

DESY-Bibliothek
17. AUG. 1966

DEUTSCHES ELEKTRONEN - SYNCHROTRON **DESY**

DESY 66/17
Juni 1966
Experimente

Die Anwendung der Maximum Likelihood Methode
in der Hochenergiephysik

von

W. P. Swanson

Deutsches Elektronen-Synchrotron DESY, Hamburg

2 HAMBURG 52 · NOTKESTIEG 1

Die Anwendung der Maximum Likelihood Methode
in der Hochenergiephysik

W. P. Swanson

Deutsches Elektronen-Synchrotron DESY, Hamburg

15. März 1966

- I. Allgemeines über die MLM.
- II. Beispiele für direkte Anwendung der MLM.
- III. Maximalisierungs-Verfahren.
- IV. Diskrete Hypothesen; Likelihood-Verhältnisse.

Abstract:

A lecture on applications of the Maximum Likelihood Method in high energy physics. Both the ordinary and extended MLM are introduced. Four practical examples are given:

- 1) The analysis of proton polarization by scattering on carbon.
- 2) The determination of spin density matrix elements.
- 3) The determination of amounts, masses and widths of resonances plotted on a triangle plot.
- 4) The determination of branching ratios in a two-step decay process.

Minimization processes are described in detail. Likelihood ratios for hypothesis discrimination are defined and the meaning of these ratios is discussed. References to other examples in the literature are given.

I. Allgemeines über die MLM:

1)
Die MLM kann folgendermassen beschrieben werden:

Man schreibt eine theoretische Verteilungsformel für den betrachteten physikalischen Prozess. Diese Formel sagt also eine Verteilung physikalisch messbarer Werte (z.B. Winkel oder invariante Massen) voraus.

Schreiben wir diese Verteilung als $P(\vec{X}, \vec{A})$, wo P die "probability" oder Wahrscheinlichkeit bedeutet. Der \vec{X} -Vektor besteht aus den messbaren Werten (z.B. Winkeln etc.), deren Verteilungen vorausgesagt sind.

P ist also die Wahrscheinlichkeit, mit der ein Ereignis mit seinen einmaligen messbaren Werten \vec{X} stattfindet.

Der Vektor \vec{A} enthält die verschiedenen Parameter der Theorie, die vielleicht unbekannt sind. Es ist der Zweck der MLM, "beste" Werte von diesen Parametern zu finden.

Wenn ein Blasenkammer-Experiment abgeschlossen ist, hat man eine Liste von Ereignissen mit je einer Gruppe von bekannten Parametern, die den Vektor \vec{X} für jedes dieser Ereignisse festlegt. Die Wahrscheinlichkeit, mit der man ein einzelnes Ereignis mit dem Vektor \vec{X}_i hat, ist $P(\vec{X}_i, \vec{A})$, und die Wahrscheinlichkeit, mit der man das ganze Experiment beobachtet hat, ist also das Produkt von sämtlichen einzelnen Wahrscheinlichkeiten.

$$\mathcal{L}(\vec{A}) = \prod_i P(\vec{X}_i, \vec{A})$$

Eine solche gesamte Wahrscheinlichkeit heisst "likelihood" und ist eine Funktion von \vec{A} .

Wenn P auf das Einheitsvolumen im \vec{X} -Raum normalisiert ist ($\int P d\vec{X} = 1$), haben wir den ersten Teil des Problems erledigt. Um zu betonen, daß P normalisiert ist, schreibe ich das Integral im Nenner.

$$\mathcal{L}(\vec{\Lambda}) = \prod_i \frac{P(\vec{X}_i, \vec{\Lambda})}{\int P(\vec{X}, \vec{\Lambda}) d\vec{X}} \quad (1)$$

Dieses Integral ist über genau den Bereich zu erstrecken, der beobachtbar ist, z.B. über genau den Raumwinkelbereich, der in dem Experiment zur Verfügung stand.

In manchen Fällen ist sowohl die Beobachtungswahrscheinlichkeit als auch die Erzeugungswahrscheinlichkeit der Ereignisse eine Funktion von \vec{X} . Dann muß diese Beobachtungswahrscheinlichkeit als multiplikativer Faktor in $P(\vec{X}, \vec{\Lambda})$ eingeschlossen werden. Referenz 2) enthält ein kompliziertes Beispiel eines solchen Problems.

Manchmal ist einer der A-Parameter nur ein multiplikativer Faktor, der die "rate" oder Häufigkeit des betrachteten Prozesses beherrscht. Solch ein Parameter wäre z.B. der totale Wirkungsquerschnitt oder die "transition rate". Offensichtlich kann ein multiplikativer Parameter solch eine "likelihood function" nicht beeinflussen, weil er gleichzeitig im Zähler und Nenner erscheint. Außerdem wäre es sinnlos, die Verteilung zu normalisieren.

In diesem Fall normalisieren wir die Wahrscheinlichkeit P überhaupt nicht, sondern betrachten einen weiteren Faktor, nämlich die Wahrscheinlichkeit, mit der Ereignisse nicht auftreten, wo sie auch in Wirklichkeit in diesem Experiment nicht aufgetreten sind. Dieser Faktor ist genau

$$P_{\text{nicht}}(\vec{\Lambda}) = \exp(-\int P(\vec{X}, \vec{\Lambda}) d\vec{X}) \quad (2)$$

und die gesamte "likelihood function" in diesem Fall ist

$$\mathcal{L}(\vec{\Lambda}) = \left[\prod_i P(\vec{X}_i, \vec{\Lambda}) \right] \cdot \exp(-\int P(\vec{X}, \vec{\Lambda}) d\vec{X})$$

Diese Methode heißt die "extended MLM". 1)

Sie werden vielleicht überrascht sein, diesen exponentiellen Faktor zu sehen. Die folgenden Betrachtungen sind eine heuristische Rechtfertigung dieses Faktors. Wenn eine "rate" zu bestimmen ist, stellen wir uns vor, dass die Wahrscheinlichkeit eines Prozesses eine normalisierte (auf eins) Verteilung $W(\vec{X}, \vec{A})$ sei, multipliziert mit einer Variablen A_0 , die die "rate" beherrscht. Lassen wir diese Variable die Anzahl vorhergesagter Ereignisse sein.

$$P(\vec{X}, A_0, \vec{A}) = A_0 W(\vec{X}, \vec{A})$$

Diese Wahrscheinlichkeit hat natürlich die Eigenschaft, daß sie, wenn sie über \vec{X} integriert wird, die vorausgesagte Anzahl A_0 ergibt. Betrachten wir A_0 als einen unbekanntem Wert, der zusammen mit \vec{A} zu bestimmen ist.

Die overall "likelihood" ist also das Produkt zweier "likelihoods", nämlich der "likelihood", mit der die N (die beobachtete Anzahl) Ereignisse auftreten, und der "likelihood", mit der die N Ereignisse so verteilt sind, wie sie in Wirklichkeit verteilt sind. Der erste erwähnte Faktor wird von der Poissonschen Verteilung vorausgesagt:

$$\mathcal{L}(A_0, N) = \frac{A_0^N e^{-A_0}}{N!} ,$$

wo A_0 in dieser Formel der theoretische Mittelwert ist, der zu bestimmen ist. N dagegen ist die Anzahl der Ereignisse, die in der Tat beobachtet worden sind. Der zweite erwähnte Faktor ist einfach

$$\prod_{i=1}^N W(\vec{X}_i, \vec{A}) .$$

Zusammen liefern diese Faktoren

$$\mathcal{L}(\vec{X}, A_0, \vec{A}) = \left| \frac{A_0^N e^{-A_0}}{N!} \right| \times \left| \prod_i^N W(\vec{X}_i, \vec{A}) \right|$$

$$= \frac{1}{N!} \left| \prod_i^N A_0 W(\vec{X}_i, \vec{A}) \right| e^{-A_0} \quad (3)$$

$$= \frac{1}{N!} \left| \prod_i^N P(\vec{X}_i, A_0, \vec{A}) \right| e^{-\int P(\vec{X}, A_0, \vec{A}) d\vec{X}}$$

Die Fakultät $N!$ ist eine unbedeutende Konstante, die die Bestimmung von A_0 bzw. \vec{A} nicht beeinflusst, und sie muß nicht berücksichtigt werden.

Beachten Sie, daß, wenn es keinen "Winkelverteilungs-Term" $W(\vec{X}, \vec{A})$ gibt, die Maximalisierung von Gleichung (3) das triviale Resultat $A_0 = N$ liefert. $W(\vec{X}, \vec{A})$ ändert also A_0 ab. Im allgemeinen wird der geänderte Wert besser sein als $A_0 = N$, weil mehr Information zu seiner Bestimmung angewendet worden ist.

