Event Weighting vs. Event Counting

Jörg Pretz[a,b,*]

[a]*Forschungszentrum Jülich, Jülich, Germany*
[b]*RWTH Aachen University, Aachen, Germany*

  E-mail: pretz@physik.rwth-aachen.de

Goal of these proceedings is to introduce a method based on event weighting in particle physics experiments. Weighting means that events are not just counted as integer numbers but are assigned a weight factor according to their importance in the analysis. This method has a close connection to the maximum likelihood method known to reach the smallest statistical error. The purpose of this document is to give a more educational overview on the subject. As an example the extraction of a beam polarization from scattered particles is discussed.

---

[*]Speaker

1. Introduction

In the analysis of particle physics experiments one often applies cuts on kinematical variables like angles and momentum. Based on the number of event fulfilling the cut criteria, physical quantities like cross section, analyzing powers or polarizations are deduced. Aim of this work is to show that this black and white picture, i.e. counting an event inside the cuts and rejecting it otherwise does in general not lead to optimal results in terms of statistical accuracy. It is shown that one can do better by assigning a weight factor to every event. A weight factor 0 corresponds to ignoring this event, a weight factor of 1 corresponds to counting the event. But an event could have a arbitrary weight factor between 0 and 1. This makes better use of information an event carries. How this is done quantitatively is shown in an educational way for the determination of particle beam polarization. This work is largely based on reference [1].

Many more examples using event weighting can be found in the literature, like the determination of the anomalous magnetic moment of the muon [2, 3], the gluon polarisation in the nucleon [4, 5], the polarized quark distribution in the nucleon [6], analyzing power in deuteron carbon scattering [7], three gauge boson coupling [8], electric or magnetic dipole moment of the τ-lepton [9, 10] and the measurement of the forward-backward asymmetry of Drell-Yan dilepton pairs [11]. In this context the term optimal observable is widely used in the literature.

This document is organized as follows. First a short introduction is given how event rates relate to a beam polarization. Then various methods to extract the beam polarization are discussed and compared.

2. Formulation of the problem

Starting point of experiments in particle physics is often the relation between the expected number of observed events N and the cross section σ containing the physics of the scattering process:

$$\left\langle \frac{\mathrm{d}N(\vartheta, \varphi)}{\mathrm{d}\Omega} \right\rangle = \mathcal{L}a(\vartheta, \varphi)\frac{\mathrm{d}\sigma(\vartheta)}{\mathrm{d}\Omega} \ . \tag{1}$$

Table 1 and figure 1 explain the variables and the coordinate system. Equation 1 holds for an unpolarized beam and target. If the beam is polarized additional contributions appear. Here we assume a spin polarized particle beam perpendicular to the momentum vector in y-direction. In this case the cross section receives additional contributions [12]:

$$\left\langle \frac{\mathrm{d}N(\vartheta, \varphi)}{\mathrm{d}\Omega} \right\rangle = \mathcal{L}a(\vartheta, \varphi)\frac{\mathrm{d}\sigma(\vartheta)}{\mathrm{d}\Omega} \left(1 + A(\vartheta)P\cos(\varphi) \right) \ . \tag{2}$$

A detailed derivation is given in appendix A.

For this paper the most interesting variables in equation 2 are the beam polarisation P, describing the degree of polarization of the beam and the analyzing power A describing how much the scattering process distinguishes between particles scattered to the left and right. As one can see from equation 2, one is only sensitive to the product of the analyzing power A and the polarization P. Here we assume that the analyzing power is known and the goal is to extract from the data, i.e. the measured event kinematics, the polarization P, with the smallest possible statistical error.
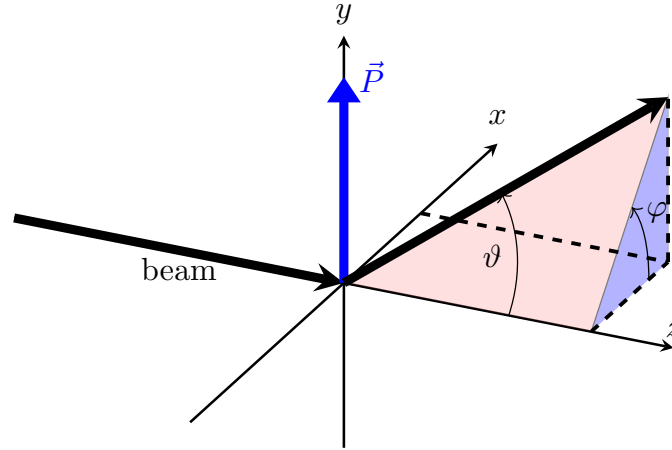
Figure 1: Definition of polar and azimuthal angles ϑ and φ and direction of the beam polarization vector. The target is located at the origin of the coodinate system.

This seems to be an easy task if all other quantities in equation 2 are known. This is often not the case. The cross section, the luminosities and the acceptance are in general not known or at least not to the required precision. Several methods to extract P are discussed in the following. These methods make use of ratios where these unknown factors drop out. To simplify the discussion we consider a fixed polar angle ϑ and an acceptance constant in φ[1]. In this case the events are distributed according to

$$\left\langle \frac{\mathrm{d}N(\varphi)}{\mathrm{d}\varphi} \right\rangle = \frac{N_0}{2\pi}(1 + AP\cos(\varphi)) \tag{3}$$

with $N_0 = \mathcal{L}a \int \frac{\mathrm{d}\sigma(\vartheta)}{\mathrm{d}\Omega}\mathrm{d}\Omega = \mathcal{L}a\sigma$. If we restrict the acceptance to the dark grey area in figure 2

---

[1]In reality this is not the case. How to extract P in this case is discussed in [1].

