

# Statistische Behandlung von Messdaten bei der Suche nach neuen Prozessen in der Hochenergiephysik

Bachelorarbeit  
zur Erlangung des akademischen Grades  
**Bachelor of Science**  
(B.Sc.)

der Universität Siegen



Department Physik

vorgelegt von  
Martin Scholl

Siegen, Juni 2023



## **Abstract**

Die vorliegende Arbeit untersucht exemplarisch das Vorgehen, mit dem Messdaten bei der Suche nach neuen Prozessen in der Hochenergiephysik statistisch ausgewertet werden können. Anhand von simulierten Zählexperimenten werden wichtige Konzepte vorgestellt, mit denen quantifiziert werden kann, wie gut die in einem Experiment beobachteten Daten mit theoretischen Vorhersagen übereinstimmen. Untersucht wird auch, welchen Einfluss der Look-Elsewhere-Effekt auf die Signifikanz einer Messung hat. Am Beispiel der Stärke eines möglichen Signals wird beschrieben, wie obere Grenzen auf Modellparameter gesetzt werden können, wenn eine Messung (noch) nicht aussagekräftig genug ist, um von einer Entdeckung sprechen zu können. Abschließend wird dargestellt, wie Störparameter in die statistische Analyse mit einbezogen werden können.

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
<b>2</b>	<b>Suche nach neuer Physik als statistischer Test</b>	<b>2</b>
2.1	Nullhypothese und Alternativhypothese . . . . .	2
2.2	Teststatistiken . . . . .	3
2.2.1	Anzahl $n$ an gezählten Ereignissen . . . . .	4
2.2.2	Likelihood-Quotient $q^{NP}$ nach Neyman und Pearson . . . . .	5
2.2.3	Profile-Likelihood-Quotient $q^{PL}$ . . . . .	6
2.3	Signifikanz und $p$ -Werte . . . . .	8
2.4	Look-Elsewhere-Effekt . . . . .	10
2.4.1	Lokaler $p$ -Wert . . . . .	11
2.4.2	Globaler $p$ -Wert . . . . .	13
<b>3</b>	<b>Ausschlussgrenzen auf die Stärke eines Signals</b>	<b>16</b>
3.1	Nullhypothese und Alternativhypothese . . . . .	16
3.2	Konfidenzintervalle und Konfidenzniveaus . . . . .	16
3.3	Obere Grenzen auf die Signalstärke . . . . .	18
3.3.1	Direkte Methode . . . . .	19
3.3.2	CLs-Methode . . . . .	22
3.3.3	Bayesianischer Ansatz . . . . .	24
<b>4</b>	<b>Anwesenheit eines Störparameters</b>	<b>28</b>
4.1	Kontrollmessung mit Seitenbändern . . . . .	28
4.2	Simulation von Pseudoexperimenten . . . . .	31
4.3	Ergebnisse . . . . .	33
<b>5</b>	<b>Zusammenfassung und Ausblick</b>	<b>37</b>
	<b>Abkürzungsverzeichnis</b>	<b>i</b>
	<b>Literatur</b>	<b>ii</b>

# 1 Einleitung

Im Oktober 1623 erschien Galileo Galileis (1564–1642) Schrift „Il Saggiatore“, in der er seine berühmt gewordene Überzeugung äußerte, dass die Sprache der Naturwissenschaft die Mathematik sei. Aus heutiger Sicht – 400 Jahre später – ließe sich hinzufügen, dass die Sprache der experimentellen Naturwissenschaft die Statistik ist [1].

Für diese These liefern der zurzeit größte Teilchenbeschleuniger Large Hadron Collider (LHC) am Forschungszentrum CERN nahe Genf und die Detektoranlagen, die an ihn angeschlossen sind, einen beeindruckenden Beleg: In dem rund 27 Kilometer langen Beschleunigerring des LHC treffen mit einer Luminosität von bis zu  $2 \cdot 10^{34} \text{ cm}^{-2}\text{s}^{-1}$  und einer Kollisionsenergie von 13,6 TeV Protonenstrahlen aufeinander [2], die z. B. im Jahr 2017 zu  $5 \cdot 10^{15}$  Kollisionen in den Teilchendetektoren ATLAS und CMS führten [3]. Viele Teilchen, die bei einer Kollision entstehen, zerfallen wieder – ihre Zerfallsspuren können mithilfe von statistischen Methoden ausgewertet werden, die Aufschluss über Art und Eigenschaften der Zerfallsprodukte geben.

Ein wesentliches Forschungsziel dieser aufwendigen Colliding-Beam-Experimente besteht darin, physikalische Prozesse aufzuspüren, die durch Theorien vorhergesagt, aber noch nicht experimentell nachgewiesen werden konnten. Die zentrale Rolle, die die Statistik bei diesem Vorhaben einnimmt, spiegelt sich in der Tatsache wider, dass die Suche nach neuen physikalischen Phänomenen einem statistischen Test zweier Hypothesen gleicht. Das mathematische Modell, auf das die Hochenergiephysik vertraut und dessen Vorhersagen durch zahlreiche Experimente bestätigt wurden, ist das Standardmodell der Teilchenphysik (SM) [4]. Das SM beschreibt erfolgreich alle bekannten Elementarteilchen und drei der vier fundamentalen Wechselwirkungen zwischen ihnen: die elektromagnetische, die schwache und die starke Wechselwirkung. Mit der Entdeckung des Higgs-Bosons im Jahr 2012 gilt das SM als komplett und doch gibt es offene Fragen, die es in seiner derzeitigen Form nicht beantworten kann.

Hier setzen alternative Modelle an, die sich gegen das SM behaupten müssen. Ein Beispiel ist die Quantengravitation, mit deren Hilfe versucht wird, die drei vom SM beschriebenen Wechselwirkungen mit der vierten fundamentalen Kraft – der Gravitation – zu vereinigen. Andere Modelle postulieren supersymmetrische Teilchen, aus denen Dunkle Materie bestehen könnte: eine Materieform, die etwa 85 % aller Materie im Universum ausmacht, für die das SM allein jedoch keine Erklärung bietet [5]. Bei all diesen Ansätzen, das SM zu erweitern oder zu ersetzen, geht es darum, mithilfe statistischer Tests zu überprüfen, inwieweit die in einem Experiment aufgenommenen Daten mit den Vorhersagen des SM verträglich sind und inwieweit sie mit den Vorhersagen der alternativen Erklärungsversuche übereinstimmen.

Ziel dieser Arbeit ist es zum einen, in simulierten Zählexperimenten nachzuvollziehen, wie die bei der Suche nach neuen Prozessen in der Hochenergiephysik gesammelten Messdaten statistisch ausgewertet werden können. Dabei wird auch das als Look-Elsewhere-Effekt bekannte Phänomen untersucht, dass statistische Fluktuationen die Signifikanz eines Messergebnisses beeinflussen können, wenn ein lokal untersuchter Parameterraum einen nicht-festgelegten Parameter enthält, z. B. die unbekannte Masse eines gesuchten Teilchens. Zum anderen soll der Fall betrachtet werden, dass eine Messung (noch) nicht zu einer Entdeckung geführt hat: Dann können Ausschlussgrenzen auf Parameter des jeweiligen statistischen Modells gesetzt werden, was am Beispiel der Stärke eines möglichen Signals dargestellt werden soll. Die Berücksichtigung eines Störparameters in der statistischen Analyse bildet den Abschluss dieser Arbeit.

## 2 Suche nach neuer Physik als statistischer Test

In diesem Kapitel werden grundlegende Konzepte vorgestellt, die für die statistische Behandlung von Messwerten bei der Suche nach neuen Prozessen in der Hochenergiephysik wichtig sind. Die Suche nach neuer Physik wird als Test zweier Hypothesen begriffen, deren grundlegende Bedeutung den weiteren Betrachtungen vorangestellt wird. Anschließend wird die Teststatistik als zentrale Prüfgröße eingeführt, mit deren Hilfe gezeigt wird, wie die Signifikanz einer Messung bestimmt werden kann. Der Abschluss dieses Kapitels widmet sich dem als Look-Elsewhere-Effekt bekannten Phänomen, dass statistische Fluktuationen irrtümlich als neue Prozesse interpretiert werden können, indem die Wahrscheinlichkeit, dass Fluktuationen auftreten, mit der Größe des untersuchten Parameterraums zunimmt.

### 2.1 Nullhypothese und Alternativhypothese

Die Suche nach neuer Physik, die über das Standardmodell der Teilchenphysik hinausgeht, kann als ein statistischer Test zweier Hypothesen betrachtet werden. Die Hypothese, die getestet wird, wird als Nullhypothese  $H_0$  bezeichnet, während die jeweils andere eine relevante Alternativhypothese  $H_1$  darstellt. Das Ziel eines statistischen Tests besteht darin, festzustellen, ob die Nullhypothese zugunsten der Alternativhypothese verworfen werden kann:

$$\text{allgemein} \left\{ \begin{array}{l} H_0: \text{Nullhypothese, die gegen eine Alternativhypothese } H_1 \text{ getestet wird} \\ H_1: \text{Alternativhypothese} \end{array} \right.$$

Um eine Messung auf eine mögliche Entdeckung hin zu untersuchen, wird die Nullhypothese  $H_0$  als diejenige Hypothese definiert, die ausschließlich bekannte Prozesse – den Untergrund – beschreibt. Mögliche Abweichungen von diesen bekannten Prozessen werden, unter der Annahme, dass die Nullhypothese wahr ist, als statistische Fluktuationen des Untergrunds interpretiert. Die Nullhypothese geht davon aus, dass der gesuchte Prozess ausbleibt und der aktuelle Wissensstand beibehalten wird – auf diese Weise repräsentiert sie das Standardmodell in der derzeitigen Form.

Die Nullhypothese  $H_0$  wird gegen eine relevante Alternativhypothese  $H_1$  getestet, die neben den bekannten Ereignissen auch neuartige Prozesse – typischerweise das Signal – beschreibt. Dabei kann es sich zum Beispiel um Prozesse handeln, die durch die Theorie vorhergesagt, aber noch nicht experimentell nachgewiesen werden konnten. So stellen zum Beispiel schwach wechselwirkende massive Teilchen (WIMP) wie das Neutralino hypothetische supersymmetrische Teilchen dar, die als Kandidaten für die Dunkle Materie infrage kommen [6]. Um sie nachzuweisen, müssen ihre Zerfälle beobachtet werden, die den gesuchten Signalprozessen entsprechen. Wie dieses Beispiel veranschaulicht, stellen Alternativhypothesen in der Hochenergiephysik häufig Erweiterungen des Standardmodells dar.