Ich betone, daß die zuerst erwähnte likelihood function zu verwenden ist, wenn es angebracht ist, die Wahrscheinlichkeit auf eins zu normalisieren und die "extended MLM", wenn eine "rate" oder ein Wirkungsquerschnitt unbekannt und zu bestimmen ist.

Wenn wir eine große Anzahl von Ereignissen haben und die entsprechenden Wahrscheinlichkeiten miteinander multiplizieren, kann dieses Verfahren leicht "machine overflow" verursachen. Aus diesem Grunde berechnet man normalerweise den Logarithmus der likelihood function $w(\vec{A}) = \ln \mathcal{L}(\vec{A})$.

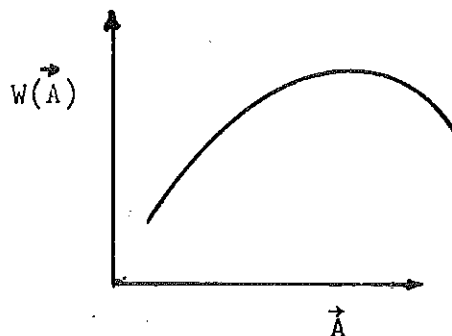
Bis jetzt haben wir uns noch nicht mit den A-Parametern befaßt. Diese sind die Werte von physikalischem Interesse, die wir aus unserem Experiment bestimmen wollen.

Um sie zu bestimmen, maximalisieren wir die gesamte Wahrscheinlichkeit $\mathcal{L}(\vec{A})$, mit der das Experiment, das in Wirklichkeit geschehen ist, vorausgesagt wird. Was dies praktisch bedeutet, ist, daß wir Werte von \vec{A} suchen mit irgendeinem möglichen Mittel, die $w(\vec{A})$ maximalisieren. Diese Werte sind die besten Werte der Parameter, die das Experiment geben kann.

Die Maximalisierung von w ist nur ein technisches Problem, das ich später diskutieren werde. Wenn dieses Maximalisierungsverfahren angewendet worden ist, können wir Abschätzungen der Standard - Abweichungen von den A-Parametern, oder, was noch nützlicher ist, die ganze Fehlermatrix

$$\langle \Delta A_i \Delta A_j \rangle$$

nach einem einfachen Verfahren berechnen. w ist natürlich eine Funktion von \vec{A} .



An der Stelle des Maximums sind alle ersten Ableitungen natürlich Null, aber die zweiten Ableitungen stehen im Zusammenhang mit der Schärfe des Maximums. Noch genauer gesagt: Wenn wir nur einen Parameter A haben, wird der Fehler in A so berechnet:

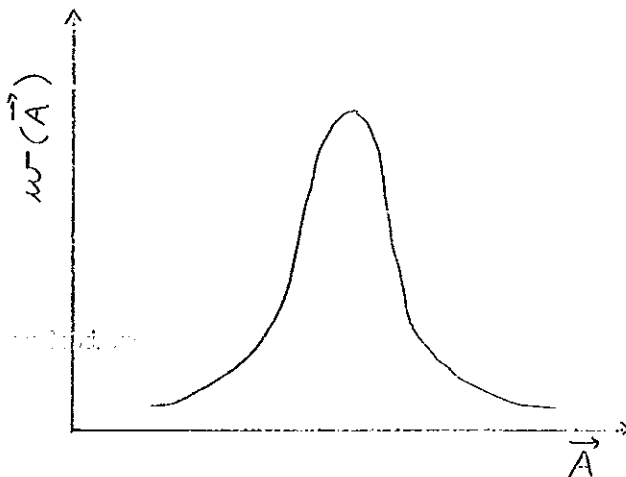
$$\langle (\Delta A)^2 \rangle = - \left(\frac{d^2 w}{d A^2} \right)^{-1}$$

und wenn wir mehrere Parameter haben, haben wir

$$\langle \Delta A_i \Delta A_j \rangle = - \left(\frac{\partial^2 w}{\partial A_i \partial A_m} \right)^{-1}_{ij}$$

Das heisst, die Fehler-Matrix ist gleich der Inversen der Matrix der zweiten Ableitungen.

Diese letzten Beobachtungen nehmen an, dass die Likelihood Funktion ähnlich einer Gaussischen Kurve in allen Variablen A ist.



Dies wird der Fall sein, wenn es genügend viele Ereignisse gibt, und die Form der Likelihood Funktion gescheit gewählt ist.

Damit ist die allgemeine Diskussion über die MLM vollständig erledigt. Die Probleme sind zweifach: Erstens muss man seine Wahrscheinlichkeit P etwas gescheit ansetzen. Diese Möglichkeit ist nur durch die Kenntnis oder Unkenntnis der zugrundeliegenden Physik beschränkt. Zweitens besteht das Problem der Maximalisierung der

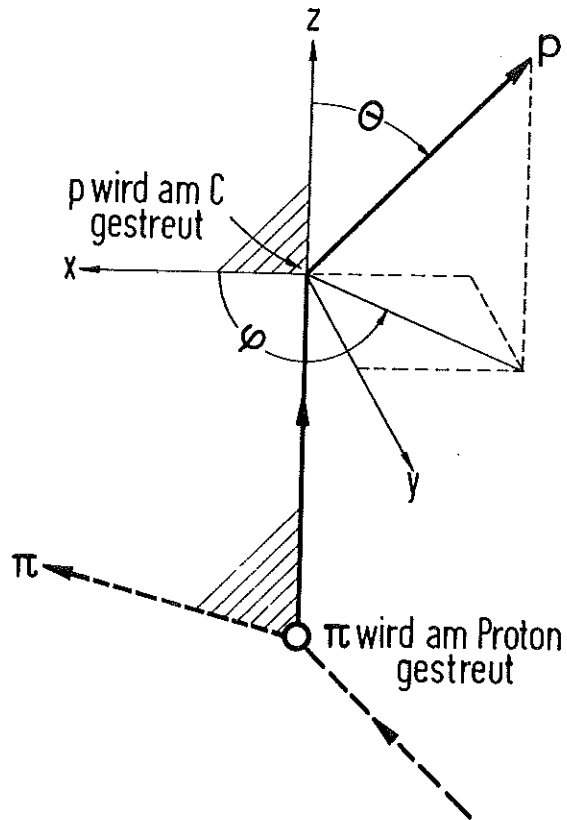


Abb.1

parallel zur Streuungs-Ebene ist, dann sind θ und ψ die normalen Polar-Koordinaten des gestreuten Protons. Es kann sein, dass die Protonen eine Spin-Polarisation haben, mit einem Polarisations-Vektor, der eine bestimmte Richtung relativ zur Streuebene (senkrecht dazu) hat.

Die Polarisation ist so definiert:

$$p = \frac{\text{AUS} - \text{EIN}}{\text{AUS} + \text{EIN}}$$
, wo EIN und AUS die Anzahl der Ereignisse mit Spin-Richtung "in die Tafel Ebene hinein" bzw. "aus der Tafel Ebene heraus" bedeuten sollen.

Die Wahrscheinlichkeit, mit der das Proton vom Kohlenstoff-Kern mit den Winkeln θ_λ und ψ_λ gestreut wird, ist proportional zu:

$$P(\theta_i, \psi_i) = 1 + p \cdot a(\theta_i) \cdot \cos \psi_i,$$

wo p (die Polarisation) zu bestimmen und $a(\theta)$ (die Analysierbarkeit) eine bekannte Funktion von θ für Proton-Kohlenstoff-Streuung ist.

Bemerken Sie, dass $P(\theta, \psi)$ schon auf das Einheitsvolumen im Raumwinkel normalisiert ist. Also ist die Likelihood nichts anderes als das Produkt über alle Ereignisse von $P(\theta_i, \psi_i)$.

$$\mathcal{L}(p) = \prod_i P(\theta_i, \psi_i)$$

Um den besten Wert der Polarisation zu bestimmen, probieren wir nur verschiedene Werte von p von 0 bis 100 %, bis wir den maximalen Wert von $\mathcal{L}(p)$ gefunden haben.