Table 1: Definitions of variables used in equations 1 and 2.

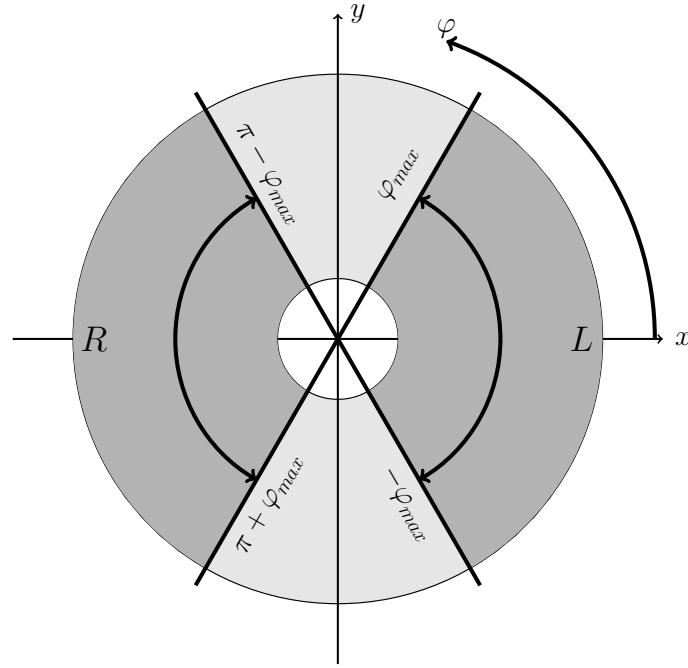| variable	meaning
$N(\vartheta, \varphi)$	number of scattered particles observed
$\langle N(\vartheta, \varphi) \rangle$	expectation value of number of events
$\sigma(\vartheta)$	unpolarized cross section
$\mathcal{L}$	luminosity
$a(\vartheta, \varphi)$	acceptance/efficiency
$\mathrm{d}\Omega$	solid angle
ϑ	polar angle
φ	azimuthal angle, $\varphi = 0$ corresponds to positive x-direction
P	beam polarization
$A(\vartheta)$	analyzing power of scattering process

Figure 2: Events are accepted in the range $-\varphi_{max} < \varphi < \varphi_{max}$ (events scattered to the left) and $\pi + \varphi_{max} < \varphi < \pi - \varphi_{max}$ (events scattered to the right). The particle beam comes out of the plane.

the probability density function (pdf) is given by:

$$p(\varphi) = \frac{1}{4\varphi_{max}}\left(1 + AP\cos(\varphi)\right) \tag{4}$$

3. Different ways to extract P

3.1 Determing P using counting rates

Events are counted in the region between $-\varphi_{max} < \varphi < \varphi_{max}$ and $\pi - \varphi_{max} < \varphi < \pi + \varphi_{max}$ as indicated in figure 2. The expectation values for the number of events in the left part of the detector is given by

$$
\begin{aligned}
\langle N_L \rangle &= \frac{N_0}{2\pi} \int_{-\varphi_{max}}^{\varphi_{max}} \left(1 + AP\cos(\varphi)\right) \mathrm{d}\varphi \\
&= \frac{N_0}{2\pi} \left(2\varphi_{max} + AP \int_{-\varphi_{max}}^{\varphi_{max}} \cos(\varphi)\right) \mathrm{d}\varphi \\
&= \frac{N_0}{\pi}\varphi_{max} \left(1 + AP\frac{\int_{-\varphi_{max}}^{\varphi_{max}} \cos(\varphi)\mathrm{d}\varphi}{2\varphi_{max}}\right) \\
&= \frac{N_0}{\pi}\varphi_{max} \left(1 + AP\overline{\cos(\varphi)}\right).
\end{aligned}
\tag{5}
$$

4

In a similar way one finds

$$\langle N_R \rangle = \frac{N_0}{\pi} \varphi_{max} \left(1 - AP \overline{\cos(\varphi)}\right) .$$ (6)

The following estimator can be used to determine P:

$$\hat{P} = \frac{1}{A \overline{\cos(\varphi)}} \frac{N_L - N_R}{N_L + N_R} .$$ (7)

N_L and N_R are the actual measured number of events. In this ratio $a, \mathcal{L}$ and σ hidden in N_0 drop out. The average $\overline{\cos(\varphi)}$ can be evaluated directly from the data, i.e.

$$
\begin{aligned}
\overline{\cos(\varphi)} &= \frac{\int_{-\varphi_{max}}^{\varphi_{max}} \cos(\varphi) \mathrm{d}\varphi}{2\varphi_{max}} = -\frac{\int_{\pi-\varphi_{max}}^{\pi+\varphi_{max}} \cos(\varphi) \mathrm{d}\varphi}{2\varphi_{max}} \\
&= \frac{\sum_L \cos(\varphi_i) - \sum_R \cos(\varphi_i)}{N_L + N_R} .
\end{aligned}
$$ (8)

Throughout the paper we assume that the number of events is large enough such that expectation values or averages like in equation 8 can be replaced to the corresponding sums over the event sample.