Werden statistische Analysen durchgeführt, deren Ziel es ist, neue Prozesse jenseits des Standardmodells zu entdecken, nehmen Nullhypothese  $H_0$  und Alternativhypothese  $H_1$  daher folgende Funktionen wahr [7]:

$$\text{Suche nach neuen Prozessen} \left\{ \begin{array}{l} H_0: \text{Hypothese, die nur Untergrundprozesse beschreibt} \\ H_1: \text{Hypothese, die Signal- und Untergrundprozesse beschreibt} \end{array} \right.$$

In der vorliegenden Arbeit wird die Hypothese, die ausschließlich bekannte Untergrundereignisse beschreibt, als *Untergrundhypothese* bezeichnet, während die Hypothese, die zusätzlich die gesuchten Signalereignisse beschreibt, als *Signalhypothese* bezeichnet wird.<sup>1</sup>

## 2.2 Teststatistiken

Im Allgemeinen werden in einem Experiment der Hochenergiephysik mehrere Variablen  $\vec{x} = (x_1, \dots, x_n)$  gemessen, die entsprechend einer Wahrscheinlichkeitsverteilung  $f(\vec{x})$  verteilt sind. Die Wahrscheinlichkeitsverteilung  $f(\vec{x} | H_0)$  beschreibt die Wahrscheinlichkeit, die Daten  $\vec{x}$  zu erhalten, wenn die Nullhypothese  $H_0$  wahr ist. Die Wahrscheinlichkeitsverteilung  $f(\vec{x} | H_1)$  hingegen beschreibt die Wahrscheinlichkeit, die Daten  $\vec{x}$  zu erhalten, wenn die Alternativhypothese  $H_1$  wahr ist. Beide Wahrscheinlichkeitsverteilungen unterscheiden sich im Allgemeinen.

Die Aufgabe eines statistischen Tests besteht darin, zu bestimmen, inwieweit die betrachteten Hypothesen mit den beobachteten Daten eines Experiments übereinstimmen. Um dafür nicht alle verfügbaren Variablen  $(x_1, \dots, x_n)$  nutzen zu müssen, deren Wahrscheinlichkeitsverteilungen komplizierte Eigenschaften haben können, wird eine Teststatistik  $t$  definiert. Sie fasst den Ausgang eines Experiments in einer reellwertigen Zahl zusammen und ist daher eine Funktion der Daten  $\vec{x}$  [8]:

$$t = t(\vec{x}). \quad (1)$$

Da die Teststatistik  $t$  von den Zufallsvariablen  $\vec{x}$  abhängt, ist sie selbst eine Zufallsvariable – sie unterliegt den Wahrscheinlichkeitsverteilungen  $f(t | H_0)$  und  $f(t | H_1)$ . Im Folgenden wird anhand von drei Teststatistiken gezeigt, wie ihre Verteilungen aus den Wahrscheinlichkeitsverteilungen der Variablen  $\vec{x}$  bestimmt werden können. Dazu wird ein Zählexperiment betrachtet, bei dem die Anzahl an Untergrundereignissen  $n_b$  und die Anzahl an Signalereignissen  $n_s$  Poissonverteilt sind. Die erwartete Anzahl an Untergrundereignissen  $\bar{b}$  und an Signalereignissen  $\bar{s}$  seien bekannt – sie entsprechen den Erwartungswerten der jeweiligen Poisson-Verteilung. Es gilt daher:

$$p_{\bar{b}}(n_b) = \frac{\bar{b}^{n_b}}{n_b!} e^{-\bar{b}} \quad (2)$$

$$p_{\bar{s}}(n_s) = \frac{\bar{s}^{n_s}}{n_s!} e^{-\bar{s}} \quad (3)$$

Da die Poisson-Verteilung reproduktiv ist, folgt die Summe zweier unabhängiger Poisson-verteilter Zufallsvariablen ebenso einer Poisson-Verteilung. Daher gilt für die Anzahl der Gesamtereignisse  $n_s + n_b$ :

$$p_{\bar{s}+\bar{b}}(n_s + n_b) = \frac{(\bar{s} + \bar{b})^{n_s+n_b}}{(n_s + n_b)!} e^{-(\bar{s}+\bar{b})}. \quad (4)$$

Der allgemeine Fall mit Variablen  $\vec{x} = (x_1, \dots, x_2)$  reduziert sich also bei den hier betrachteten Zählexperimenten auf nur eine Variable, die der Anzahl  $n$  an Ereignissen entspricht. Für die Teststatistik gilt damit:

$$t = t(n). \quad (5)$$

<sup>1</sup> In der Literatur werden häufig die Begriffe „Nur-Untergrund-Hypothese“ und „Signal-plus-Untergrund-Hypothese“ verwendet. Sie werden in der vorliegenden Arbeit aus Gründen der besseren Lesbarkeit ersetzt.

Beide Wahrscheinlichkeitsverteilungen,  $f(n | H_0)$  und  $f(n | H_1)$ , folgen im betrachteten Beispiel außerdem dem gleichen Verteilungstyp, nämlich einer Poisson-Verteilung.

### 2.2.1 Anzahl $n$ an gezählten Ereignissen

Die einfachste Teststatistik, mit der ein Zählexperiment beschrieben werden kann, entspricht der Anzahl  $n$  an gezählten Ereignissen selbst [8]:

$$t = t(n) = n. \quad (6)$$

Die Verteilung der Teststatistik  $n$  ist eine Poisson-Verteilung mit dem Parameter  $\bar{b}$ , falls die Untergrundhypothese wahr ist, und eine Poisson-Verteilung mit dem Parameter  $\bar{s} + \bar{b}$ , falls die Signalthypothese wahr ist. *Abbildung 1* zeigt die auf 1 normierten Teststatistik-Verteilungen für den Fall, dass acht Untergrundereignisse bzw. sechs Signalereignisse erwartet und zehn Ereignisse beobachtet wurden. Bei letzteren kann in der Praxis nicht unterschieden werden, ob es sich um Untergrund- oder Signalereignisse handelt. Die Anzahl an beobachteten Ereignissen wird mit einem tiefgestellten  $obs$  gekennzeichnet:

$$t_{obs} = t(n_{obs}) = n_{obs}. \quad (7)$$

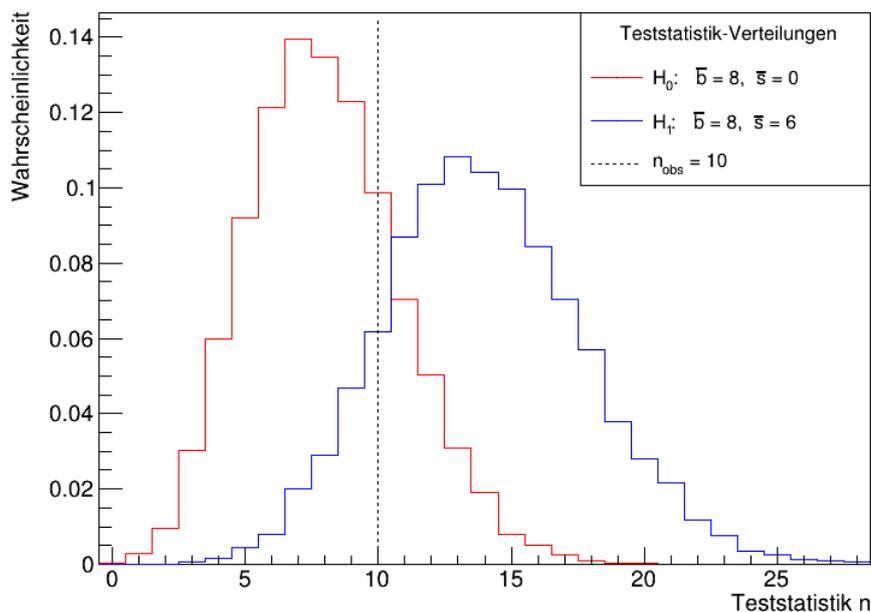


Abbildung 1: Auf 1 normierte Verteilungen der Teststatistik  $n$  für den Fall, dass acht Untergrundereignisse erwartet werden, wenn die Untergrundhypothese wahr ist (rot, linke Verteilung) und zusätzlich sechs Signalereignisse erwartet werden, wenn die Signalthypothese wahr ist (blau, rechte Verteilung). Die gestrichelte Linie markiert die Anzahl  $n_{obs}$  der Ereignisse, die in einem Experiment beobachtet wurden. In diesem Beispiel gilt  $n_{obs} = 10$ .

Abbildung 1 veranschaulicht außerdem, dass es sich bei einer Teststatistik um eine monotone Funktion der Zählvariablen  $n$  handelt: Der Ausgang eines Zählperiments ist umso besser mit der Untergrundhypothese verträglich, je geringer die Anzahl an beobachteten Ereignissen  $n_{obs}$

ist. Je mehr Ereignisse  $n_{obs}$  beobachtet werden, desto eher stimmt das Messergebnis hingegen mit den Vorhersagen der Signalthypothese überein.

