Procedure to normalize MC samples to data

Pierre Chatagnon, Rafayel Paremuzyan

August 2, 2023

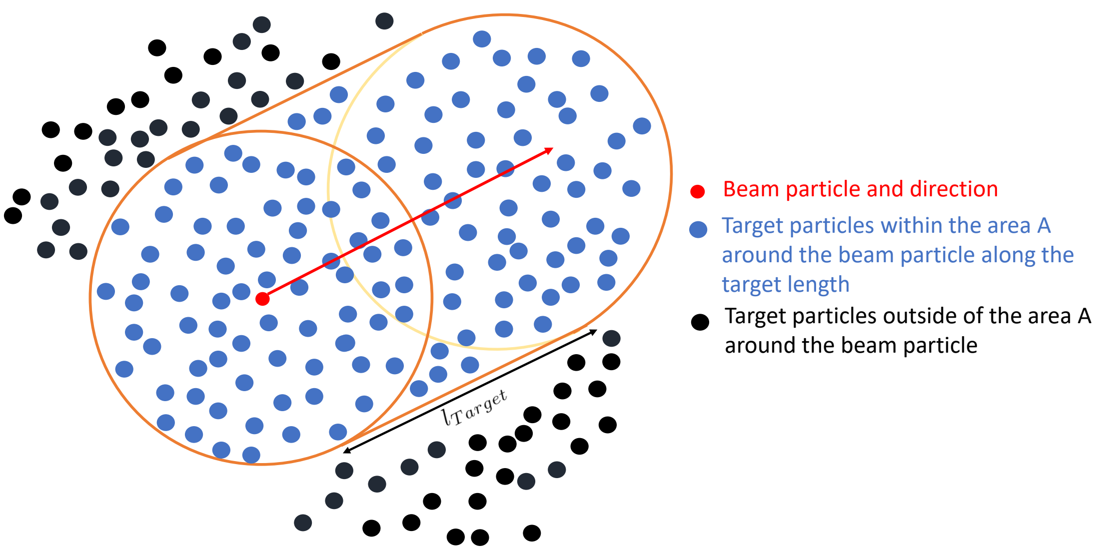
Abstract

This document describes the steps needed to properly normalize data to MC based on the integrated luminosity of the data sample under analysis.

1 Integrated luminosity

1.1 General formula

In order to normalize data to MC, the integrated luminosity of the dataset under scrutiny must be determined first. In the case where the beam diameter is much smaller than the target transverse size, the luminosity is defined as the number of beam particle multiplied by the number of target particle within a given area. This is describe schematically in Figure 1. Reversely it can also be seen as the number of event per unit of cross section that will be produce for a given dataset. It has the unit of a inverse squared length and is usually reported in cm^{-2} or b^{-1}. For fixed-target experiment as CLAS12, the integrated luminosity depends on the total beam charge on target and the target parameters (length and density).



$$\mathcal{L}_{\mathcal{INT}} = N_{beam} \times \frac{N_{Target}}{A} = N_{beam} \times n_{Target}$$

Figure 1: Schematic description of the luminosity concept in a fixed target experiment. Each beam particle can interact with any target particle within a certain transverse area along the whole length of the target.

It can be written as:

$$\mathcal{L}_{\mathcal{INT}} = N_{beam} \times n_{Target} = \frac{Q}{e} \times \frac{l \cdot \rho \cdot N_t \cdot N_A \cdot C}{M_t}, \tag{1}$$

where N_{beam} is the total number of electron that crossed the target and n_{Target} is the number of "target" particle per area in the target of CLAS12. Each of these two quantities can be expressed in terms of the following quantities:

- $\mathcal{L_{INT}}$ [in pb^{-1}]: the total integrated luminosity
- Q [in C]: the total accumulated beam charge for the dataset under scrutiny
- l [in cm]: the length of the target cell
- ρ [in $g.cm^{-3}$]: the density of the target material
- N_t [No units]: the number of "target" atom in the target molecule
- N_A [in mol^{-1}]: the Avogadro number ($6.02 \times 10^{23} mol^{-1}$)
- C [No units]: Conversion factor from cm^{-2} to pb^{-1} (10^{-36})
- e [in C] : the electron charge ($1.602 \times 10^{-19} C$)
- M_t [in $g.mol^{-1}$]: Molar mass of the target material

1.2 Application to RG-A

The formula above can be applied to the RG-A target by using the following inputs:

- l [in cm]: 5 cm
- ρ [in $g.cm^{-3}$]: 0.0708 $g.cm^{-3}$ (Density of liquid di-hydrogen)
- N_t [No units]: 2, the number of "target" proton in dihydrogen
- N_A [in mol^{-1}]: the Avogadro number ($6.02 \times 10^{23} mol^{-1}$)
- C [No units]: Conversion factor from cm^{-2} to pb^{-1} (10^{-36})
- e [in C] : the electron charge ($1.602 \times 10^{-19} C$)
- M_t [in $g.mol^{-1}$]: 2.016 $g.mol^{-1}$ Molar mass of dihydrogen