Simple Gaussian error propagation for equation 7 leads to a statistical error for P, assuming $PA \ll 1$,

$$\sigma_{\hat{P}}^2 = \frac{1}{\left(A \overline{\cos(\varphi)}\right)^2 N} ,$$ (9)

where $N = N_L + N_R$ is the total number of events.

In this context it is convenient to define the figure of merit (FOM) as

$$\text{FOM} = \sigma_{\hat{P}}^{-2} = \left(A \overline{\cos(\varphi)}\right)^2 N$$ (10)

The dashed curve in figure 3 shows the FOM as a function of φ_{max}. The curve shows an interesting behavior. Increasing φ_{max} the FOM increases as one would expect by adding more and more events. But from 67 degrees on the FOM decreases. Thus by adding events the statistical error decreases, indicating that this is not the optimal way to analyze the data. The reason is the following. Increasing φ_{max} one adds more data where $|\cos(\vartheta)|$ is small. Looking at equation 2 one realizes that the smaller $|\cos(\vartheta)|$ the less sensitive one is to P. Adding these events dilutes the sample and leads to an increase of the statistical error on $\hat{P}$. The following section discusses a more efficient way to analyze the data.

3.2 Determining P using event weighting

Now we consider the following estimator

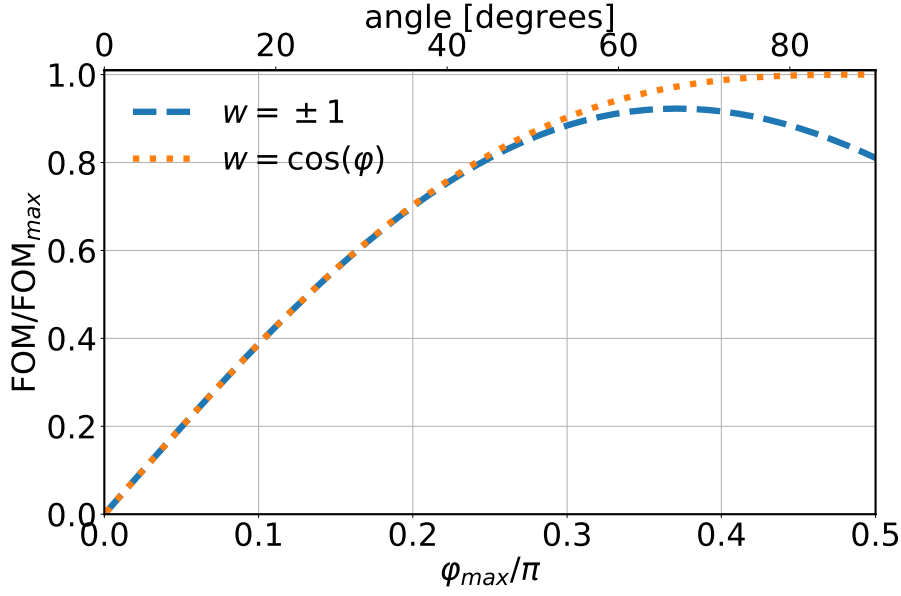$$\hat{P} = \frac{1}{A} \frac{\sum_{L,R} w(\varphi_i)}{\sum_{L,R} w(\varphi_i) \cos(\varphi_i)} .$$ (11)

Figure 3: Figure of merit (FOM) as a function of φ_{max} for the two weighting factors.

Here every event is assigned a, for the moment almost arbitrary, weight factor $w(\varphi)$. The only constraint is that $w(\varphi) = -w(\varphi + \pi)$. In this case equation 11 provides an unbiased estimator for P. This can be verified by looking at the expectation values of the sums in equation 11:

$$AP\left\langle \sum_i w(\varphi_i)\cos(\varphi_i) \right\rangle = \left\langle \sum w(\varphi_i) \right\rangle$$

$$APN \int_{acc} w(\varphi_i)\cos(\varphi_i)p(\varphi)\mathrm{d}\varphi = N \int_{acc} w(\varphi_i)p(\varphi)\mathrm{d}\varphi$$

$$AP \int_{acc} w(\varphi_i)\cos(\varphi_i)(1 + \cancel{PA\cos(\varphi)})\mathrm{d}\varphi = \int_{acc} w(\varphi_i)(\cancel{1} + PA\cos(\varphi))\mathrm{d}\varphi . \qquad (12)$$

Since the cosine function has the same properties as w, $\cos(\varphi) = -\cos(\varphi + \pi)$), the crossed out terms in equation 12 vanish, leaving on both sides $AP \int w(\varphi)\cos(\varphi)\mathrm{d}\varphi = 4\varphi_{max}\overline{w\cos(\varphi)}$. Note that there is a subtle difference between average denoted as $\overline{(\ldots)}$ and expectation values $\langle \ldots \rangle$ discussed further in appendix B.

