ROOT: Errors and Weights

Baylor Mini-Symposium
December 10th, 2009
Karen Bland
In Root

• In practice, here’s the main idea...
  – Use any time you’re weighting/scaling or doing an operation such as adding or dividing histograms
  – If in doubt, use Sumw2() -- it doesn’t hurt. The weights are 1 by default.

• How do you use it?
  TH1F* ha = new TH1F("ha", "Guassian Histo A", 100, -6, 6);
  ha->FillRandom("gaus", 10000);
  ha->Sumw2();
In Root: The function `Sumw2()`

- Create structure to store sum of squares of weights
- If histogram is already filled, the sum of squares of weights is filled with the existing bin contents
- The error per bin will be computed as `sqrt(sum of squares of weight)` for each bin.
- This function is automatically called when the histogram is created if the static function `TH1::SetDefaultSumw2` has been called before
Demo of Scaling, Adding, Dividing

• See Demo...
• In Summary:
  – Scaling -> Use Sumw2()
  – Adding/Subtracting -> Use Sumw2() if a histogram is weighted
  – Multiplying/Dividing -> Use Sumw2() if a histogram is weighted
Quick Stat Review

• In general:
  – Mean = \( u = \Sigma (x_i \Pr(x_i)) \)
  – Var = \( \sigma^2 = \Sigma [ (x_i - u)^2 \Pr(x_i)] = \Sigma [ x_i^2 \Pr(x_i)] - u^2 \)
  – Standard Deviation = \( \sigma = \sqrt{\sigma^2} \)

• Propagation of error:

Poisson Distribution

- An approximation of binomial when average # of successes is much smaller than possible number (μ<<n because p<<1)
- Generally a good approximation for counting experiments where data represent # items observed per unit time interval P_B(x,n,p) for binomial would be prob of x events per unit time interval out of n possible events, each w/ a prob of p occurring, but...
  - the larger number n of possible events makes exact evaluation impossible
  - Often n and p unknown; instead maybe average number of possible events observed per unit time, μ, or estimate
- Appropriate distribution for describing experiments in which the possible values of data are strictly bounded on one side but not on other
- Advantage of distribution is that it is specified completely by the mean, μ, and and the standard deviation, σ, is uniquely determined by mean (so that means also good for describing nearly symmetric curves)

\[
P_p(x,μ) = \frac{μ^x}{x!} e^{-μ}
\]

\[
σ^2 = μ
\]

\[
σ = \sqrt{μ}
\]
Statistical Fluctuations: Introduction

• “Poisson distribution is appropriate for describing spread of data points in counting experiments where the number of items detected per unit time interval constitute the observations. For such experiments, the measurements fluctuate from observation to observation not because of any imprecision in measuring the time interval or because of an inexactness in counting the number of events occurring the interval, but because random samples of events distributed randomly in time contain numbers of events which fluctuate from sample to sample.

• In any given time interval there is a finite nonzero chance of observing any integral number of events. The probability for observing any specific number of counts is given by the Poisson probability function, with a specified mean \( \mu \) and a resulting uncertainty \( \sigma = \text{Sqrt}(\mu) \). Since the fluctuations in the observations result from the statistical nature of the distribution, they are classified as statistical fluctuations, and the resulting errors in the final determination are classified as statistical errors.” - Bevington
Statistical Fluctuations: Mean and Standard Deviation

• The best estimate of the mean for unequal uncertainties for the Poisson distribution (using the method of least squares):
  \[ \mu \approx \frac{\Sigma(x_i/\sigma_i^2)}{\Sigma(1/\sigma_i^2)} = \frac{\Sigma(w_ix_i)}{\Sigma(w_i)} \]

• But the uncertainties \(\sigma_i\) associated with the data must be determined from the actual measurements \(x_i'\):
  \[ \sigma_i \approx \sqrt{(x_i' - \text{bar})} = \sqrt{\text{(sample mean of original values)}} \]
  where \(x_i' - \text{bar}\) is the average of all the measurements that are made with the same uncertainty \(\sigma_i\) as the measurement \(x_i'\).