Perceel, diagram

Automatisch gegenereerde beschrijving

Afbeelding met tekst, Perceel, lijn, diagram

Automatisch gegenereerde beschrijving

Cumulativee

Toevoegen

Invloed van as op het histogram

Uitbreidingen en verfijningen.