Two cases are of interest

- simple event counting:

$$w = \begin{cases} 1, & \text{for } -\varphi_{max} < \varphi < \varphi_{max} \\ -1, & \text{for } \pi - \varphi_{max} < \varphi < \pi + \varphi_{max} \\ 0, & \text{otherwise} \end{cases}$$

- event weighting:

$$w = \begin{cases} \cos(\varphi), & \text{for } -\varphi_{max} < \varphi < \varphi_{max} \text{ and } \pi - \varphi_{max} < \varphi < \pi + \varphi_{max} \\ 0, & \text{otherwise} \end{cases}$$
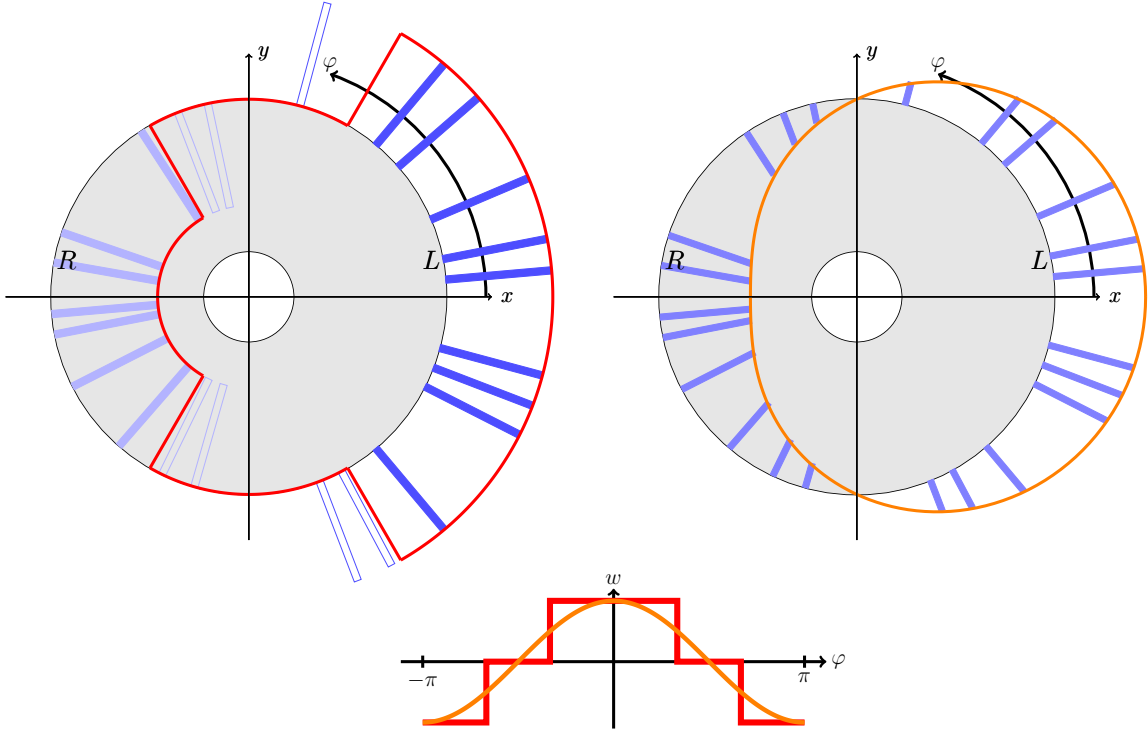
Figure 4: Left: event counting, Right: Weighting events with $\cos(\varphi)$. In case of event counting events are only accepted in the indicated φ range. For the weighting method no cut is applied. The graph at the bottom shows the two weighting factors as a function of φ.

The two cases are indicated in a pictorial way in figure 4. Note that in the case $w = \pm 1$ the estimator in equation 11 coincides with equation 7, if $\overline{\cos\varphi}$ is evaluated from data (see equation 8).

Although both cases lead to an unbiased estimator for P, they lead to different statistical accuracy. This will be discussed in the next subsection.

3.3 Comparison of the two weighting methods

The statistical accuracy or FOM can be obtained by simple error propagation starting from equation 11. Details of the calculations are given in appendix C. Here we just quote the result (again neglecting contributions of order $(AP)^2$ and higher)

$$\text{FOM}^{\pm 1} = NA^2 \overline{\cos(\varphi)}^2, \tag{13}$$

$$\text{FOM}^{\cos(\varphi)} = NA^2 \overline{\cos(\varphi)^2}. \tag{14}$$

The ratio of the two FOMs is

$$\frac{\overline{\cos(\varphi)^2}}{\left(\overline{\cos(\varphi)}\right)^2} \geq 1.$$

I.e. weighting the events with $\cos(\vartheta)$ leads in general to a higher FOM compared to the counting rate asymmetry ($w = \pm 1$). The comparison of the FOMs is shown in figure 3 by the solid and the dashed curve. At small φ_{max} the difference between the two methods is very small. From

$\varphi_{max} \approx 67$ degrees on the FOM of the counting method starts to decrease because the sample is diluted by events at small $\cos(\varphi)$. Using $w = \cos(\varphi)$ the FOM still increases. This means that one can include all events without decreasing the FOM. Events at $\varphi = \pm\pi/2$ receive a weight factor 0 and are thus effectively not counted because they don't carry any information about the polarization. This is also reflected in the fact that the dotted curve in figure 3 has slope zero at $\varphi = \pi/2$. In the counting method one should apply a cut at $\varphi_{max} \approx 67$ degrees in order to maximize the FOM. Comparing the maxima of both curves, the gain in FOM is about 8%.