### 2.2.2 Likelihood-Quotient $q^{NP}$ nach Neyman und Pearson

Ein statistischer Test zweier Hypothesen ist umso mächtiger, je größer die Wahrscheinlichkeit ist, die Nullhypothese  $H_0$  zu verwerfen, wenn sie tatsächlich falsch ist. Dies kann durch eine optimale Trennung der Hypothesen  $H_0$  und  $H_1$  erreicht werden. Die Teststatistik, die dies ermöglicht, ist nach dem Lemma von Neyman und Pearson gegeben durch das Verhältnis der Likelihoods  $L(n_{obs} | H_0)$  und  $L(n_{obs} | H_1)$ , mit denen die beobachteten Daten  $n_{obs}$  erhalten werden, wenn die jeweils betrachtete Hypothese wahr ist [9]. Die Likelihood-Funktionen haben im Falle eines Zähl-experiments mit Poisson-verteilter Zufallsvariable die folgende Form:

$$L(n_{obs} | H_0) = \frac{\bar{b}^{n_{obs}}}{n_{obs}!} e^{-\bar{b}} \quad (8)$$

$$L(n_{obs} | H_1) = \frac{(\bar{s} + \bar{b})^{n_{obs}}}{n_{obs}!} e^{-(\bar{s} + \bar{b})} \quad (9)$$

Zur leichteren mathematischen Handhabung wird dem Verhältnis der Likelihoods ein konstanter Faktor (-2) und der natürliche Logarithmus vorangestellt. Da ein konstanter Faktor die Werte der Teststatistik nur verschiebt und der natürliche Logarithmus eine monotone Funktion im Intervall  $(0, \infty)$  ist, liefert dies äquivalente Ergebnisse. Die Werte  $q^{NP}$ , aus denen sich eine Teststatistik-Verteilung zusammensetzt, sind damit gegeben durch:

$$t = q^{NP} = -2 \ln \frac{L(n | H_1)}{L(n | H_0)} = -2 n \ln \left( 1 + \frac{s}{b} \right) + 2 s. \quad (10)$$

Um die Verteilung der Teststatistik  $q^{NP}$  unter der Hypothese  $H_0$  zu erhalten, werden Werte  $n = n_b$  aus der Poisson-Verteilung mit Erwartungswert  $\bar{b}$  aus Gleichung 2 gezogen, und mit dem jeweils untersuchten Parameter  $\bar{s}$  für die Signalthypothese in Gleichung 10 eingesetzt. Für die Verteilung der Teststatistik unter der Hypothese  $H_1$  werden analog Werte  $n = n_s + n_b$  aus der Poisson-Verteilung mit Erwartungswert  $\bar{s} + \bar{b}$  aus Gleichung 4 gezogen, und mit dem gleichen Parameter  $\bar{s}$  in Gleichung 10 eingesetzt. *Abbildung 2* zeigt die so erhaltenen normierten Teststatistik-Verteilungen für wiederum acht erwartete Untergrund- bzw. sechs erwartete Signallereignisse und zehn beobachtete Ereignisse. Dabei entspricht ein Wert  $n_{obs}$  nun dem Wert  $q_{obs}^{NP}$ , der entsprechend Gleichung 11 bestimmt werden kann:

$$q_{obs}^{NP} = -2 \ln \frac{L(n_{obs} | H_1)}{L(n_{obs} | H_0)} = -2 n_{obs} \ln \left( 1 + \frac{s}{b} \right) + 2 s. \quad (11)$$

Aus der Definition der Teststatistik  $q^{NP}$  in Gleichung 10 folgt, dass große Werte von  $q^{NP}$  gut mit der Untergrundhypothese  $H_0$  verträglich sind, während kleiner werdende  $q^{NP}$ -Werte für eine immer bessere Übereinstimmung mit der Signalthypothese  $H_1$  sprechen. Das spiegelt sich in *Abbildung 2* wider, in der zu erkennen ist, dass die Hypothesen  $H_0$  und  $H_1$  ihre Seiten getauscht haben.

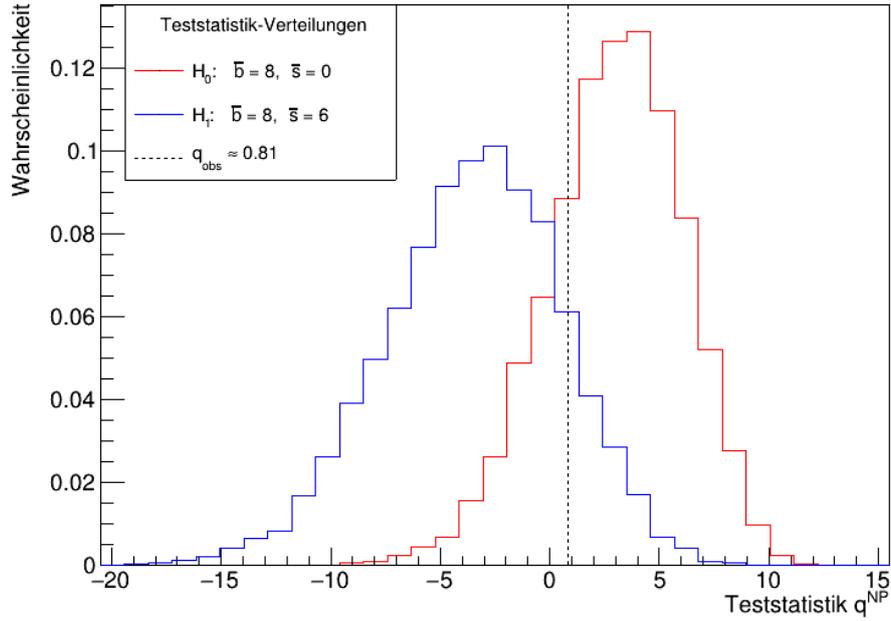


Abbildung 2: Auf 1 normierte Verteilungen der Teststatistik  $q^{NP}$  für den Fall, dass acht Untergrundereignisse erwartet werden, wenn die Untergrundhypothese wahr ist (rot, rechte Verteilung) und zusätzlich sechs Signalereignisse erwartet werden, wenn die Signalhypothese wahr ist (blau, linke Verteilung). Die gestrichelte Linie markiert den beobachteten Wert  $q_{obs}^{NP}$ , der zehn beobachteten Ereignissen entspricht. Im Unterschied zu den Wahrscheinlichkeitsverteilungen der Teststatistik  $n$ , entsprechen nun kleinere Werte der Teststatistik  $q^{NP}$  einer besseren Übereinstimmung mit der Signalhypothese.

### 2.2.3 Profile-Likelihood-Quotient $q^{PL}$

Bisher wurde die Anzahl an erwarteten Untergrundereignissen  $\bar{b}$  als bekannt vorausgesetzt. In der Praxis kann es allerdings sein, dass der wahre Untergrund unbekannt ist und der Parameter  $\bar{b}$  als Störparameter in die statistische Behandlung der Messdaten aufgenommen werden muss. Eine Möglichkeit, dies zu tun, ist die Wahl einer Teststatistik, die auf einem Profile-Likelihood-Verhältnis basiert [10]:

$$q^{PL} = -2 \ln \frac{L(s | \hat{b}(s))}{L(\hat{s} | \hat{b})}. \quad (12)$$

Dabei ist  $\hat{b}$  der Wert des Störparameters  $\bar{b}$ , der die Likelihood  $L(s | b)$  für eine bestimmte Signalstärke  $s$  maximiert. Dadurch hängt die Profile-Likelihood  $L(s | \hat{b}(s))$  nicht vom Störparameter  $\hat{b}$  ab, sondern nur von dem Parameter  $s$ . Die Größen  $\hat{s}$  und  $\hat{b}$  sind diejenigen Werte für die Anzahl an Signalereignissen  $s$  und Untergrundereignissen  $b$ , bei denen die Wahrscheinlichkeit für die beobachteten Stichprobenwerte am größten ist.  $L(\hat{s} | \hat{b})$  entspricht daher dem globalen Maximum der Likelihood und es gilt  $L(\hat{s} | \hat{b}) \geq L(s | \hat{b}(s))$ . Der Profile-Likelihood-Quotient nimmt daher Werte im Intervall  $[0,1]$  an: Je näher dieser Wert bei 1 ist, desto besser stimmt der Wert der Signalstärke  $s$  mit den beobachteten Daten überein; je näher er bei 0 ist, desto weniger gut ist diese Übereinstimmung.

Für den hier betrachteten Fall, dass das Ziel der statistischen Analyse die Entdeckung eines neuen physikalischen Phänomens ist, stellt die Untergrundhypothese die getestete Nullhypothese  $H_0$  dar. Für sie gilt  $s = 0$  und für den Zähler in Gleichung 12 folgt  $L(s = 0 | \hat{b}(s = 0))$ . Für den Nachweis eines positiven Signals wird außerdem die Annahme getroffen, dass die Anwesenheit neuer Signalprozesse die Anzahl an Ereignissen relativ zu den erwarteten Untergrundereignissen nur erhöhen und nicht verringern kann, sodass der Parameter für die Signalstärke nicht negativ ist. Zusammengefasst ergibt sich damit für die Teststatistik  $q^{PL}$ , die für die Suche nach einem positiven Signal in der Hochenergiephysik genutzt werden kann:

$$q^{PL} = \begin{cases} -2 \ln \frac{L(s=0 | \hat{b}(s=0))}{L(\hat{s} | \hat{b})}, & \hat{s} \geq 0 \\ 0, & \hat{s} < 0 \end{cases} \quad (13)$$

Die mathematische Herleitung des Profile-Likelihood-Quotienten für eine unbekannte Anzahl an zu erwartenden Untergrundereignissen wird in Kapitel 4 behandelt. Zunächst soll der Spezialfall betrachtet werden, dass der Parameter  $b$  bekannt und konstant ist. In diesem Fall entspricht das Profile-Likelihood-Verhältnis im Wesentlichen dem inversen Likelihood-Verhältnis nach Neyman und Pearson – mit dem Unterschied, dass die konstante Signalstärke  $s$  durch ihren besten Schätzwert  $\hat{s}$  ersetzt wird:

$$q^{PL} = \begin{cases} -2 \ln \frac{L(s=0|b)}{L(\hat{s}|b)} = -2n \ln \left( \frac{b}{\hat{s}+b} \right) - 2\hat{s}, & \hat{s} \geq 0 \\ 0, & \hat{s} < 0 \end{cases} \quad (14)$$

Dieser Schätzer wird bestimmt, indem das Maximum der logarithmierten Likelihood  $L(s | b)$  ermittelt wird. Dazu wird diese an der Stelle  $s = \hat{s}$  abgeleitet und gefordert, dass die Ableitung gleich null ist:

$$\begin{aligned} \left. \frac{\partial \ln L(s | b)}{\partial s} \right|_{s=\hat{s}} &= \left. \frac{\partial \ln \left( \frac{(s+b)^n}{n!} e^{-(s+b)} \right)}{\partial s} \right|_{s=\hat{s}} \\ &= \left. \frac{\partial (\ln \left( \frac{(s+b)^n}{n!} \right) + \ln (e^{-(s+b)}))}{\partial s} \right|_{s=\hat{s}} \\ &= \left. \frac{\partial (n \ln (s+b) - \ln (n!) - s - b)}{\partial s} \right|_{s=\hat{s}} \quad (15) \\ &= \frac{n}{\hat{s} + b} - 1 \stackrel{!}{=} 0 \\ &\implies \hat{s} = n - b. \end{aligned}$$

Eingesetzt in Gleichung 14 lautet die Profile-Likelihood-Teststatistik für die Suche nach einem positiven Signal im Spezialfall einer bekannten Anzahl an erwarteten Untergrundereignissen:

$$q^{PL} = \begin{cases} -2n \ln(b) + 2b + 2n \ln(n) - 2n, & n \geq b \\ 0, & n < b \end{cases} \quad (16)$$

Aufgrund der auf das Intervall  $[0, 1]$  beschränkten Werte des Profile-Likelihood-Verhältnisses nimmt die Teststatistik  $q^{NP}$  ausschließlich nicht-negative Werte an, was in *Abbildung 3* an den Teststatistik-Verteilungen für die Nullhypothese  $H_0$  und für die Alternativhypothese  $H_1$  ersichtlich ist. Auch in diesem Beispiel wurden acht Untergrundereignisse bzw. sechs Signalereignisse erwartet und zehn Ereignisse beobachtet.