Beispiel 2

Nehmen wir an, dass ein Rho-Meson in irgendeiner Reaktion erzeugt worden ist. Sagen wir, es ist in der Reaktion $\gamma p \rightarrow \rho p \rightarrow p \pi^+ \pi^-$ erzeugt worden. Danach zerfällt das ρ in $\pi^+ \pi^-$. Das ρ ist in einem Zustand von Spin-Ausrichtung, der völlig beschrieben werden kann mit einer Spin-Dichte-Matrix ⁴⁾ ρ_{ij}

Im Ruhesystem des ρ haben die Zerfallprodukte des ρ eine Verteilung

$$W(\theta, \varphi; \rho_{ij}) = \frac{3}{4\pi} \left\{ \frac{1}{2} (1 - \rho_{00}) + \frac{1}{2} (3\rho_{00} - 1) \cos^2 \theta - \rho_{1,-1} \sin^2 \theta \cos 2\varphi - \sqrt{2} \operatorname{Re} \rho_{10} \sin 2\theta \cos \varphi \right\}$$

θ und φ sind natürlich Parameter der einzelnen beobachteten Ereignisse. Die ρ_{ij} sind die Parameter, die man aus dem Experiment bestimmen möchte, um Information über den Erzeugungsmechanismus zu bekommen. Um dies zu leisten, maximalisiert man diese Summe über alle beobachteten Ereignisse k :

$$w(\rho_{ij}) = \ln \mathcal{L}(\rho_{ij}) = \sum_k \ln W(\theta_k, \varphi_k; \rho_{ij})$$

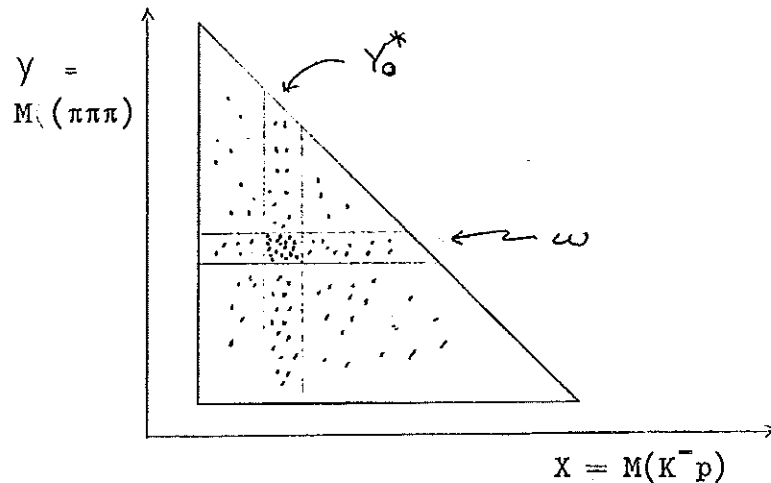
Bemerken Sie, dass $W(\theta, \varphi)$ schon normalisiert ist mit dem Faktor $3/4\pi$.

Dies Problem könnte auch mit der MLS gelöst werden, wenn wir die Daten in zweidimensionalen Intervallen in θ und φ teilten.

Dieses Verfahren wäre aber weniger elegant, würde grössere Fehler liefern und wäre von der Willkürlichkeit der Intervalleinteilung abhängig. Eine vollständige Behandlung dieses Problems ist in einem Artikel von J. Friedmann und R. Ross zu finden ⁵⁾. Andere Methoden, mit denen man die Spindichte-Matrixelemente bestimmen kann, sind in einem Artikel von N. Schmitz zu finden ⁶⁾.

Beispiel 3

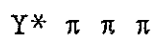
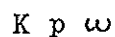
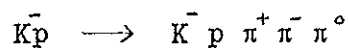
Nehmen wir an, dass wir die Reaktion ($Kp \rightarrow Kp\pi\pi$) untersuchen. In Illinois haben wir in dieser Reaktion für eine Gesamtschwerpunktenergie von 2.5 GeV starke Υ_0^* (1520) und ω Erzeugung gefunden ⁷⁾. In solch einem Fall ist es nützlich, den Endzustand mit einem Triangle plot zu zeigen.



In diesem plot ist die Masse der Kp -Kombination als Abszisse aufgetragen und die $(\pi\pi)$ -Masse als Ordinate.

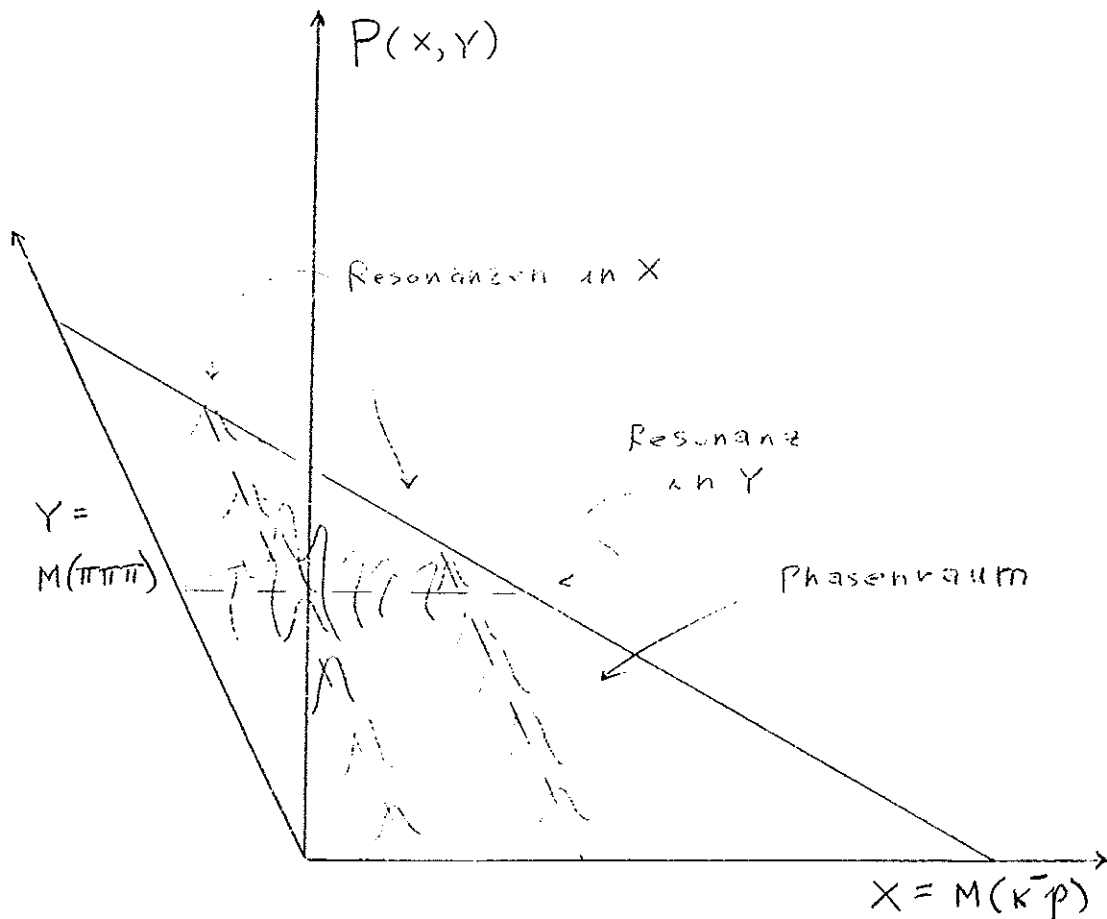
Dieser plot zeigt ganz klare Anhäufungen, die den zwei erwähnten Resonanzen entsprechen. Das dichte Gebiet im overlap-Bereich deutet an, dass wir gleichzeitige Erzeugung von diesen beiden Resonanzen haben.

Es gibt also mindestens diese vier möglichen Prozesse



Sicher gibt es Ereignisse, die nach reinem Phasenraum verteilt sind und sicher gibt es gleichzeitige $Y^* \omega$ - Erzeugung und vielleicht andere Endzustände, aber wie oft kommen sie vor?

Um dieses Problem zu lösen, haben wir eine Wahrscheinlichkeit als Funktion der zwei Koordinaten, die da aufgetreten sind, geschrieben. Stellen Sie sich vor, dass eine dritte Achse, die Wahrscheinlichkeitsachse, aus der X-Y-Ebene herausragt. Dann ist der Phasenraum ein gebeugtes Dach, das das Dreieck deckt, und am Rande verschwindet.



Nach Hagedorn ⁸⁾ ist die vom Phasenraum vorausgesagte Wahrscheinlichkeit

$$P_{PR}(X,Y) = \left[R_2(X; m_1, m_2) \cdot R_3(Y; m_3, m_4, m_5) \cdot R_2(E; X, Y) \right] / \int \dots dX dY$$

Die Wahrscheinlichkeit einer Resonanz in X, d.h. zwischen den K^- und p, ist proportional der Phasenraumwahrscheinlichkeit P, multipliziert mit einem Breit-Wignerschen Faktor. Diese Wahrscheinlichkeitsfläche sieht wie ein glattes Gebirge aus.

$$P_X(X,Y) = \frac{P_{PR}(X,Y)}{\left[(X-M_R)^2 + (\sigma/2)^2 \right]} \int \dots dXdY$$

Die Wahrscheinlichkeit einer Resonanz in Y, d.h. unter den drei Pionen, wird analog in Y geschrieben.