In the next subsection we show that the choice $w = \cos(\varphi)$ leads to the largest reachable FOM.

3.4 Connection to Maximum Likelihood Estimator

In this subsection it is shown, that the maximum likelihood method leads to same estimator for $\hat{P}$ as the weighting method with $w = \cos(\varphi)$. Since the maximum likelihood estimator is known, at least in the large N limit, to reach the minimum variance bound, this also proves that by weighting with $w = \cos(\varphi)$ one reaches the smallest statistical error or the largest FOM.

The log-likelihood function for the event distribution in equation 2 reads

$$\ell = \frac{1}{4\varphi_{max}} \sum_{i=1}^{N} \ln\left(1 + AP\cos(\varphi)\right) . \tag{15}$$

Since the number of events N is not fixed here, one should use the extended maximum likelihood method. Since N does not depend on the parameter P to be estimated, this is not important here.

The likelihood estimator is derived from

$$\frac{\partial \ell}{\partial P} = \sum_i \frac{A\cos(\varphi_i)}{1 + AP\cos(\varphi_i)} \overset{!}{=} 0 .$$

Assuming $AP \ll 1$, one arrives at an analytic expression

$$\hat{P} = \frac{\sum_i A\cos(\varphi)}{\sum_i (A\cos(\varphi))^2} = \frac{1}{A}\frac{\sum_i \cos(\varphi_i)}{\sum_i \cos^2(\varphi_i)} ,$$

which is identical to the expression for $w = \cos(\varphi)$ in equation 11.

Concerning the FOM, one finds

$$
\begin{aligned}
\text{FOM}^{LH} &= -\left\langle \frac{\partial^2 \ell}{\partial P^2} \right\rangle \\
&= \left\langle \frac{A^2\cos(\varphi)^2}{(1 + AP\cos(\varphi))^2} \right\rangle \\
&= \frac{N}{4\varphi_{max}} \int \frac{A^2\cos(\varphi)^2}{1 + AP\cos(\varphi)}\mathrm{d}\varphi \\
&\approx \frac{N}{4\varphi_{max}} \int A^2\cos(\varphi)^2 \left(1 - AP\cos(\varphi) + (AP\cos(\varphi))^2\right)\mathrm{d}\varphi \\
&= NA^2 \left(\overline{\cos(\varphi)^2} + (AP)^2\overline{\cos(\varphi)^4}\right) \\
&= NA^2 \overline{\cos(\varphi)^2}\left(1 + (AP)^2\frac{\overline{\cos(\varphi)^4}}{\overline{\cos(\varphi)^2}}\right) ,
\end{aligned}
\tag{16}
$$

which agrees with the FOM derived for $w = \cos(\varphi)$ in appendix C using error propagation, if one uses the approximation $(1 + \epsilon) \approx (1 - \epsilon)^{-1}$.

4. Summary and Outlook

In this note it has been shown that simply counting events is not always the best way to analyze data. One can minimize the statistical error on a polarization measurement by assigning to every event an appropriate weight factor. Concerning the statistical accuracy it was shown that a weighting factor can be found which leads to the same estimator as the maximum likelihood method, known to reach the smallest statistical error. There are cases where the likelihood may not be applied directly, e.g. when the pdf is not completely known like in case of unknown acceptance and luminosity factor. In this cases an estimator using event weighting can often still be found [4, 7].

The method can be generalized. If the detector acceptance is not restricted to one value of ϑ, the analyzing power $A(\vartheta)$ can also be included in the optimal weight factor:

$$w = A(\vartheta) \cos(\varphi) \,.$$

In general, if events are distributed according to

$$p(\vec{x}) \propto (1 + \beta(\vec{x})P)$$

where $\vec{x}$ denotes a set of variables, the optimal weight factor is $w = \beta(\vec{x})$ at least for $\beta P \ll 1$. The optimal weight for the arbitrary βP is discussed in [13].

A. From event rates to cross section

This appendix shows the derivation of equation

$$N(\vartheta, \varphi) = a(\vartheta, \varphi)\mathcal{L}\sigma(\vartheta)\left(1 + PA(\vartheta)\cos(\varphi)\right) \tag{17}$$

The number of particles scattered to the left and right are given by

$$N(\varphi = 0) = a\rho\ell\left(n^{\uparrow}\sigma_{\uparrow,L} + n^{\downarrow}\sigma_{\downarrow,L}\right) = a\rho\ell\left(n^{\uparrow}\sigma_L + n^{\downarrow}\sigma_R\right) \tag{18}$$

$$N(\varphi = \pi) = a\rho\ell\left(n^{\uparrow}\sigma_{\uparrow,R} + n^{\downarrow}\sigma_{\downarrow,R}\right) = a\rho\ell\left(n^{\uparrow}\sigma_R + n^{\downarrow}\sigma_L\right) \tag{19}$$

Here $n^{\uparrow}$ ($n^{\downarrow}$) denotes the number of beam particles with spin pointing upwards (downwards). ρ is the target density and ℓ its length. $\sigma_L^{\uparrow}$ is the cross section for a beam particle with spin upwards scattered to the left. Equivalent definitions hold for the other cross sections appearing in equation 18 and 19. The second equal sign hold because of the rotational symmetry:

$$\sigma_{\downarrow,L} = \sigma_{\uparrow,R} =: \sigma_R \quad \text{and} \quad \sigma_{\downarrow,R} = \sigma_{\uparrow,L} =: \sigma_L$$