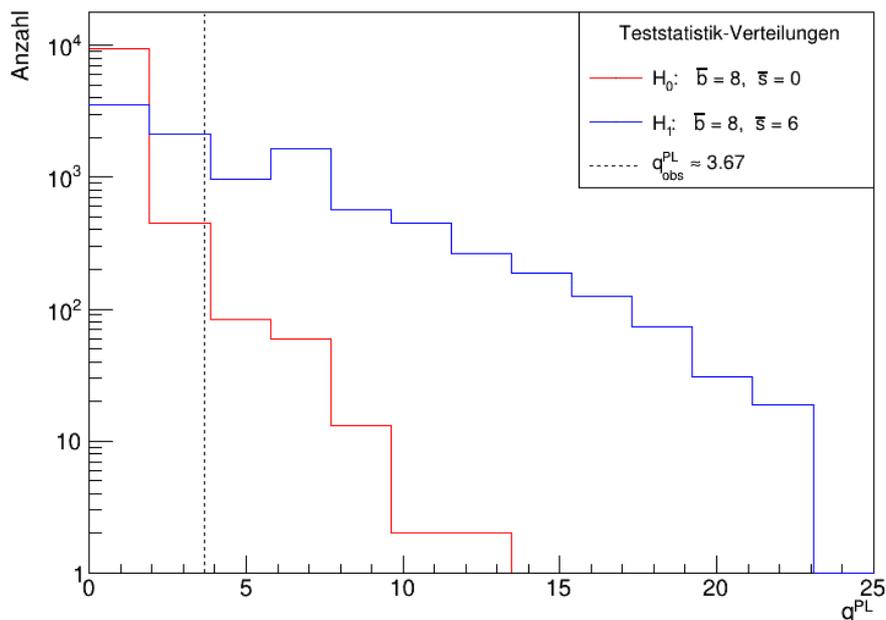


Abbildung 3: Verteilungen der Teststatistik  $q^{PL}$  für den Fall, dass acht Untergrundereignisse erwartet werden, wenn die Untergrundhypothese wahr ist (rot, linke Verteilung) und zusätzlich sechs Signalereignisse erwartet werden, wenn die Signalthypothese wahr ist (blau, rechte Verteilung). Die gestrichelte Linie markiert den beobachteten Wert  $q_{obs}^{PL}$ , der zehn beobachteten Ereignissen entspricht.

### 2.3 Signifikanz und $p$ -Werte

Sowohl die Nullhypothese  $H_0$  als auch die Alternativhypothese  $H_1$  machen eine Aussage darüber, wie wahrscheinlich es ist, einen bestimmten Ausgang eines Experiments zu beobachten, wenn die jeweils betrachtete Hypothese wahr ist. Die Wahrscheinlichkeit, Daten zu erhalten, die genauso oder weniger verträglich mit den Vorhersagen der Nullhypothese sind als die beobachteten Daten, wird  $p$ -Wert genannt. Im Zusammenhang mit der Suche nach neuen Prozessen wird in den folgenden Betrachtungen ein kleines  $b$  hinzugefügt, um ihn von den  $p$ -Werten zu unterscheiden, die bei der Bestimmung von Ausschlussgrenzen in Kapitel 3 genutzt werden. Für die  $p_b$ -Werte der drei vorgestellten Teststatistiken gelten die Gleichungen 17–19, *Abbildung 4* veranschaulicht außerdem den  $p_b$ -Wert für die Teststatistik  $n$  bei einem gewählten Signifikanzniveau von  $\alpha_b = 0.10$ :

$$p_b = P(n \geq n_{obs}) = \sum_{n=n_{obs}}^{\infty} \frac{\bar{b}^n}{n!} e^{-\bar{b}} \quad (17)$$

$$p_b = P(q^{NP} \geq q_{obs}^{NP}) = \int_{-\infty}^{q_{obs}^{NP}} f(q^{NP} | H_0) dq^{NP} \quad (18)$$

$$p_b = P(q^{PL} \geq q_{obs}^{PL}) = \int_{q_{obs}^{PL}}^{\infty} f(q^{PL} | H_0) dq^{PL} \quad (19)$$

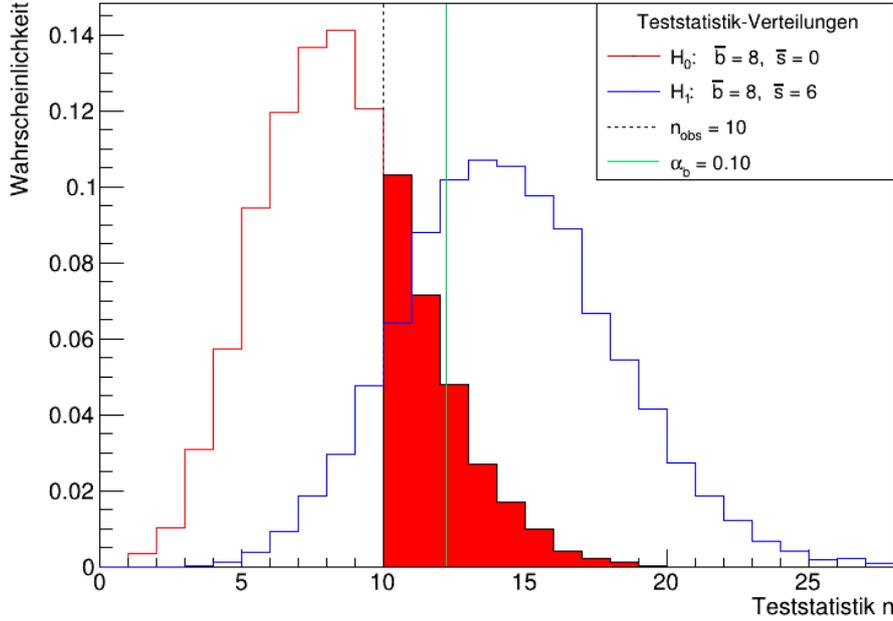


Abbildung 4: Die rot eingefärbte Fläche unter der Nullhypothese  $H_0$  veranschaulicht den  $p_b$ -Wert am Beispiel der Teststatistik  $n$ : Er entspricht der Wahrscheinlichkeit, ein genauso oder weniger wahrscheinliches Messergebnis zu erhalten als das beobachtete, wenn die Nullhypothese wahr ist. In diesem Beispiel liegt der  $p_b$ -Wert für zehn beobachtete Ereignisse (schwarz gestrichelte Linie) bei  $p_b \approx 0.29$  und ist damit größer als das zuvor festgelegte Signifikanzniveau  $\alpha_b = 0.10$  (grüne Linie). Die Messung ist daher nicht statistisch signifikant bzgl. eines Verwerfens von  $H_0$ .

Ist der  $p_b$ -Wert kleiner als ein vor der Messung festgelegtes Signifikanzniveau  $\alpha_b$ , gilt die Messung als statistisch signifikant und die Nullhypothese wird zugunsten der Signalthypothese verworfen. Dabei entspricht das Signifikanzniveau der Wahrscheinlichkeit, die Nullhypothese  $H_0$  zu verwerfen, obwohl sie wahr ist (Fehler 1. Art); sie spiegelt damit den Fehler wider, der bei der Suche nach einer Entdeckung toleriert wird. In der Teilchenphysik ist es Konvention, diesen Fehler so zu wählen, dass er folgendem Signifikanzniveau entspricht:

$$\alpha_b = 2.87 \cdot 10^{-7}. \quad (20)$$

Aus dieser Festlegung folgt, dass die Nullhypothese erst dann verworfen und von einer Entdeckung gesprochen wird, wenn die Wahrscheinlichkeit gleich oder kleiner ist als 0.0000287 %, dass ein beobachtetes Signal allein durch statistische Fluktuationen bekannter Untergrundprozesse zustande gekommen ist. Häufig wird der  $p_b$ -Wert auch in eine äquivalente Größe übersetzt, die die Signifikanz durch die Anzahl  $Z$  an Standardabweichungen angibt, die eine Gauß-verteilte Variable von ihrem Erwartungswert entfernt ist:

$$Z = \Phi^{-1}(1 - p_b). \quad (21)$$

Dabei ist  $\Phi^{-1}$  die inverse kumulative Verteilungsfunktion, d. h. das Quantil, der Standardnormalverteilung. Ein  $p_b$ -Wert von  $2.87 \cdot 10^{-7}$  entspricht dabei einer Signifikanz von  $Z = 5\sigma$  – sie muss mindestens erreicht werden, um in der Teilchenphysik von einer Entdeckung sprechen zu können [4].