Gleichzeitige Resonanz-Erzeugung ähnelt einem Finger, der aus der Ebene herausragt. Lediglich ist die Phasenraumwahrscheinlichkeit P_{PR} multipliziert mit zwei Breit-Wignerschen Faktoren, die X bzw. Y entsprechen.

Alle Verteilungen sind auf das Einheitsvolumen in X und Y durch die Integrale normalisiert. Die Gesamtwahrscheinlichkeit, mit der ein Ereignis an der Stelle X,Y stattfindet, ist also eine Funktion von $A_{PR} \dots$ durch A_{XY} , bei der diese A's die Anzahl der Ereignisse sind, die tatsächlich der entsprechenden Verteilung gehören. Die Verteilungen sind die, die gerade definiert worden sind.

$$\begin{aligned}
 & P_i (A_{PR}, A_X, A_Y, A_{XY}; X_i, Y_i) \\
 &= A_{PR} \cdot \left\{ \begin{array}{l} \text{Normalisierte} \\ \text{Wahrscheinlichkeit nach Phasenraum} \end{array} \right\} \\
 &+ A_X \cdot \left\{ \begin{array}{l} \text{Normalisierte} \\ \text{Wahrscheinlichkeit für X-Resonanz Erzeugung} \end{array} \right\} \\
 &+ A_Y \cdot \left\{ \begin{array}{l} \text{Normalisierte} \\ \text{Wahrscheinlichkeit für Y-Resonanz Erzeugung} \end{array} \right\} \\
 &+ A_{XY} \cdot \left\{ \begin{array}{l} \text{Normalisierte} \\ \text{Wahrscheinlichkeit für gleichzeitige Resonanz Erzeugung} \end{array} \right\}
 \end{aligned}$$

In diesem Falle wollen wir "rates" (Anzahl) bestimmen. Also wenden wir die "extended MLM" an und schreiben

$$\mathcal{L}(A_{PR} \dots A_{XY}) = \left[\prod_i P_i \right] e^{-(A_{PR} + \dots + A_{XY})}$$

Da jede der Verteilungen für sich in dieser Formel normalisiert ist, ist das Integral der Gesamtwahrscheinlichkeit nur die Totalanzahl der Ereignisse $A_{PR} + \dots + A_{XY}$. Diese Summe ist ungefähr - - jedoch im Allgemeinen nicht genau der beobachteten Anzahl gleich. Wenn \mathcal{L} maximalisiert ist, haben wir die gewünschten Werte der verschiedenen A's gefunden.

Sie können sich leicht vorstellen, dass wir, wenn wir Anpassungen an die Histogramme gemacht hätten, wir Informationen über die exakte Stelle jedes einzelnen Ereignisses verloren hätten. So würden die Fehler dementsprechend grösser werden.

Dieses Programm, das TRIKOT ⁹⁾ (eine Abkürzung von TRIKOTNIK) heisst, ist in Illinois geschrieben und von Herrn Behrens für das DESY-Monitor-system übersetzt worden. Ein ähnliches Programm wurde von Friedmann und Ross in Berkeley aufgestellt, welches mehr mögliche Endzustände behandeln kann ¹⁰⁾. Herr P. Söding erstellt jetzt ein ähnliches Programm für den Prozess $\gamma p \rightarrow p\pi\pi$ ¹¹⁾.

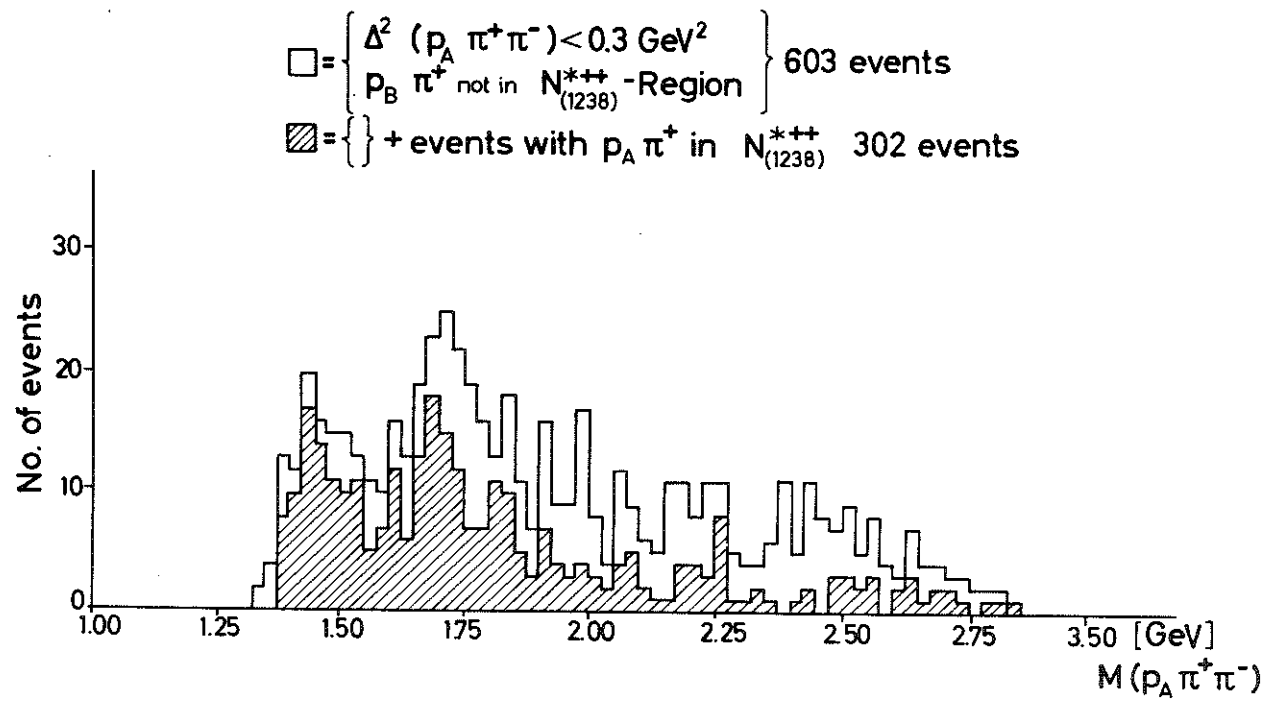
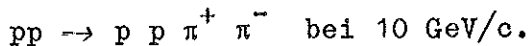


Abb.2

Beispiel 4

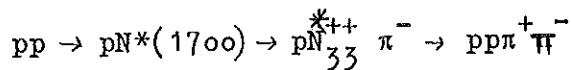
Betrachten wir die Reaktion, die von der Cambridge-Hamburg-Collaboration ¹²⁾ untersucht wird, nämlich



Eine bedeutende Eigenschaft dieser Reaktion ist zwei "enhancements" in der $p\pi^+\pi^-$ -Massen-Verteilung um 1450 MeV bzw. 1700 MeV. (Abb.2)

Das schraffierte Histogramm besteht aus den Ereignissen, deren $p\pi^+$ -Kombination im $N^*_{3/2^-}$ Isobar Massenbereich liegt. Diese Daten deuten also an, dass wir neben anderen Dingen auch eine Resonanz um 1700 MeV erzeugen, die hinterher in $N^{*++}\pi^-$ zerfällt.

Also ist die Reaktionskette folgende:



Wir möchten das Verzweigungsverhältnis $N^*(1700) \rightarrow \frac{N^*_{33} \pi}{N^*_{33} \pi + p\pi^+\pi^-}$

aus diesen Daten abschätzen. Deshalb schreiben wir eine entsprechende drei-dimensionale Massenverteilung.

$$P(M(p_B \pi^+ \pi^-), M(p_B \pi^+), M(p_B \pi^-)) =$$

$$\left\{ \frac{p \cdot M(p_B \pi^+) \cdot M(p_B \pi^-)}{E} \right\} \left\{ \frac{1}{[(M(p\pi^+\pi^-) - 1700)^2 + (\Gamma/2)^2]} \right\} \left\{ \frac{1}{1} \right\}$$