The beam polarization is defined as

$$P = \frac{n^{\uparrow} - n^{\downarrow}}{n^{\uparrow} + n^{\downarrow}}, \tag{20}$$

and the analyzing power is given by

$$A = \frac{\sigma_L - \sigma_R}{\sigma_L + \sigma_R}. \tag{21}$$

This allows one to write

$$
\begin{aligned}
N(\varphi = 0) &= a\rho\ell\left(n^{\uparrow}\sigma_R + n^{\downarrow}\sigma_L\right) \\
&= a\rho l\frac{1}{2}\left(n^{\uparrow}\sigma_R + n^{\downarrow}\sigma_L + n^{\uparrow}\sigma_R + n^{\downarrow}\sigma_L\right) \\
&= a\rho\ell\frac{1}{2}\left(n^{\uparrow}\sigma_R + n^{\uparrow}\sigma_L + n^{\downarrow}\sigma_R + n^{\downarrow}\sigma_L\right. \\
&\qquad\qquad \left. + n^{\uparrow}\sigma_R - n^{\uparrow}\sigma_L - n^{\downarrow}\sigma_R + n^{\downarrow}\sigma_L\right) \\
&= a\rho\ell\frac{1}{2}\left((n^{\uparrow} + n^{\downarrow})(\sigma_R + \sigma_L) + (n^{\uparrow} - n^{\downarrow})(\sigma_R - \sigma_L)\right) \\
&= a\rho\ell\ \underbrace{\frac{1}{2}(\sigma_R + \sigma_L)}_{=\sigma}\left((n^{\uparrow} + n^{\downarrow}) + (n^{\uparrow} - n^{\downarrow})\underbrace{\frac{\sigma_R - \sigma_L}{\sigma_R + \sigma_L}}_{=A}\right) \\
&= a\rho\ell\ \underbrace{(n^{\uparrow} + n^{\downarrow})}_{=n}\sigma\left(1 + \underbrace{\frac{n^{\uparrow} - n^{\downarrow}}{n^{\uparrow} + n^{\downarrow}}}_{=P}A\right) \\
&= a\mathcal{L}\sigma\left(1 + PA\right).
\end{aligned}
$$

with the luminosity $\mathcal{L} = \rho \ell n$. For the right side ($\varphi = \pi$) on finds correspondingly

$$N(\varphi = \pi) \quad = \quad a\mathcal{L}\sigma \left(1 - PA\right).$$

Again using rotational symmetry and properly normalized we find for arbitrary azimuthal angles equations 2 and 3:

$$\left\langle \frac{\mathrm{d}N(\varphi)}{\mathrm{d}\varphi} \right\rangle = \frac{a\mathcal{L}\sigma}{2\pi}(1 + AP\cos(\varphi)). \tag{22}$$

B. Expectation values and averages

Note that there is a difference between the average denoted be $\overline{(\ldots)}$ and $\langle \ldots \rangle$ for odd powers of the cosine function. The expectation value is defined as

$$
\begin{aligned}
\left\langle \cos(\varphi)^2 \right\rangle &= \int_{acc} \cos(\varphi)^2 p(\varphi)\mathrm{d}\varphi \\
&= \frac{1}{4\varphi_{max}} \int_{acc} \cos(\varphi)^2 \left(1 + AP\cos(\varphi)\right) \mathrm{d}\varphi \\
&= \frac{1}{4\varphi_{max}} \left(\int_{-\varphi_{max}}^{\varphi_{max}} \cos(\varphi)^2 \left(1 + AP\cos(\varphi)\right) \mathrm{d}\varphi \right. \\
&\quad \left. + \int_{\pi-\varphi_{max}}^{\pi+\varphi_{max}} \cos(\varphi)^2 \left(1 + AP\cos(\varphi)\right) \mathrm{d}\varphi \right) \\
&= \frac{\int_{-\varphi_{max}}^{\varphi_{max}} \cos(\varphi)^2 \mathrm{d}\varphi}{2\varphi_{max}} = \overline{\cos(\varphi)^2}
\end{aligned} \tag{23}
$$

In general for even powers one obtains:

$$\left\langle \cos(\varphi)^{2n} \right\rangle = \overline{\cos(\varphi)^{2n}}, \quad n = 0, 1, 2, \ldots. \tag{24}$$

For odd powers on the other hand we find

$$
\begin{aligned}
\left\langle \cos(\varphi) \right\rangle &= \int_{acc} \cos(\varphi) p(\varphi)\mathrm{d}\varphi \\
&= \frac{1}{4\varphi_{max}} \int_{acc} \cos(\varphi) \left(1 + AP\cos(\varphi)\right) \mathrm{d}\varphi \\
&= \frac{1}{4\varphi_{max}} \left(\int_{-\varphi_{max}}^{\varphi_{max}} \cos(\varphi) \left(1 + AP\cos(\varphi)\right) \mathrm{d}\varphi \right. \\
&\quad \left. + \int_{\pi-\varphi_{max}}^{\pi+\varphi_{max}} \cos(\varphi) \left(1 + AP\cos(\varphi)\right) \mathrm{d}\varphi \right) \\
&= 0 + AP \left\langle \cos^2 \varphi \right\rangle.
\end{aligned} \tag{25}
$$