## 2.4 Look-Elsewhere-Effekt

Gilt die Suche nach neuen Prozessen bisher unentdeckten Teilchen, ist es im Allgemeinen der Fall, dass ein Nachweis indirekt über ihre Zerfälle erfolgen muss, da die Teilchen selbst zu schnell zerfallen. Zu diesem Zweck untersuchen große Teilchendetektoren wie der ATLAS-Detektor am Europäischen Forschungszentrum für Teilchenphysik CERN bei Genf die Zerfallsmuster, die nach einer Teilchenkollision in Beschleunigeranlagen entstehen: Sie messen unter anderem die Energie und den Impuls der beteiligten Zerfallsprodukte, aus denen jeweils die invariante Masse des Mutterteilchen berechnet werden kann. Die Verteilung der invarianten Masse wird auf resonante Überhöhungen hin untersucht, die ein Hinweis auf den Zerfall – und damit auf die Existenz – des gesuchten Teilchens sein können. Ist jedoch nicht bekannt, an welcher Stelle des untersuchten Massenspektrums ein Überschuss an Ereignissen zu erwarten ist, weil z. B. die Masse eines neuen Teilchens unbekannt ist, muss bei der statistischen Auswertung der Daten berücksichtigt werden, dass eine solche Überhöhung auch an jeder anderen Stelle des Massenspektrums auftreten kann – allein aufgrund der Größe des untersuchten Parameterraums. Das Phänomen, dass ein derartiger Ereignisüberschuss als scheinbares Signal missverstanden werden kann, wird als Look-Elsewhere-Effekt bezeichnet [11].

Um den Look-Elsewhere-Effekt genauer zu untersuchen, wird ein Beispielspektrum betrachtet, das zehn Bins eines beliebigen Massenfensters umfasst. Die Anzahl der Bins werde mit  $n_{bins}$  bezeichnet. Außerdem werden folgende Annahmen gemacht:

- das echte Signal befinde sich in Bin 5 (Signalbin)
- die Anzahl an erwarteten Signalereignissen sei  $\bar{s} = 6$
- die Anzahl an erwarteten Untergrundereignissen pro Bin sei  $\bar{b} = 8$
- die Wahrscheinlichkeitsdichteverteilungen des Untergrunds  $\rho_b(x)$  und des Signals  $\rho_s(x)$  seien jeweils Gleichverteilungen, d. h. es gelte:

$$\rho_b(x) = \begin{cases} \frac{1}{10}, & 0 \leq x \leq 10 \\ 0, & \text{sonst} \end{cases} \quad (22)$$

$$\rho_s(x) = \begin{cases} 1, & 4 \leq x \leq 5 \\ 0, & \text{sonst} \end{cases} \quad (23)$$

Der erste Schritt besteht darin, den von der Nullhypothese  $H_0$  vorhergesagten Untergrund zu generieren. Dazu wird die Anzahl der zu nehmenden Messwerte für jedes Pseudoexperiment aus einer Poisson-Verteilung mit Erwartungswert  $\lambda = n_{bins} \cdot \bar{b}$  gezogen. Jedem Untergrundereignis wird anschließend gemäß Gleichung 22 ein zufälliger Wert des betrachteten Massenbereichs zugeordnet. Das Eintragen dieser Werte in ein Histogramm liefert für ein Pseudoexperiment die in *Abbildung 5* mit *nur Untergrund* gekennzeichnete Verteilung.

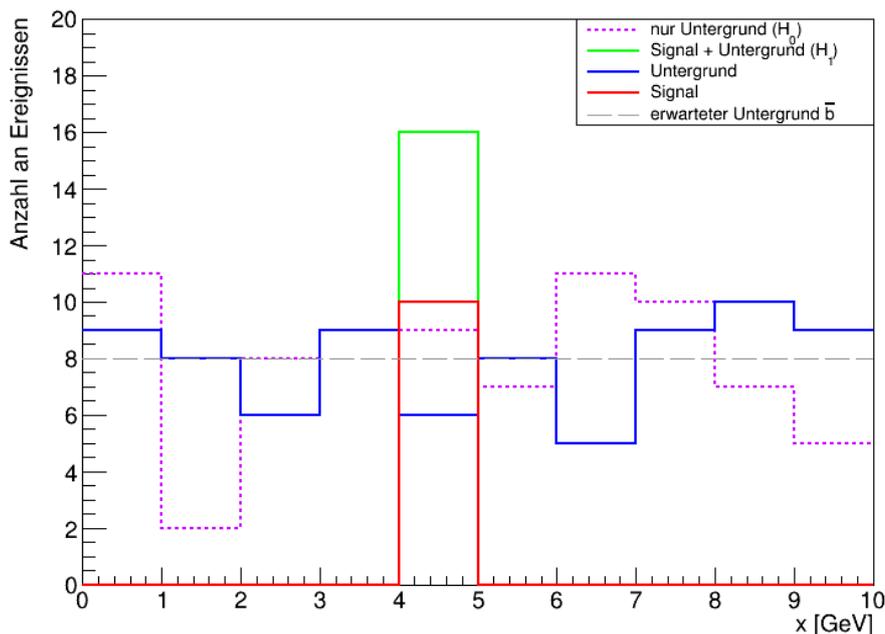


Abbildung 5: Beispielhaftes Massenspektrum eines Pseudoexperiments, bei dem das echte Signalbin (Bin 5) die höchste Anzahl an Ereignissen aufweist. In diesem Fall wird das Signal korrekt identifiziert.

Die von der Alternativhypothese  $H_1$  vorhergesagte Verteilung setzt sich zusammen aus einem analog erzeugten Untergrund – blau dargestellt (durchgezogene Verteilung) –, auf den im Signalbin ein Signal (rot, unten) hinzuaddiert wird. Die Anzahl der Signalereignisse wird hierfür aus einer Poisson-Verteilung mit Erwartungswert  $\bar{s}$  entnommen. Die Summe aus Untergrund- und Signalereignissen im Signalbin ist grün (oben) dargestellt. Abbildung 5 zeigt den Fall, der am häufigsten eintritt: Das Signalbin weist die höchste Anzahl an Ereignissen auf, sodass das Signal richtig identifiziert wird. *Abbildung 6* zeigt die ebenfalls mögliche Situation, dass das echte Signal durch Fluktuationen unterdrückt wird: Aufgrund der höheren Ereigniszahlen in der Umgebung des Signalbins, wird das Signal nun an anderer als der tatsächlichen Stelle des betrachteten Massenintervalls vermutet. Wie oft das in dem hier betrachteten Beispiel passiert, lässt sich anhand der in *Abbildung 7* abgebildeten relativen Signalstärke veranschaulichen: In etwa 63 % der 10000 durchgeführten Pseudoexperimente wird das echte Signalbin richtig identifiziert, in etwa 37 % wird das Signal in einem anderen Bin vermutet. *Abbildung 8* stellt ergänzend für verschiedene Signalstärken und pro Bin erwartete Untergrundereignisse die Wahrscheinlichkeiten dar, mit denen das Signalbin jeweils korrekt ausgewählt wird.

Interessant ist daher die Frage, wie sich der beschriebene Look-Elsewhere-Effekt auf die Signifikanz einer Messung auswirkt. Dazu wird im Folgenden am Beispiel des Profile-Likelihood-Quotienten  $q^{PL}$  der  $p_b$ -Wert zunächst unter der Annahme berechnet, dass das Signalbin bekannt ist. Dieser  $p_b$ -Wert wird als lokaler  $p$ -Wert bezeichnet, da sich die berechnete Signifikanz auf eine bestimmte Position im Massenspektrum bezieht. Im Anschluss wird der  $p_b$ -Wert noch einmal berechnet, wobei berücksichtigt wird, dass ein Ereignisüberschuss auch an jeder anderen Position des Massenspektrums auftreten kann. Das führt schließlich auf den Begriff des globalen  $p$ -Werts.

### 2.4.1 Lokaler $p$ -Wert

Ist das Signalbin bekannt, wird die Signifikanzanalyse für eine feste Masse  $x$  durchgeführt. Die in Kapitel 2.2.3 vorgestellte Profile-Likelihood-Teststatistik nimmt daher für die Suche nach einem positiven Signal entsprechend Gleichung 16 für einen festen Wert  $x$  folgende Form an:

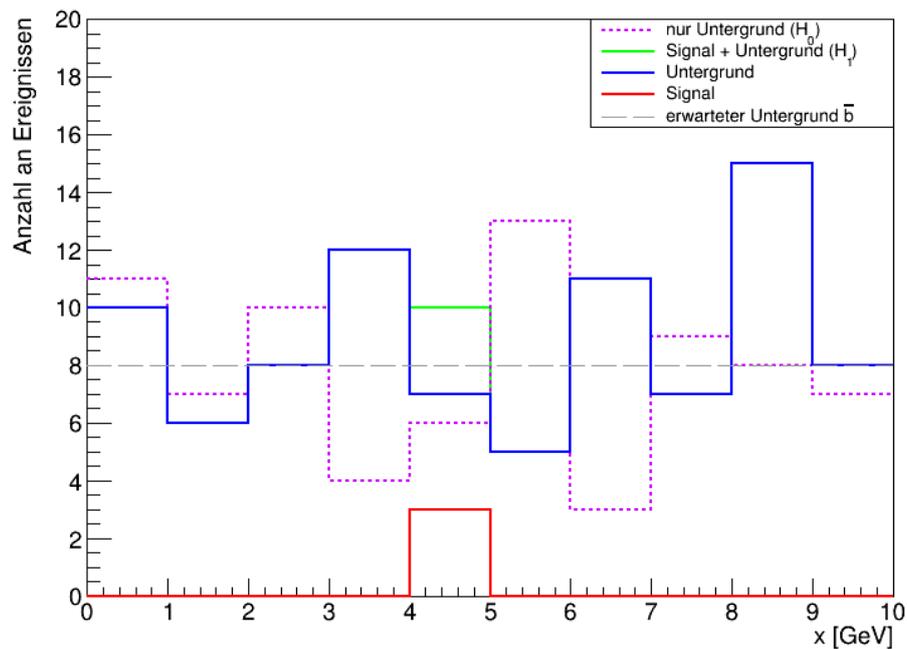


Abbildung 6: Beispielhaftes Spektrum eines Pseudoexperiments, bei dem das echte Signalbin (Bin 5) nicht die höchste Anzahl an Ereignissen aufweist, da der Untergrund in der Umgebung des Signalbins höher fluktuiert. In diesem Fall wird das echte Signal an einer anderen Position im Massenspektrum vermutet (in Bin 9) und daher nicht korrekt identifiziert.