↑ Phasenraum-Faktor

↑ Erzeugung des $N^*(1700)$

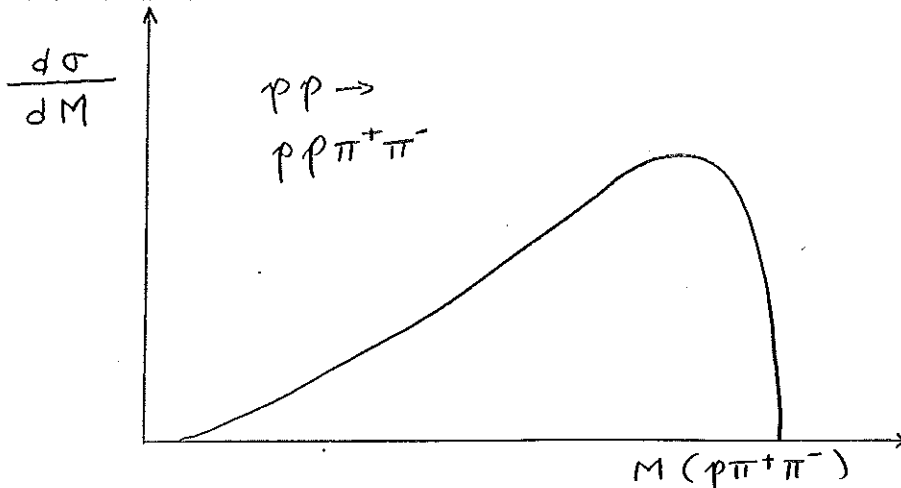
↑ unkorrelierter Zerfall des $N^*(1700)$

für kein $N^*(1700)$ $\left\{ \frac{1}{1} \right\}$

$$\left\{ \frac{1}{[(M(p\pi^+) - 1238)^2 + (\Gamma/2)^2]} \right\}$$

für $N^*(1700) \rightarrow N^{*++}(1238) \pi^-$

Der erste Faktor ist ein Phasenraumfaktor. Hier sind E und \mathcal{P} die Gesamtschwerpunktenergie bzw. der Schwerpunktimпульs der $\rho \pi^+ \pi^-$ -Kombination. Dieser Faktor, wenn er über $M(\rho_B \pi^-)$ und $M(\rho_B \pi^+)$ integriert worden ist, sagt eine $\pi^+ \pi^-$ Massenverteilung voraus, die wohl bekannt ist:



Die Erzeugung einer $N^*(1700)$ Resonanz ist proportional zum zweiten Faktor, einem Breit-Wignerschen Faktor.

Daneben schreiben wir dann die Wahrscheinlichkeit, mit der die $N^*(1700)$ Resonanz zerfällt. Wenn der Zerfall ganz unkorreliert ist, (d.h. keinen Isobarzwischenzustand hat), ist diese Wahrscheinlichkeit unabhängig von der invarianten Masse der Zerfallprodukte π^+ . Also schreiben wir nur die Konstante "eins". Aber wenn die Resonanz durch $N^*\pi$ zerfällt, ist der dritte Faktor nicht mehr eins, sondern ein zweiter Breit-Wignerscher Faktor, der den Zerfall beschreibt. Also ersetzen wir die "Eins" durch diesen Faktor.

Ausserdem haben wir einen Untergrund, der betrachtet werden muss. Um diesen Untergrund zu behandeln, schreiben wir eine Konstante (eins), wo der erste Breit-Wignersche Term steht. Aber beachten wir noch die zwei Zerfallsmöglichkeiten, da der Untergrund auch entweder unkorreliert sein oder $N^*(1238)$ enthalten kann.

Insgesamt haben wir also vier mögliche Verteilungen.

Diese vier Verteilungen sind Funktionen sowohl von $M(p\pi)$ als auch von $M(p\pi^+)$. Normalisieren wir alle Verteilungen und machen wir eine lineare Kombination, wo die A's wieder die Anzahl der Ereignisse der verschiedenen Erzeugungs- und Zerfallsprozesse sind. Die Gesamtwahrscheinlichkeit ist dann

$$\mathcal{L}(A_j) = \left\{ \prod_{i=1}^{NE} \left[\sum_{j=1}^4 A_j P_j(M(p_B \pi^+ \pi^-), M(p_B \pi^+), M(p_B \pi^-)) \right] \right\} \times e^{-(A_1 + A_2 + A_3 + A_4)}$$

Dieses Programm heisst MITOSIS und erlaubt, die Masse und Breite des $N^*(1700)$ auch als freien Parameter zu bestimmen.¹³⁾

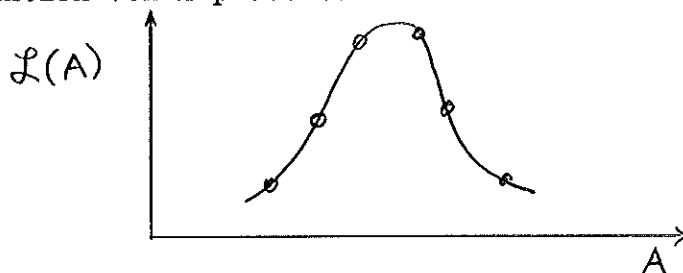
III. Maximalisierungs-Verfahren

Bis jetzt haben wir nur Beispiele der Formulierung der MLM nach physikalischen Grundsätzen diskutiert. Jetzt will ich kurz das technische Problem der Maximalisierung erwähnen.

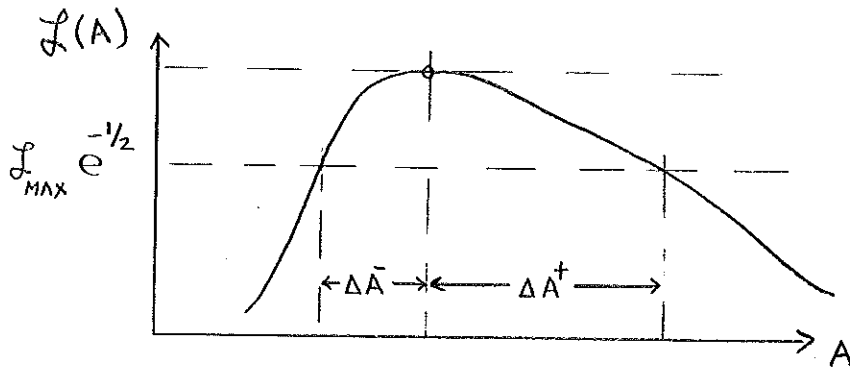
Im Fall des direkten Gebrauchs der MLM gibt es einige Spezialfälle.

1) Wenn man nur einen Parameter A hat, ist es am besten, wenn man

\mathcal{L} als Funktion von A plottet:



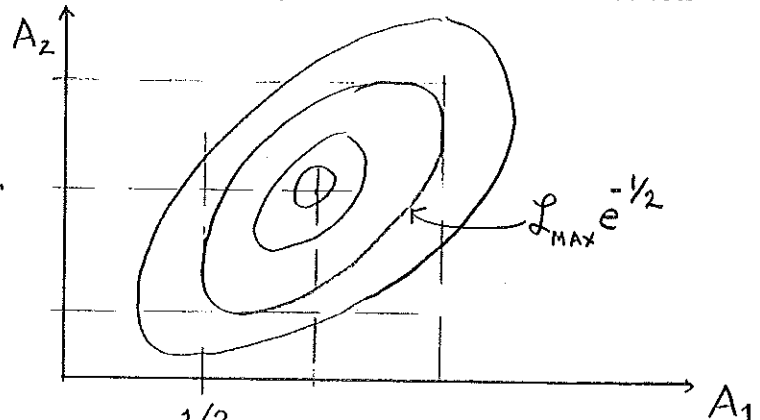
In günstigen Fällen (genügend grosse Statistik und ein realistisches physikalisches Modell) wird \mathcal{L} ähnlich einer Gaussischen Kurve sein. Wenn in ^(un) günstigen Fällen die Likelihood-Kurve asymmetrisch ist, ist es wichtiger, die ganze Kurve zu zeigen.



In jedem Fall, wenn man den Fehler zu A bestimmen will, misst man die zwei Punkte, wo \mathcal{L} ^{auf} den $e^{-1/2}$ -ten Teil abgenommen hat. In dem ungünstigen Fall, der hier gezeichnet ist, würde man asymmetrische Fehler angeben.

Wenn das Problem zwei Parameter A_1 und A_2 hat, ist es am besten, einen zwei-dimensionalen Plot zu machen, um so viel Information wie möglich zu zeigen.

In diesem Plot sind die Kurven von konstantem \mathcal{L} .
Wenn es nötig ist,
Fehler in den zwei Parametern anzugeben,



zeichnet man nur eine Kurve von $e^{-1/2}$ und schätzt die Fehler in der Weise ab, die hier gezeichnet worden ist. Die Fehler können selbstverständlich hier auch asymmetrisch sein.

Wenn man mehrere Parameter hat, ist ein Minimalisierungs- oder Maximalisierungs-programm erforderlich. Ein Minimalisierer kann natürlich als Maximalisierer funktionieren, wenn nur ein Minuszeichen vor der w-Funktion eingeführt wird. Ein solches Programm ist eines der nützlichsten und gewaltigsten Anpassungsprogramme, das jemals entwickelt worden ist. Wenn man ein einziges gutes Minimalisierungsprogramm hätte, könnte man auf die meisten seiner anderen Anpassungs-Programme verzichten und Wochen von Programm-aufwand sparen.