In the evaluation of the FOM in equation 10 and 13

$$\overline{\cos(\varphi)} = \frac{\int_{-\varphi_{max}}^{\varphi_{max}} \cos(\varphi)\mathrm{d}\varphi}{2\varphi_{max}} = \frac{-\int_{\pi-\varphi_{max}}^{\pi+\varphi_{max}} \cos(\varphi)\mathrm{d}\varphi}{2\varphi_{max}} \tag{26}$$

occurred. For odd powers one obtains

$$\left\langle \cos(\varphi)^{2n+1} \right\rangle = AP \left\langle \cos(\varphi)^{2n+2} \right\rangle \neq \overline{\cos(\varphi)^{2n+1}}, \quad n = 0, 1, 2, \ldots. \tag{27}$$

Table 2 list a few expectation values and averages used in the derivations.

	$w = \pm 1$	$w = \cos(\varphi)$
$\langle w \rangle$	$AP \overline{\cos(\varphi)}$	$AP \left\langle \cos(\varphi)^2 \right\rangle$
$\left\langle w^2 \right\rangle$	1	$\left\langle \cos(\varphi)^2 \right\rangle$
$\langle wc \rangle$	$\overline{\cos(\varphi)}$	$\left\langle \cos(\varphi)^2 \right\rangle$
$\left\langle w^2 c \right\rangle$	$AP \left\langle \cos(\varphi)^2 \right\rangle$	$AP \left\langle \cos(\varphi)^4 \right\rangle$
$\left\langle w^2 c^2 \right\rangle$	$\left\langle \cos(\varphi)^2 \right\rangle$	$\left\langle \cos(\varphi)^4 \right\rangle$

Table 2: Various expectation values for the two weighting factors.

C. Error Propagation

Starting from the expression in equation 11 ($w_i \equiv w(\varphi_i)$)

$$\hat{P} = \frac{1}{A} \frac{\sum_i w_i}{\sum_i w_i \cos(\varphi_i)} \,, \tag{28}$$

we define

$$W = \sum_i w_i \quad \text{and} \tag{29}$$

$$V = \sum_i v_i = \sum_i w_i \cos(\varphi_i) \,. \tag{30}$$

First we calculate the covariance matrix for V and W:

$$\text{Cov}(W, V) = \langle WV \rangle - \langle W \rangle \langle V \rangle \tag{31}$$

$$= \left\langle \sum_i w_i \sum_j v_j \right\rangle - \left\langle \sum_i w_i \right\rangle \left\langle \sum_j v_j \right\rangle \tag{32}$$

$$= \left\langle \sum_{i=j} w_i v_i + \sum_{i \neq j} w_i v_j \right\rangle - \left\langle \sum_i w_i \right\rangle \left\langle \sum_j v_j \right\rangle \tag{33}$$

$$= \langle N \rangle \langle WV \rangle + \langle N(N-1) \rangle \langle W \rangle \langle V \rangle - \langle N \rangle^2 \langle W \rangle \langle V \rangle \tag{34}$$

$$= \langle N \rangle \langle WV \rangle + \left(\left\langle N^2 \right\rangle - \langle N \rangle - \langle N \rangle^2 \right) \langle W \rangle \langle V \rangle \,. \tag{35}$$

Assuming a Poisson distribution for N, the term in parentheses vanishes. Thus the elements of the covariance matrix of W and V are given by:

$$\text{cov}(W, V) = N \left\langle w^2 \cos(\varphi) \right\rangle = \sum_i w_i^2 \cos(\varphi_i) \,,$$

$$\sigma_W^2 = N \left\langle w^2 \right\rangle = \sum_i w_i^2 \,,$$

$$\sigma_V^2 = N \left\langle w^2 \cos(\varphi)^2 \right\rangle = N \sum_i w_i^2 \cos(\varphi_i)^2 \,.$$

Error propagation on equation 28 leads to (using the shorthand notation $c = \cos(\varphi)$)

$$
\begin{aligned}
\sigma_{\hat{P}}^2 &= \left(\frac{\partial \hat{P}}{\partial W}, \frac{\partial \hat{P}}{\partial V} \right) \begin{pmatrix} \sigma_W^2 & \text{cov} \\ \text{cov} & \sigma_V^2 \end{pmatrix} \begin{pmatrix} \frac{\partial \hat{P}}{\partial W} \\ \frac{\partial \hat{P}}{\partial V} \end{pmatrix} \\
&= \frac{N}{A^2} \left(\frac{1}{V}, -\frac{W}{V^2} \right) \begin{pmatrix} \langle w^2 \rangle & \langle w^2 c \rangle \\ \langle w^2 c \rangle & \langle w^2 c^2 \rangle \end{pmatrix} \begin{pmatrix} \frac{1}{V} \\ -\frac{W}{V^2} \end{pmatrix} \\
&= \frac{1}{NA^2} \left(\frac{\langle w^2 \rangle}{\langle wc \rangle^2} - 2 \frac{\langle w^2 c \rangle \langle w \rangle}{\langle wc \rangle^3} + \frac{\langle w^2 c^2 \rangle \langle w \rangle^2}{\langle wc \rangle^4} \right)
\end{aligned}
\tag{36}
$$

In the last line we replaced $W = N \langle w \rangle$ and $V = N \langle wc \rangle$.