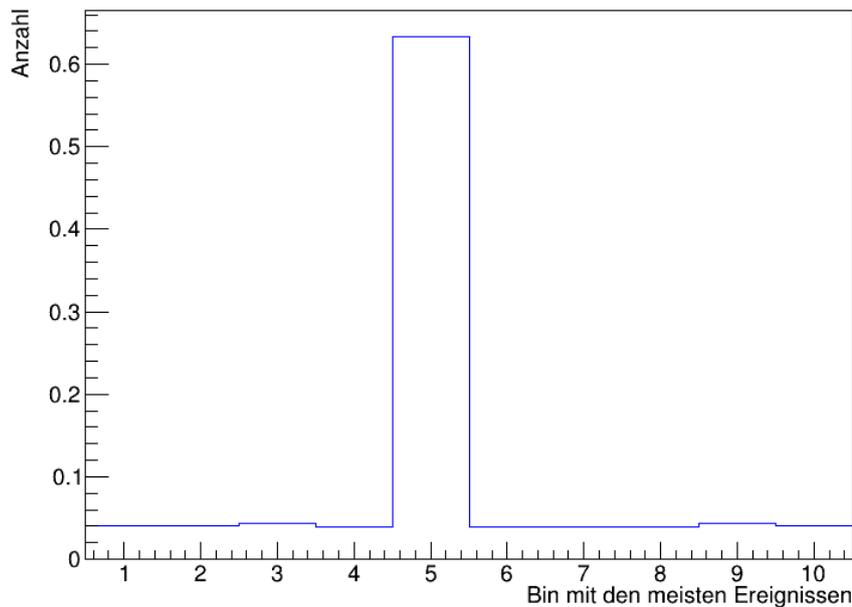


Abbildung 7: Die relative Signalstärke: In 10000 durchgeführten Pseudoexperimenten mit acht erwarteten Untergrundereignissen pro Bin und sechs erwarteten Signalereignissen im Signalbin (Bin 5) weist das Signalbin in etwa 63 % der Fälle die höchste Anzahl an gezählten Ereignissen auf.

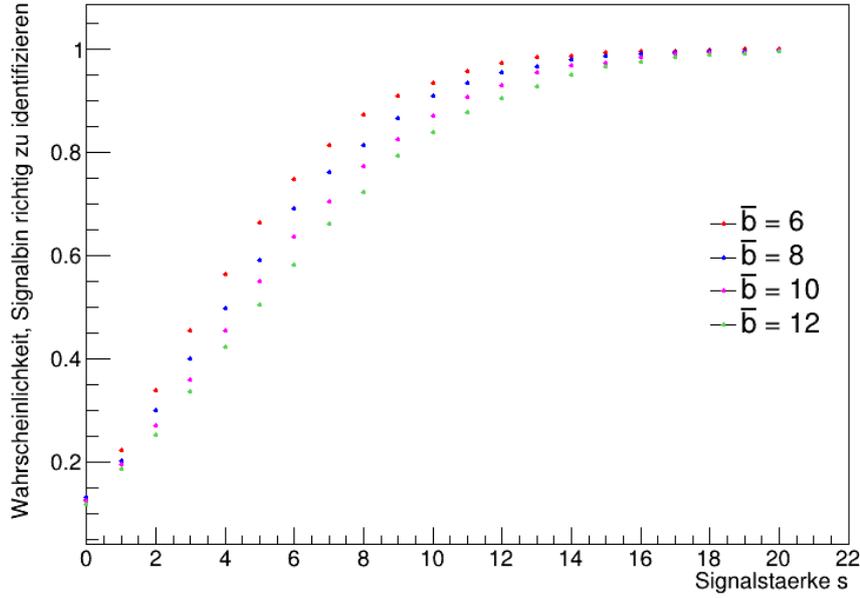


Abbildung 8: Die für verschiedene Signalstärken  $s$  und pro Bin erwartete Untergrundeignisse  $\bar{b}$  berechneten Wahrscheinlichkeiten, das Signalbin in einem Zählerexperiment mit flachem Untergrund und zehn Bins richtig zu identifizieren.

$$q_{\text{lokal}}^{PL} = \begin{cases} -2 \ln \frac{L(s=0|b)}{L(\hat{s}(x)|b)} = -2 n_{\text{sig}} \ln(b) + 2b + 2 n_{\text{sig}} \ln(n_{\text{sig}}) - 2 n_{\text{sig}}, & n_{\text{sig}} \geq b \\ 0 & , \quad n_{\text{sig}} < b \end{cases} \quad (24)$$

In einer Serie an Pseudoexperimenten wird jeweils eine Untergrund-Verteilung generiert und die Anzahl  $n_{\text{sig}}$  der Ereignisse im Signalbin gezählt. Zusammen mit der Anzahl an erwarteten Untergrundeignissen  $\bar{b}$  liefert das Einsetzen in Gleichung 24 für jedes durchgeführte Pseudoexperiment einen Wert  $q_{\text{lokal}}^{PL}$  der Teststatistik. Die zugehörige Teststatistik-Verteilung für acht erwartete Untergrundeignisse und zehn beobachtete Ereignisse entspricht der Verteilung in *Abbildung 9*. Der Anteil der Werte  $q_{\text{lokal}}^{PL} \geq q_{\text{obs}}^{PL}$  entspricht dem lokalen  $p$ -Wert, für den in diesem Beispiel  $p_{\text{lokal}} \approx 0.036$  gilt.

#### 2.4.2 Globaler $p$ -Wert

Ist die Position des Signals im betrachteten Massenspektrum nicht bekannt, ist der beste Schätzer für die Signalstärke  $\hat{s}$  eine Funktion der variablen Masse  $x$  – ihr Schätzer  $\hat{x}$  entspricht der Position des Bins mit der höchsten Zählzahl. Die Berücksichtigung der Ereigniszahlen im gesamten Massenspektrum spiegelt sich in folgender Anpassung der Teststatistik wider:

$$q_{\text{global}}^{PL} = \begin{cases} -2 \ln \frac{L(0|\hat{b}(0))}{L(\hat{s}(\hat{x})|\hat{b})} = -2 n_{\text{max}} \ln(b) + 2b + 2 n_{\text{max}} \ln(n_{\text{max}}) - 2 n_{\text{max}}, & n_{\text{max}} \geq b \\ 0 & , \quad n_{\text{max}} < b \end{cases} \quad (25)$$

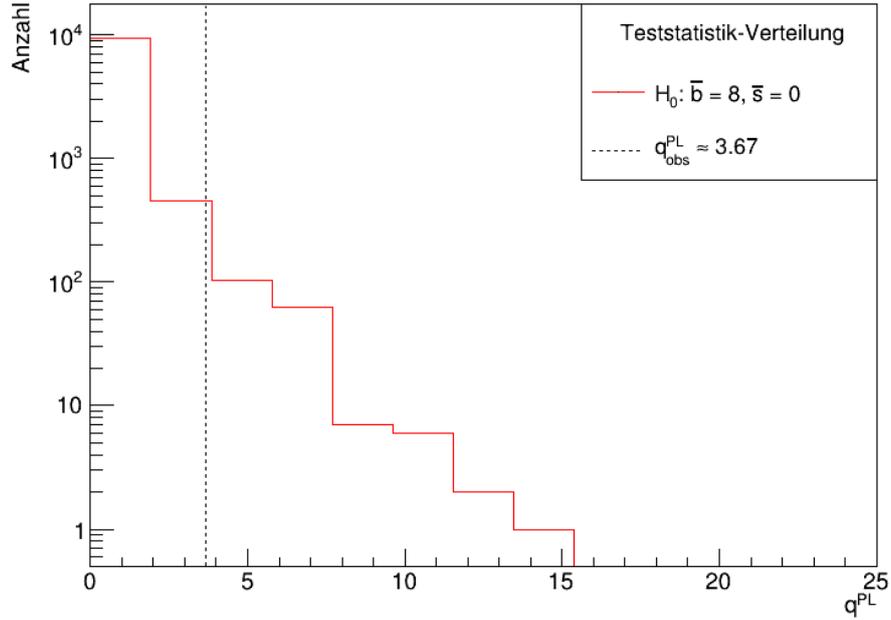


Abbildung 9: Teststatistik-Verteilung unter der Nullhypothese  $H_0$  ( $\bar{b} = 8$ ), die mithilfe des Profile-Likelihood-Quotienten  $p^{PL}$  und 10000 durchgeführten Pseudoexperimenten erzeugt wurde – ohne Berücksichtigung des Look-Elsewhere-Effekts. Die schwarz gestrichelte Linie markiert den Wert  $q_{obs}^{PL}$ , der einer beobachteten Anzahl von zehn Ereignissen entspricht ( $n_{obs} = 10$ ). Der lokale  $p$ -Wert beträgt etwa 0.036.

Die Verteilung der auf diese Weise modifizierten Teststatistik wird analog bestimmt, mit dem Unterschied, dass nun in jedem Pseudoexperiment die Anzahl an Ereignissen  $n_i$  in jedem Bin  $i$  gezählt wird und die höchste Anzahl  $n_{max} = \max\{n_i\}$  zur Berechnung der  $q_{global}^{PL}$ -Werte genutzt wird. Die so erzeugte Teststatistik-Verteilung ist in *Abbildung 10* dargestellt. Der unter dieser Teststatistik-Verteilung berechnete  $p_b$ -Wert heißt globaler  $p$ -Wert, da er die Signifikanz in allen Bins im untersuchten Massenspektrum berücksichtigt. In diesem Beispiel gilt  $p_{global} \approx 0.303$ .