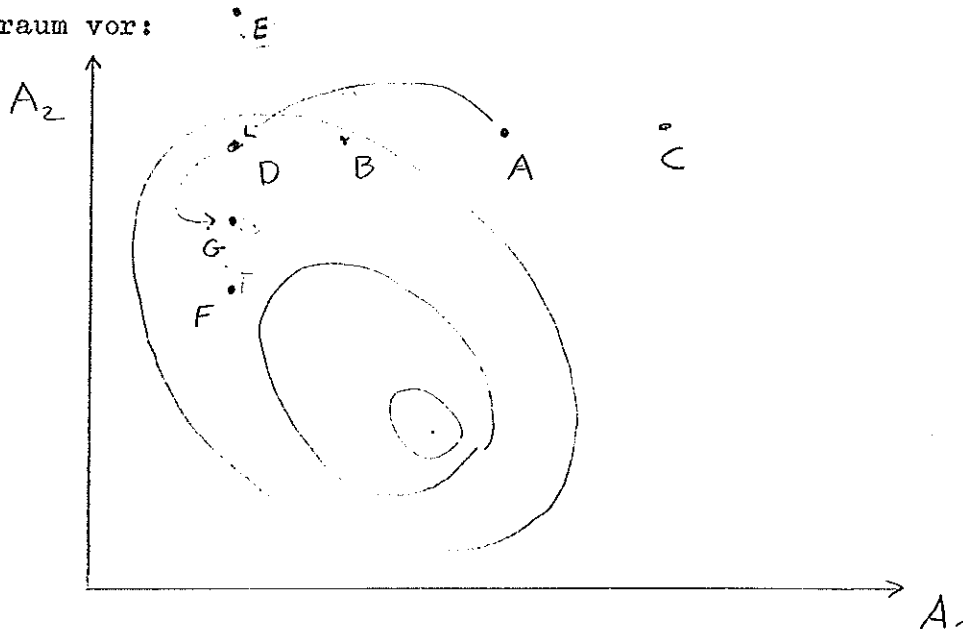
Wie funktioniert ein Minimalisierungsprogramm? Am einfachsten und langsamsten wählt man zufällig oder systematisch Werte der Parameter und testet, ob sie die Funktion kleiner machen. Im Gegensatz dazu ist die schnellste und bekannteste Methode wahrscheinlich die folgende:

An irgendeiner Stelle (sie muss schon in der Nähe des Minimums sein) berechnet das Programm alle ersten und zweiten Ableitungen von $w(\vec{A})$. Nennen wir sie $w_i(\vec{A})$ und $w_{ij}(\vec{A})$. Dann liefert eine einfache Verwendung einer Taylorischen Reihe Änderungen, die erforderlich sind, um die Minimalstelle anzunähern (die Newtonsche Methode).

$$\Delta A_i = - w_{ij}^{-1} w_j$$

Wiederholte Anwendung dieser Näherung wird normalerweise erforderlich, um die minimale Stelle mit ausreichender Genauigkeit zu finden.

Eine andere, oft verwendete Methode ist die, für die einzelnen Parameter nacheinander Minima zu suchen mit einer parabolischen Annäherung. Stellen Sie sich das folgende Verhalten im Parameter-Hyperraum vor:



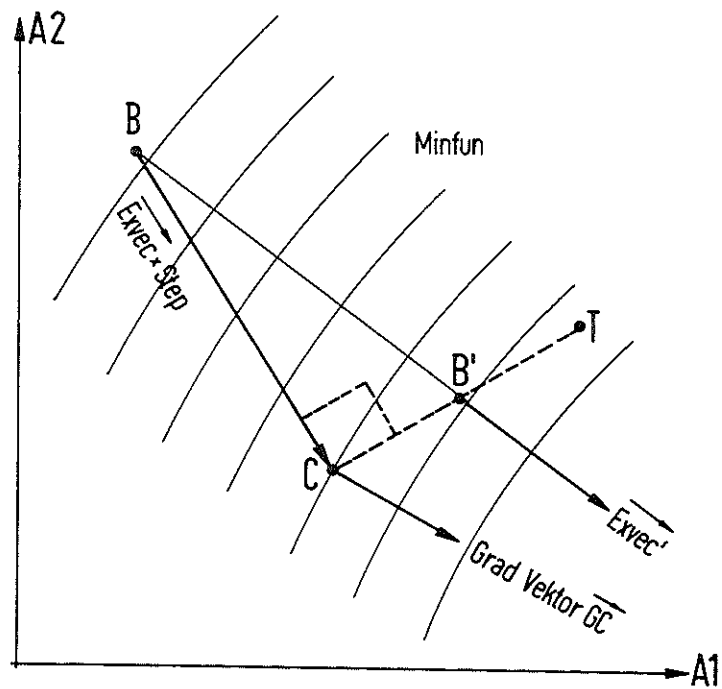


Abb.3

Ein Programm wie dieses, das von W.E. Humphrey geschrieben worden ist ¹⁴⁾, kann angewendet werden, um nicht nur ein Minimum zu finden, sondern auch um eine Entdeckungsreise durch den Hyper-Parameterraum zu machen. Dieses Programm heisst MINFUN. Die folgende Abbildung (3) erklärt, wie es funktioniert. Der Ausgangspunkt heisst B. Stellen wir uns vor, dass das Programm schon ahnt, in welcher Richtung die Schlucht liegt. (Wenn es dies nicht ahnt, berechnet es bloss den Gradienten bei B). Die vermutete Richtung heisst \vec{EXVEC} . Es macht einen willkürlichen Schritt in diese Richtung zu einem neuen Punkt C. Es berechnet dort den Gradienten \vec{GC} . In der $\vec{BC}-\vec{GC}$ Hyperebene wählt es einen weiteren willkürlichen Punkt T in einer Richtung, die senkrecht zu \vec{BC} liegt. Mit der Kenntnis des Gradienten \vec{GC} an der Stelle C und der Höhe des Punktes T bestimmt es eine Parabel. Das Extremum dieser Parabel fällt jeweils ungefähr mit der Talsohle zusammen.

Dieser Talsohlenpunkt ist ein neuer Ausgangspunkt B' und die Richtung $\vec{BB'}$ ist die neue Suchrichtung, die \vec{EXVEC} ersetzt.

Stellen Sie sich vor, dass Sie eine Binde vor den Augen tragen. Mit einem langen Stock müssen Sie so schnell wie möglich von einem Berg ins Tal kommen. Die Methode, die hier beschrieben wird, besteht darin, einen Punkt in der vermuteten Richtung zu wählen, und dort mit dem Stock herumzukratzen, um die am Ort der Spitze des Stockes vorhandene Gradientenrichtung zu finden. Dann ertastet man einen Nachbarpunkt, nur um dessen Höhe zu bestimmen. Mit dieser Auskunft errechnet man schnell die vermutete Stelle der Talsohle. Dann läuft man zu dieser Stelle und wiederholt das Verfahren mit einem kürzeren Stock.

MINFUN kann angewandt werden entweder, um das nächste Minimum zu finden (dies heisst die "converging mode"), oder ständig mit beliebiger konstanter Schrittweite Schluchten zu folgen, um einen ganzen Hyper- raum zu erforschen. MINFUN ist von Herrn J. Diaz für das DESY-Monitor- system übersetzt worden und steht bei ihm zur Verfügung.

Diese MINIMIZERS sind sehr gewaltige Werkzeuge zum Anpassen, und ich empfehle, dass Sie sie benutzen. Sie können selbstverständlich auch für einfache Kurven-Fittings angewandt werden. Sie haben den Vorteil, dass man, wenn man einen Parameter, der eine unlineare Rolle spielt, freilassen will, (z.B. die Masse oder Breite einer Resonanz), es ohne weiteres tun kann.