Given the properties of w ($w(\varphi) = -w(\pi + \varphi)$), one finds according to table 2:

$$
\langle w \rangle = AP \langle wc \rangle , \tag{37}
$$
$$
\langle w^2 c \rangle = AP \langle w^2 c^2 \rangle . \tag{38}
$$

This leads to

$$
\text{FOM} = NA^2 \frac{\langle wc \rangle^2}{\langle w^2 \rangle} \left(1 - (AP)^2 \frac{\langle w^2 c^2 \rangle}{\langle w^2 \rangle} \right)^{-1} . \tag{39}
$$

Finally for the FOMs read

$$
\text{FOM}^{\pm 1} = NA^2 \overline{\cos(\varphi)}^2 \left(1 - (AP)^2 \overline{\cos(\varphi)^2} \right)^{-1} , \tag{40}
$$

$$
\text{FOM}^{\cos(\varphi)} = NA^2 \overline{\cos(\varphi)^2} \left(1 - (AP)^2 \frac{\overline{\cos(\varphi)^4}}{\overline{\cos(\varphi)^2}} \right)^{-1} . \tag{41}
$$

To leading order equations 40 and 41 agree with the results given in equations 13 and 14. Moreover the FOM for the case $w = \cos(\varphi)$ coincides with the FOM derived for the maximum likelihood method in equation 16.

As explained in section 3.2 the estimator

$$
\hat{P} = \frac{1}{\overline{\cos(\varphi)}} \frac{N_L - N_R}{N_L + N_R}
$$

leads to

$$
\text{FOM} = NA^2 \overline{\cos(\varphi)}^2 \left(1 - (AP)^2 \overline{\cos(\varphi)^2} \right)^{-1} ,
$$

if $\overline{\cos(\varphi)}$ is evaluated from a sum over events.

In case $\overline{\cos(\varphi)}$ is supposed to be known, e.g. from the integral $\int \cos(\varphi) d\varphi$, the FOM is slightly different at order $(AP)^2$ where now a factor $\overline{\cos(\varphi)}^2$ instead of $\overline{\cos(\varphi)^2}$ appears.

$$
\text{FOM} = NA^2 \overline{\cos(\varphi)}^2 \left(1 - (AP)^2 \overline{\cos(\varphi)}^2 \right)^{-1} .
$$

References

[1] Pretz, J. and Müller, F., "Extraction of Azimuthal Asymmetries using Optimal Observables," *Eur. Phys. J.*, vol. C79, no. 1, p. 47, 2019.

[2] B. Abi *et al.*, "Measurement of the Positive Muon Anomalous Magnetic Moment to 0.46 ppm," *Phys. Rev. Lett.*, vol. 126, no. 14, p. 141801, 2021.

[3] G. W. Bennett *et al.*, "Statistical equations and methods applied to the precision muon (g-2) experiment at BNL," *Nucl. Instrum. Meth. A*, vol. 579, pp. 1096–1116, 2007.

[4] M. Alekseev *et al.*, "Direct Measurement of the Gluon Polarisation in the Nucleon via Charmed Meson Production," 2008.

[5] J. Pretz and J.-M. Le Goff, "Simultaneous Determination of Signal and Background Asymmetries," *Nucl. Instrum. Meth.*, vol. A602, pp. 594–596, 2009.

[6] J. Pretz, "Improved Method to extract Nucleon Helicity Distributions using Event Weighting," *JINST*, vol. 12, no. 02, p. P02007, 2017.

[7] F. Müller *et al.*, "Measurement of deuteron carbon vector analyzing powers in the kinetic energy range 170-380 MeV," *Eur. Phys. J. A*, vol. 56, no. 8, p. 211, 2020.

[8] M. Diehl and O. Nachtmann, "Optimal observables for the measurement of three gauge boson couplings in e+ e- —> W+ W-," *Z. Phys. C*, vol. 62, pp. 397–412, 1994.

[9] D. Atwood and A. Soni, "Analysis for magnetic moment and electric dipole moment form-factors of the top quark via $e + e^- \to t\bar{t}$," *Phys. Rev.*, vol. D45, pp. 2405–2413, 1992.

[10] W. Bernreuther, L. Chen, and O. Nachtmann, "Electric dipole moment of the tau lepton revisited," *Phys. Rev. D*, vol. 103, no. 9, p. 096011, 2021.

[11] A. Bodek, "A simple event weighting technique for optimizing the measurement of the forward-backward asymmetry of Drell-Yan dilepton pairs at hadron colliders," *Eur. Phys. J. C*, vol. 67, pp. 321–334, 2010.

[12] G. G. Ohlsen and P. W. Keaton, "Techniques for measurement of spin-1/2 and spin-1 polarization analyzing tensors," *Nucl. Instrum. Meth.*, vol. 109, pp. 41–59, 1973.

[13] J. Pretz, "Comparison of methods to extract an asymmetry parameter from data," *Nucl. Instrum. Meth.*, vol. A659, pp. 456–461, 2011.