Ein Vergleich der Teststatistiken zeigt, dass diejenige Teststatistik, die den Look-Elsewhere-Effekt berücksichtigt, breiter ist als diejenige, die aus einer Analyse bei einer bestimmten Masse hervorgeht. Diese Beobachtung spiegelt sich in dem größer gewordenen Median der Verteilung wider. In der Konsequenz erhöht sich der Anteil der Werte  $q_{global}^{PL} \geq q_{obs}^{PL}$ : Der globale  $p$ -Wert im vorliegenden Beispiel ist etwa um den Faktor 8 größer als der lokale  $p$ -Wert. Dieser Faktor wird auch als Trial Factor bezeichnet; mit ihm kann die Auswirkung des Look-Elsewhere-Effekts auf die beobachtete Signifikanz ausgedrückt werden:

$$F_{trial} = \frac{\int_{t_{obs}} f(t_{global} | H_0) dt_{global}}{\int_{t_{obs}} f(t_{lokal} | H_0) dt_{lokal}} = \frac{p_{global}}{p_{lokal}} \quad (26)$$

$$\iff p_{global} = F_{trial} \cdot p_{lokal} \quad (27)$$

Offensichtlich reduziert die Berücksichtigung des Look-Elsewhere-Effekts die scheinbare Signifikanz einer Messung, die sich aus der Berechnung des lokalen  $p$ -Werts ergibt. Anders formuliert, erhöht der Look-Elsewhere-Effekt die Likelihood eines Peaks im Massenspektrum, die notwendig ist, um einen bestimmten  $p_b$ -Wert zu erreichen [12]. Die Berücksichtigung dieses Effekts in der

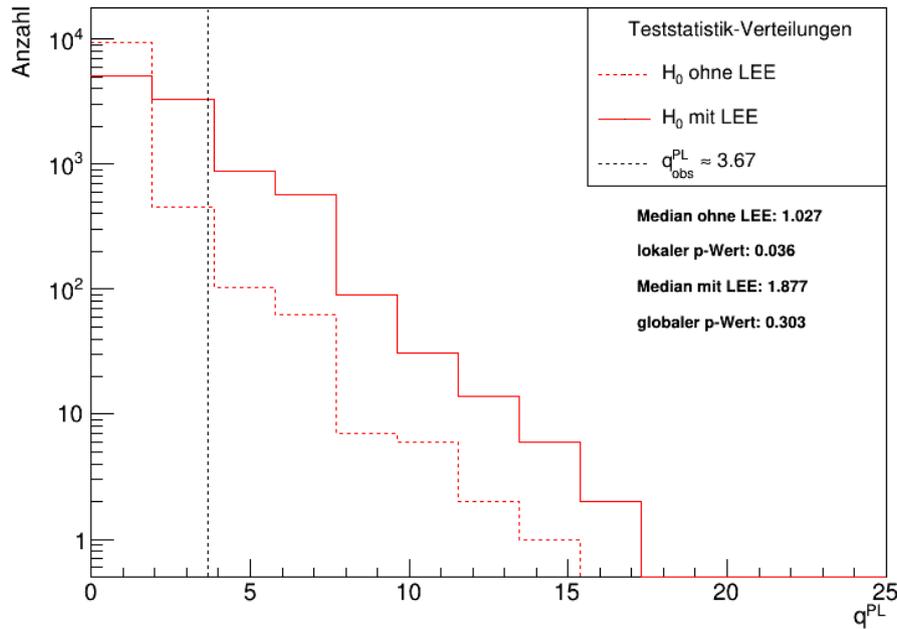


Abbildung 10: Teststatistik-Verteilungen unter der Nullhypothese  $H_0$  ( $\bar{b} = 8$ ), die mithilfe des Profile-Likelihood-Quotienten  $p^{PL}$  und 10000 durchgeführten Pseudoexperimenten erzeugt wurden – einmal ohne Berücksichtigung des Look-Elsewhere-Effekts (gestrichelte Verteilung) und einmal mit (durchgezogene Verteilung). Der Vergleich zeigt, dass der Look-Elsewhere-Effekt (LEE) den Median der Verteilung erhöht und so den lokalen  $p$ -Wert des beobachteten Messwerts (schwarz gestrichelte Linie  $\hat{=} n_{obs} = 10$ ) vergrößert: hier etwa um den Faktor 8,42.

statistischen Auswertung der Messdaten verringert daher das Risiko, eine nur scheinbar signifikante Überhöhung über dem Untergrund eines Massenspektrums fälschlicherweise als echtes Signal zu interpretieren.

## 3 Ausschlussgrenzen auf die Stärke eines Signals

Kann die Untergrund-Hypothese nicht verworfen werden, da der  $p_b$ -Wert eines durchgeführten Experiments größer ist als das vorher festgelegte Signifikanzniveau  $\alpha_b$ , können mithilfe der beobachteten Daten obere Grenzwerte  $s_{up}$  auf die Signalstärke gesetzt werden.

In diesem Kapitel wird anhand von drei Methoden vorgestellt, wie sich diese Ausschlussgrenzen bestimmen lassen. Dafür ist es von grundlegender Bedeutung, die Nullhypothese und die Alternativhypothese entsprechend des sich geänderten Vorhabens neu zu definieren. Im Anschluss werden Konfidenzintervalle und Konfidenzniveaus eingeführt, mit denen schließlich obere Grenzen auf die Signalstärke bestimmt werden.

### 3.1 Nullhypothese und Alternativhypothese

Ein wesentlicher Unterschied zu der bisher betrachteten Suche nach einer Entdeckung besteht darin, dass für die Bestimmung von oberen Ausschlussgrenzen auf die Signalstärke nicht mehr die Untergrundhypothese gegen eine Signalthypothese getestet wird, sondern eine Signalthypothese gegen die Untergrundhypothese. Die Signalthypothese übernimmt damit die Funktion der Nullhypothese  $H_0$  und die Untergrundhypothese die der relevanten Alternativhypothese  $H_1$  [10]:

$$\text{Ausschlussgrenzen} \begin{cases} H_0: \text{Hypothese, die Signal- und Untergrundprozesse beschreibt} \\ H_1: \text{Hypothese, die nur Untergrundprozesse beschreibt} \end{cases}$$

Analog zu dem in Kapitel 2 beschriebenen Vorgehen, besteht das Ziel darin, die Nullhypothese  $H_0$  zugunsten der Alternativhypothese  $H_1$  zu verwerfen. Ein kleiner  $p_{s+b}$ -Wert entspricht nun der Wahrscheinlichkeit, ein Messergebnis zu erhalten, das genauso oder weniger verträglich mit der untersuchten Signalthypothese ist. Ein vorher festgelegtes Signifikanzniveau  $\alpha_{s+b}$  entscheidet darüber, ob eine Messung statistisch signifikant ist und die Signalthypothese verworfen wird. *Abbildung 11* zeigt analog zu *Abbildung 4* die auf 1 normierten Verteilungen der Teststatistik  $n$  unter der Nullhypothese und der Alternativhypothese sowie ein beispielhaftes Messergebnis  $n_{obs}$  mit dem entsprechenden  $p_{s+b}$ - und  $p_b$ -Wert. Trotz der Ähnlichkeit der beiden Abbildungen und der beibehaltenen Farben für Signal- und Untergrundhypothese, ist es wichtig zu beachten, dass die beiden Hypothesen ihre Funktionen getauscht haben: Die Signalthypothese ist nun die getestete Nullhypothese  $H_0$  und die Untergrundhypothese die relevante Alternativhypothese  $H_1$ .

### 3.2 Konfidenzintervalle und Konfidenzniveaus

Um Konfidenzintervalle und Konfidenzniveaus einzuführen, wird im Folgenden ein Zählexperiment betrachtet, bei dem  $\bar{b}$  Untergrundereignisse erwartet und  $n_{obs}$  Ereignisse beobachtet wurden. Zwischen letzteren soll nicht unterschieden werden können, ob es sich um Untergrund- oder Signalereignisse handelt. Der beobachtete  $p_b$ -Wert unter der Untergrundhypothese sei größer als das zuvor festgelegte Signifikanzniveau  $\alpha_b$ , sodass die Messung in Bezug auf eine mögliche Entdeckung nicht als statistisch signifikant angesehen werden kann.

Um nun eine obere Ausschlussgrenze  $s_{up}$  auf die Stärke eines möglichen Signals zu setzen, wird eine Signalthypothese  $H_0$  bestimmt, die mit der Anzahl an beobachteten Ereignissen  $n_{obs}$  gerade noch verträglich ist. Das bedeutet, dass – unter der Annahme, dass die Signalthypothese wahr ist – die Wahrscheinlichkeit bei  $\bar{b}$  erwarteten Untergrundereignissen  $n_{obs}$  oder weniger Ereignisse

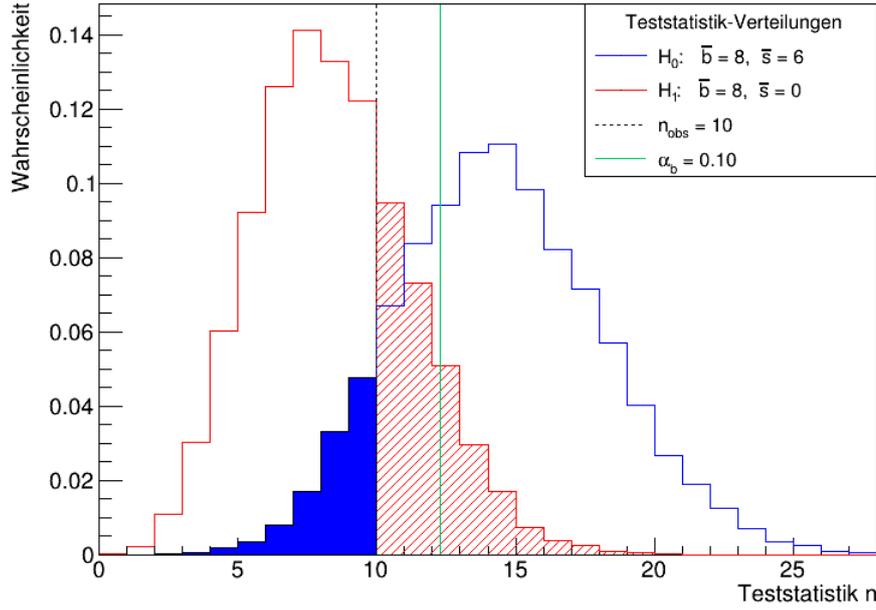


Abbildung 11: Auf 1 normierte Teststatistik-Verteilungen der Untergrundhypothese (rot, linke Verteilung) und der Signalthypothese (blau, rechte Verteilung). Ist der bei der Suche nach einer Entdeckung beobachtete  $p_b$ -Wert (rot schraffiert) größer als das Signifikanzniveau  $\alpha_b$  (grün durchgezogene Linie), können obere Ausschlussgrenzen auf die Signalstärke bestimmt werden. Dazu muss eine Signalstärke ermittelt werden, für die der  $p_{s+b}$ -Wert (blau eingefärbt) dem vorher festgelegten Signifikanzniveau  $\alpha_{s+b}$  entspricht.

zu beobachten, gerade dem Signifikanzniveau  $\alpha_{s+b}$  entspricht:

$$\alpha_{s+b} = P(n \leq n_{obs} | H_0) = \sum_{n=0}^{n_{obs}} \frac{(s_{up} + \bar{b})^n}{n!} e^{-(s_{up} + \bar{b})}. \quad (28)$$

Die so erhaltene Signalthypothese beschreibt ein Signal der Stärke  $s = s_{up}$ . Für alle Signalthypothesen mit  $\bar{s} > s_{up}$  gilt dann, dass ihre  $p_{s+b}$ -Werte, bei gleicher Anzahl an beobachteten Ereignissen  $n_{obs}$ , kleiner sind als das Signifikanzniveau  $\alpha_s$ . Diese Signalthypothesen werden daher zugunsten der Untergrundhypothese  $H_1$  verworfen und damit Signale der Signalstärken  $s > s_{up}$  ausgeschlossen<sup>2</sup>[7].