Replacing the numerical values in the general formula, the integrated luminosity for RG-A reads:

$$\mathcal{L_{INT}} = 1316.875 \times 10^3 \times Q[\text{in C}]\ pb^{-1} \tag{2}$$

or

$$\mathcal{L_{INT}} = 1316.875 \times Q[\text{in mC}]\ pb^{-1} \tag{3}$$

1.3 Application to proton scattering on RG-B target

The formula above can be applied to proton scattering experiments on the RG-B target by using the following inputs:

- l [in cm]: 5 cm
- ρ [in $g.cm^{-3}$]: 0.162 $g.cm^{-3}$ (Density of liquid D2)
- N_t [No units]: 2, the number of proton in D2
- N_A [in mol^{-1}]: the Avogadro number ($6.02 \times 10^{23} mol^{-1}$)
- C [No units]: Conversion factor from cm^{-2} to pb^{-1} (10^{-36})
- e [in C] : the electron charge ($1.602 \times 10^{-19} C$)
- M_t [in $g.mol^{-1}$]: 4.02 $g.mol^{-1}$ Molar mass of D2

Replacing the numerical values in the general formula, the integrated luminosity for RG-B reads:

$$\mathcal{L_{INT}} = 1514.34 \times 10^3 \times Q[\text{in C}]\ pb^{-1} \tag{4}$$

or

$$\mathcal{L_{INT}} = 1514.34 \times Q[\text{in mC}]\ pb^{-1} \tag{5}$$

1.4 Application to free proton scattering on RG-C ammonia target

- l [in cm]: 5 cm
- ρ [in $g.cm^{-3}$]: 0.817 $g.cm^{-3}$ (Density of solid ammonia at -80°C) Need to be adjusted with the effective density of the RG-C target taking into account the fact that the target is made of beeds of ammonia in an helium bath
- N_t [No units]: 3, the number of "target" free proton in NH3
- N_A [in mol^{-1}]: the Avogadro number ($6.02 \times 10^{23} mol^{-1}$)
- C [No units]: Conversion factor to pb^{-1} (10^{-36})
- e [in C] : the electron charge ($1.602 \times 10^{-19} C$)
- M_t [in $g.mol^{-1}$]: 17.03 $g.mol^{-1}$ Molar mass of ammonia

Replacing the numerical values in the general formula, the integrated luminosity for RG-C reads:

$$\mathcal{L}_{\mathcal{INT}} = 2704.16 \times 10^3 \times Q[\text{in C}]\ pb^{-1} \tag{6}$$

or

$$\mathcal{L}_{\mathcal{INT}} = 2704.16 \times Q[\text{in mC}]\ pb^{-1} \tag{7}$$

NB: This formula applies after nuclear subtraction has been performed. Only the free proton luminosity is reported here. If ones wants to study the scattering on any proton of the target, N_t must be changed to $17 = 2 \times 7\ protons\ in\ N + 3 \times 1\ proton\ in\ H$

2 Monte-Carlo re-weighting

There are two types for Monte-Carlo generator: unweighted and weighted MC. In the first case, each event produced by the Monte-Carlo can be treated as a data-event. In the second case, each MC event is assigned a weight that must be weight the event when filling histograms. In each case the normalization data/MC is slightly different.

2.1 Monte-Carlo with un-weighted events

In the majority of Monte-Carlo event generators, each generated event do not carry any weight and can be treated exactly as a data event. This is the case of GRAPE. In this case, when using the generator, users are provided with the total cross-section of the process over the phase-space where events have been produced. This cross-section is referred as σ_{GEN_TOT} [in pb]. The total number of generated events for this given phase-space is denoted N_{GEN}.

To normalize an un-weighted MC sample to a given dataset with integrated luminosity $\mathcal{L}_{\mathcal{INT}}$ [in pb^-1], one should assign to each MC event a weight equal to:

$$\omega = \frac{\mathcal{L}_{\mathcal{INT}} \cdot \sigma_{GEN\_TOT}}{N_{GEN}}. \tag{8}$$

The weight factor ω is constant. Hence, one can, instead of re-weighting each event, scale all MC histograms by this same factor.

2.2 Monte-Carlo with weighted events

Some generator provide weight for each event. Usually these weight corresponds to the cross-section at the given kinematics of the events multiply by a phase-space factor accounting for the fact that the event is randomly generated over a certain limited phase-space. This is the case of TCSGen and JPsiGen.

To normalize a weighted MC sample to a given dataset with integrated luminosity $\mathcal{L}_{\mathcal{INT}}$ [in pb^-1], one should assign to each MC event a weight equal to:

$$\omega = \frac{\mathcal{L}_{\mathcal{INT}} \cdot \omega_{GEN}}{N_{GEN}}, \tag{9}$$

where ω_{GEN} is the weight provided by the generator.