IV. Diskrete Hypothesen; Likelihood-Verhältnisse

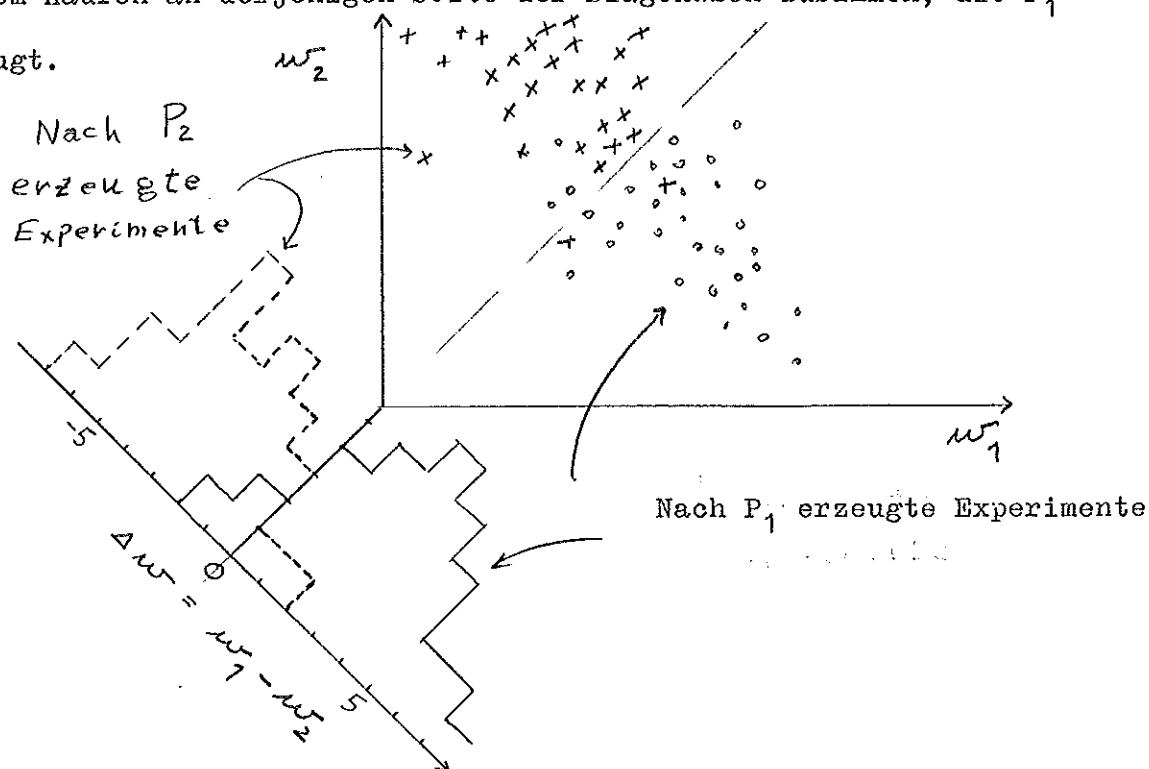
Es kommt oft vor, dass man eine Entscheidung zwischen zwei (oder mehreren) diskreten Hypothesen treffen möchte, anstatt beste Werte kontinuierlicher Parameter zu bestimmen. Nennen wir die zwei entsprechenden, auf eins normierten Verteilungen $P_1(\vec{X})$ bzw. $P_2(\vec{X})$. Eine Grösse, die oft verwandt worden ist, um die zwei Hypothesen zu unterscheiden, ist das Likelihood-Verhältnis:

$$L.V. \equiv \frac{\mathcal{L}_1}{\mathcal{L}_2} \equiv \prod_i \frac{P_1(\vec{X}_i)}{P_2(\vec{X}_i)}$$

Aus praktischen Gründen ist es oft besser, mit dem Logarithmus zu arbeiten: $\Delta w = \ln \mathcal{L}_1 - \ln \mathcal{L}_2$. Wenn ein experimentelles Likelihood-Verhältnis gross (oder klein) ist, kann man im allgemeinen sagen, dass das Experiment Hypothese 1 (bzw. 2) vorzieht.

Um die Bedeutung des Likelihood-Verhältnisses eines gegebenen Experiments zu verstehen, ist es oft nützlich, künstliche Experimente nach dem Monte Carlo - Verfahren herzustellen und die Verteilung von Δw zu untersuchen. Künstliche Experimente können natürlich mit FAKE¹⁵⁾ hergestellt werden. Jedes Ereignis kann mit der Grösse $P(\vec{X}_1)/P_{MAX}(\vec{X})$ gewichtet werden. Nur wenn diese Zahlgrösse als eine zwischen Null und Eins gleichverteilt erzeugte Zufallszahl ist, wird das Ereignis behalten. So kann ein Satz Ereignisse mit je einem Binärgewicht nach der Verteilung $P(\vec{X})$ erzeugt werden.

Der folgende Plan wird vorgeschlagen: Nehmen wir an, dass das wirkliche Experiment aus N Blaskammer-Ereignissen besteht. Ein Monte-Carlo-Rechenmaschinen-Programm wird geschrieben, das Ereignisse entweder nach der Verteilung $P_1(\vec{X})$ oder der Verteilung $P_2(\vec{X})$ erzeugen kann. Eine grosse Anzahl (einige Hundert) nach P_1 simulierter "Experimente" von je N Ereignissen wird hergestellt. Die Likelihoods \mathcal{L}_1 und \mathcal{L}_2 dieser "Experimente" werden berechnet und auf eine w_1 - w_2 -Ebene geplottet. Solche "Experimente" scharen sich normalerweise in einem Haufen an derjenigen Seite der Diagonalen zusammen, die P_1 bevorzugt.



Eine grosse Anzahl nach P_2 erzeugter "Experimente" wird ebenso behandelt. Die x-en der Abbildung repräsentieren die zweite Art von "Experiment". Beachten Sie, dass die diagonalen Linien in diesem Plot Linien konstanten Δw sind. Ein Histogramm von Δw kann also leicht aus diesem $w_1 - w_2$ Plot erhalten werden.

Das wirkliche Experiment kann natürlich auch als ein Punkt geplottet werden. Hat man Pech, so liegt dieser Punkt zwischen beiden Anhäufungen, jedoch mit etwas Glück liegt er innerhalb eines Haufens, aber weit vom zweiten entfernt. Wenn er nicht im Bereich einer der Anhäufungen liegt, sollte man beide betrachteten Hypothesen als verdächtig ansehen.

Es ist natürlich zu empfehlen, dass diese Analyse vor Ausführung des wirklichen Experiments gemacht wird, um das Experiment besser planen zu können. Solche Betrachtungen helfen, eine vernünftige, geplante Statistik des beabsichtigten Experiments zu wählen.

Mit der Auskunft dieses Plots kann man das Ergebnis seines Experimentes folgendermassen ausdrücken: "Unser Experiment bevorzugt Hypothese 1 mit einem Likelihood Verhältnis (Zahl),
Wenn diese Hypothese tatsächlich richtig ist, so ist zu erwarten, dass bei häufiger Wiederholung des gleichen Experiments in A o/o der Fälle Hypothese 1 mit einem grösseren Likelihood -Verhältnis bevorzugt wird. Wäre stattdessen Hypothese 2 richtig, so erwarten wir, dass das beobachtete Likelihood Verhältnis bei Wiederholung des Experiments nur in B o/o aller Fälle überschritten würde."

Hoffentlich ist A eine Zahl zwischen etwa 5 und 95 o/o, und B eine sehr kleine Zahl.

Ein anderes mögliches Verfahren besteht darin, die theoretischen Werte

$$\bar{w} = \int P(\vec{x}) \ln P(\vec{x}) d\vec{x} \quad \text{und}$$
$$\langle (w - \bar{w})^2 \rangle = \int P(\vec{x}) (w - \ln P(\vec{x}))^2 d\vec{x}$$

für alle möglichen Hypothesen zu berechnen und mit den entsprechenden experimentellen Werten zu vergleichen.

Es ist üblich, das Likelihood-Verhältnis als "betting odds" (Gewinnchancen) zu bezeichnen. Diese Bezeichnung kann mit einer Analogie gerechtfertigt werden. Stellen wir uns vor, dass wir Roulette mit zwei Sorten von Rädern spielen. Die erste Sorte hat gleichförmige Wahrscheinlichkeit für die Zahl 1 - 36 und die zweite hat eine ganz andere aber bekannte Wahrscheinlichkeitsverteilung. Nehmen wir an, dass bei der zweiten Art die Wahrscheinlichkeit für das Auftreten einer Zahl proportional zu dieser Zahl ist. Um das Problem zu vereinfachen, nehmen wir an, dass keines der beiden Räder eine "Null" enthält.

Bei normalem Roulette hat man nur ein Rad der uniformen Art und man setzt auf eine beliebige Zahl. Aber das jetzige Spiel ist so, dass das Casino zufällig ein Rad wählt, ohne das der Spieler weiss, welches. Das Rad wird gedreht und die Zahl wird angesagt. Dann soll der Spieler wetten, welches Rad benutzt wurde.

Die a priori Chancen sind 1:1, weil das Rad zufällig gewählt wurde und der Spieler keinerlei Information hat, bevor die Zahl angesagt wird. Wenn z.B. die erste Zahl eine "Drei" ist, kann man die a posteriori "betting odds" leicht berechnen:

$$\frac{\frac{1}{36}}{\frac{3}{1+2+\dots+36}} \approx 6.$$

Man kann sich leicht überzeugen, dass, wenn ein Rad zufällig gewählt und gedreht und dieses Verfahren sehr oft wiederholt wird, die "Drei" ungefähr sechsmal so oft auftaucht, wenn das "uniforme" Rad gedreht wird, als wenn das "proportionale" Rad gedreht wird. In diesem Sinne ist es vernünftig, das Likelihood-Verhältnis als "betting odds" zu bezeichnen, und zwar mit Geld nach diesen odds zu wetten.

Wenn dasselbe Rad noch einmal gedreht wird, können diese "odds" (ungefähr sechs) als a priori "betting odds" für die zweite Umdrehung des Spiels betrachtet werden.

Für die dritte Drehung des Rades könnte man das Produkt der ersten zwei Likelihood-Verhältnisse als a priori "betting odds" betrachten ... usw. Die Analogie mit einem Blasenkammerexperiment mit N Ereignissen (N Drehungen desselben Rades) ist augenfällig. In der Natur haben wir nicht Modelle, die zufällig gewählt sind, sondern nur ein richtiges Modell. Aber die Tatsache, dass wir keine Auskunft über dieses Modell haben, bevor das erste Ereignis analysiert wird, zwingt uns, die a priori "betting odds" als 1:1 zu betrachten.

Obwohl das Gesamt-Likelihood-Verhältnis des Experiments in diesem Sinne "betting odds" entspricht, finde ich es besser, das Ergebnis eines Experiments nach der obigen sorgfältigen Aussage auszudrücken, als nur die "betting odds" zu geben. Dadurch hat man genauere Informationen über die Zuverlässigkeit seines Experiments, weil solche Fragen wie "Wie typisch ist mein Resultat?" bzw. "Wie oft würde ein wiederholtes Experiment irreführend sein?" besser beantwortet werden können.

Im Falle eines Blasenkammer-Experiments ist es wichtig zu untersuchen, wie die einzelnen Ereignisse zum Likelihood-Verhältnis beitragen. Es besteht die Gefahr, dass ein experimenteller Bias "Ereignisse" erzeugen konnte, die in einem nach einer der Hypothesen sehr unwahrscheinlichen \bar{X} -Gebiete liegen. Solche "Ereignisse" können das Likelihood-Verhältnis stark beeinflussen und ein irreführendes Resultat verursachen. Bleibt jedoch das Likelihood-Verhältnis gross (oder klein), nachdem man diejenigen Ereignisse, die für sich allein die grössten (kleinsten) Likelihood-Verhältnisse haben, aus dem Sample entfernt hat, so kann man sich der Entscheidung etwas sicherer sein.

Man sollte auch den Einfluss des Untergrundes und der anderen Parameter des Experiments sorgfältig untersuchen, etwa durch wiederholte Bestimmung des Likelihood-Verhältnisses mit jeweils anderen Werten dieser Parameter.

Ein Fall, der für diese Behandlung geeignet ist, ist die Spin-Paritätsbestimmung des Y_1^* (1660)¹⁶⁾.

Der Verfasser hat die MLM durch den Artikel von J. Orear kennengelernt und grossen Nutzen aus der Zusammenarbeit mit D.C. Gates, T.L. Jenkins, R.W. Kenney, Ph. Eberhard und U.E. Kruse, und aus Gesprächen mit R.D. Eandi und P. Söding gezogen.

Ich danke auch P. Söding, R. Schäfer und meiner Frau Gisela für kritische Durchsicht des deutschen Textes und Frau H. Wischmeier für die Vorbereitung des Druckes.

Dieser Bericht war ein Vortrag, der im Rahmen eines Seminars von E. Lohrmann über Methoden der Blaskammerphysik am II. Institut für Experimentalphysik der Universität Hamburg gehalten wurde.

Abbildungen

- 1) Bestimmung der Polarisation des Rückstoss-Protons in der Pion-Proton-Streuung.
- 2) Experimentelle $p\pi^+\pi^-$ -Massenverteilung in der Reaktion $pp \rightarrow pp\pi^+\pi^-$ bei 10 GeV/c. Das schraffierte Histogramm ist die entsprechende Verteilung für solche $p\pi^+\pi^-$ -Kombinationen, bei denen die Masse des $p\pi^+$ -Systems im N^* -Bereich liegt.
- 3) Das Minimalisierungs-Verfahren (Ravine-Methode) des Programms MINFUN.

Zitierte Literatur

1. Jay Orear, Notes on Statistics for Physicists, UCRL-8417 (1958).
2. W.P. Swanson, D.C. Gates, T.L. Jenkins and R.W. Kenney, Phys. Rev. 137, B 1188 (1964).
3. R.D. Eandi, T.J. Devlin, R.W. Kenney, P.G. McManigal and B.J. Moyer, Phys. Rev. 136, B 1187 (1964).
4. J.D. Jackson, Nuovo Cimento 34, 1644 (1964)
5. J.H. Friedman and R.R. Ross, Phys. Rev. Lett. 16, 485 (1966).
6. N. Schmitz, Peripheral Collisions, Proceedings of the 1965 Easter School for Physicists at Bad Kreuznach, CERN 65-24 (1965).
7. D.W. Mortara, et al. University of Illinois, to be published.
8. R. Hagedorn, Relativistic Kinematics, W.A. Benjamin, Inc. New York (1963) Kapitel 7.
9. W.P. Swanson, Informelle Beschreibung des Programms TRIKOT, DESY (1965).
10. J. Friedman and R. Ross, Maximum Likelihood Estimate of Resonant State Production in Multiparticle Final States, Lawrence Radiation Laboratory, Physics Notes P-102 (1964).
M. Ferro-Luzzi et al., Nuovo Cimento 39, 417 (1965).
11. P. Söding, DESY, private Kommunikation.
12. S.P. Almeida, et.al., Oxford International Conference on High Energy Physics, (Sept. 1965).
13. W.P. Swanson, Informelle Beschreibung des Programms MITOSIS, DESY, (1966).
14. W.E. Humphrey, A General Minimizing Routine-MINFUN, Lawrence Radiation Laboratory, Physics Note P-6 (1962).
R.J. Pankhurst, MINFUN-Program to Minimize a General Function of N Variables, CERN Program Library Supplementary Write-up D-502 (1964).
15. G.R. Lynch, Program FAKE: Monte Carlo Simulation of Bubble Chamber Events. UCRL-10335 (1962).
16. A. Leveque, et al. Phys. Letters 18, 69 (1965)
Ph. Eberhard, L.R.L. - Illinois Collaboration, private Diskussion (1965).

Allgemeine Diskussionen über die MLM

Harald Cramér, Mathematical Methods of Statistics, (Princeton Univ. Press, Princeton, New Jersey, (1958), p. 498.

Johann Pfanzagl, Allgemeine Methodenlehre der Statistik II, (Sammlung Göschen Band 747/747a, Walter de Gruyter u. Co. Berlin, (1962) p. 66.

Frank T. Solmitz, Analysis of Experiments in Particle Physics, Ann. Reviews of Nuclear Science, Vol. 14, p.375 (1964).

Derek J. Hudson, Statistics Lectures II: Maximum Likelihood and Least Squares Theory, CERN 64-18, Data Handling Division, p. 123 (1964).

N.P. Klepikov and S.N. Sokolov, Analysis and Planning of Experiments by the Method of Maximum Likelihood, (Pergamon Press, Frankfurt/Main) (1961.)

F. Solmitz, Notes on the least squares and maximum likelihood methods, University of Chicago, Dec. (1954) unpublished.

Beispiele der Anwendung der MLM

C. Baltay, et al., Phys. Rev. 140, B 1030 (1965).
Anwendung der MLM, um die Lebensdauer eines Teilchens zu bestimmen.

J.A. Anderson, et al., Phys. Rev. Lett. 14, 475 (1965)
Bestimmung der Parameter des Zerfalls $K_1^0 \rightarrow \pi^+ \pi^- \pi^0$.

G.W. Meissner, et al. PRL 16, 278 (1966)
Anwendung der MLM um den K_1-K_2 Massenunterschied zu bestimmen.

Camerini, et al. PRL 14, 989 (1965)
Anwendung der MLM, um Zeitumkehrinvarianz im $K_{\mu 3}$ - Zerfall zu untersuchen.

G.P. Steck, A Statistical Problem in the Theory of Counters, Sandia Corp.
SCOC - 2964, U.S.A.E.C. TID-17493 (Office of Technical Services, Dept. of
Commerce, Washington).
Analyse allgemeiner Zähler-Theorie nach der MLM.

J. H. Friedman und R.R. Ross, Phys. Rev.Lett. 16, 485 (1966).
Gleichzeitige Bestimmung von Resonanz-Erzeugung- und Zerfall-Parametern
nach MLM.

Maximalisierungs-Verfahren

W.C. Davidon, A Variable Metric Method for Minimization,
Argonne National Laboratory, ANL-5990 (Rev.) (1959).

A.H. Rosenfeld and W.E. Humphrey, Analysis of Bubble Chamber Data,
Ann.-Rev.Nucl. Science 13, p. 126 (1963).

Douglass J. Wilde, Optimum Seeking Methods (Prentice-Hall, Inc. 1964)