Die Wahrscheinlichkeit, die Signalthypothese mit  $\bar{s} = s_{up}$  irrtümlich zu verwerfen, wenn sie wahr ist, entspricht dem gewählten Signifikanzniveau  $\alpha_{s+b}$ . Daraus folgt, dass der Wert  $s_{up}$  von einem Ensemble an unter gleichen Bedingungen durchgeführten Experimenten in  $CL = 1 - \alpha_s$  Prozent aller Fälle korrekt bestimmt wird und das endliche Intervall mit oberer Grenze  $s_{up}$  die wahre Signalstärke enthält. Diese Wahrscheinlichkeit  $CL$  wird als Konfidenzniveau bezeichnet, das beschriebene Intervall als Konfidenzintervall.

Ein Konfidenzniveau gibt an, mit welcher Wahrscheinlichkeit bei der zuvor durchgeführten Suche nach einer Entdeckung ein kleinerer als der beobachtete  $p_{b,obs}$ -Wert unter der Untergrundhypo-

<sup>2</sup> Je nach Definition einer oberen Ausschlussgrenze, ist es auch üblich, Signalstärken auszuschließen, für die  $s \geq s_{up}$  gilt. Die hier genutzte Definition wurde Ref. [7] entnommen.

these hätte beobachtet werden müssen, wenn das gesuchte Signal die Stärke  $s = s_{up}$  besitzt [13]:

$$CL = P(n \geq n_{obs} | H_0) = \sum_{n=n_{obs}}^{\infty} \frac{(\bar{s} + \bar{b})^n}{n!} e^{-(\bar{s} + \bar{b})} = 1 - \alpha_{s+b}. \quad (29)$$

Je höher diese Wahrscheinlichkeit ist, desto aussagekräftiger ist die gesuchte Ausschlussgrenze  $s_{up}$ . Beispielsweise hätte bei einem Konfidenzniveau von  $CL = 95\%$  – und einem damit verbundenen Signifikanzniveau von  $\alpha_{s+b} = 5\%$  – mit einer Wahrscheinlichkeit von  $95\%$  ein kleinerer  $p_b$ -Wert beobachtet werden müssen, wenn das Signal die Stärke  $s = s_{up}$  hat. Die blau schraffierte Fläche unter der Signalthypothese in *Abbildung 12* veranschaulicht ein solches  $95\%$ -Konfidenzniveau.

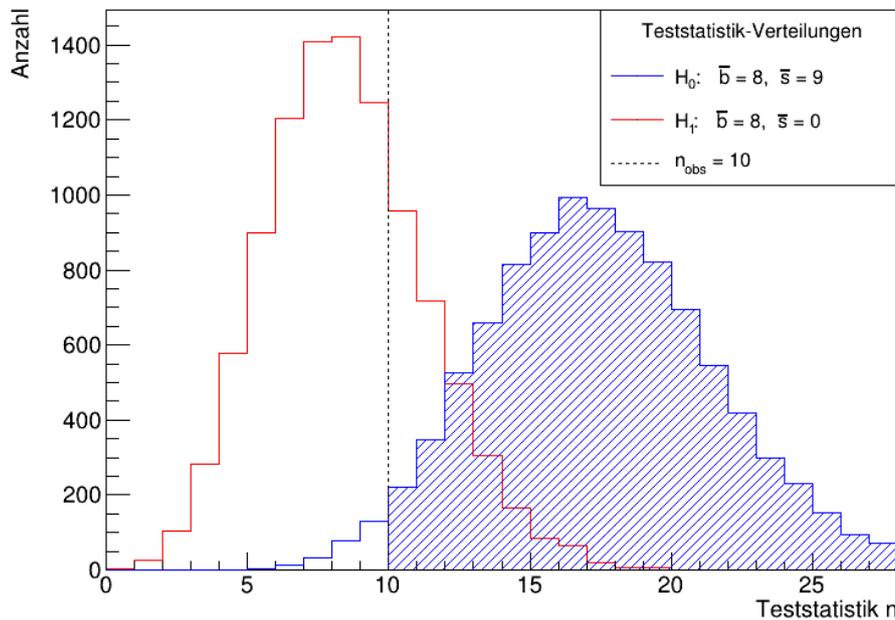


Abbildung 12: Die blau schraffierte Fläche unter der Signalthypothese  $H_0$  veranschaulicht das Konfidenzniveau  $CL = 95\%$ . Sie entspricht der Wahrscheinlichkeit, einen kleineren als den unter der Untergrundhypothese (rot, linke Verteilung) beobachteten  $p_b$ -Wert zu erhalten, wenn das gesuchte Signal die Stärke  $s = s_{up}$  hat.

Abbildung 13 zeigt den Zusammenhang zwischen Konfidenzniveaus und den Ausschlussgrenzen  $s_{up}$  für den Fall, dass acht Untergrundereignisse erwartet und zehn Ereignisse beobachtet wurden. Das häufig gewählte Konfidenzniveau  $CL = 95\%$ , das auch in den folgenden Kapiteln verwendet wird, ist durch eine rote Linie gekennzeichnet.

### 3.3 Obere Grenzen auf die Signalstärke

Im Folgenden werden für das Konfidenzniveau  $CL = 95\%$  die oberen Ausschlussgrenzen  $s_{up}$  auf die Signalstärke bestimmt. Dies wird auf drei Arten erreicht: durch eine direkte Methode, die sich unmittelbar aus der Definition des Konfidenzniveaus ergibt, durch die CLs-Methode und mithilfe des Bayesianischen Ansatzes. Als Teststatistik wird die Anzahl  $n$  an Ereignissen genutzt. Zusätzlich wird die direkte Methode auf den Likelihood-Quotienten  $q^{NP}$  nach Neyman und Pearson angewendet.

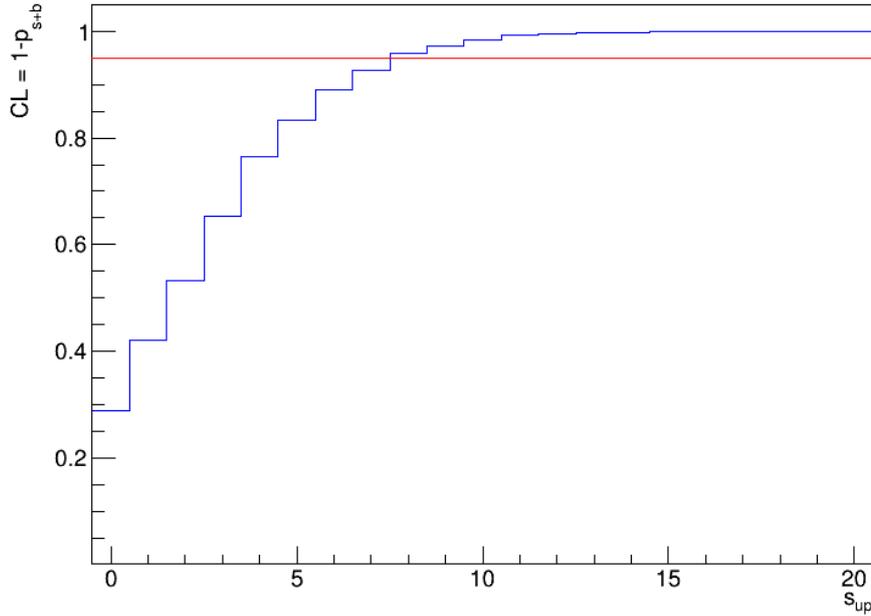


Abbildung 13: Zusammenhang zwischen Konfidenzniveaus und unterschiedlichen oberen Ausschlussgrenzen  $s_{up}$  bei acht erwarteten Untergrundereignissen  $\bar{b}$  und zehn beobachteten Ereignissen  $n_{obs}$ . Die rote Linie markiert das häufig gewählte Konfidenzniveau  $CL = 95\%$ .

### 3.3.1 Direkte Methode

Wie in Kapitel 3.2 beschrieben, besteht die Aufgabe darin, eine Signalstärke  $s_{up}$  zu bestimmen, die – für eine gegebene Anzahl an erwarteten Untergrundereignissen  $\bar{b}$  und eine Anzahl an beobachteten Ereignissen  $n_{obs}$  – gerade noch nicht zugunsten der Untergrundhypothese  $H_1$  verworfen wird. Wann dies der Fall ist, wird durch das gewählte Konfidenzniveau  $CL$  festgelegt. In dem hier betrachteten Fall von  $CL = 95\%$  gilt für die gesuchte Signalthypothese nach Gleichung 29:

$$CL = P(n \geq n_{obs} | H_0) = \sum_{n=n_{obs}}^{\infty} \frac{(s_{up} + \bar{b})^n}{n!} e^{-(s_{up} + \bar{b})} = 95\%. \quad (30)$$

Ebenso möglich ist es, die äquivalente Wahrscheinlichkeit dafür zu berechnen, dass die Wahrscheinlichkeit, weniger als  $n_{obs}$  Ereignisse zu beobachten, wenn die Signalthypothese mit Stärke  $s_{up}$  wahr ist, gerade dem Signifikanzniveau  $\alpha_{s+b} = 1 - CL$  entspricht:

$$\alpha_{s+b} = P(n < n_{obs} | H_0) = \sum_{n=0}^{n_{obs}} \frac{(s_{up} + \bar{b})^n}{n!} e^{-(s_{up} + \bar{b})} = 5\%. \quad (31)$$

Um dies zu erreichen, wird die erwartete Signalstärke  $\bar{s}$  für die jeweils betrachtete Anzahl an erwarteten Untergrundereignissen  $\bar{b}$  solange bei einem gegebenen Wert für  $n_{obs}$  variiert, bis Gleichung 31 erfüllt ist. Das so gefundene Signal hat dann eine Stärke von  $s = s_{up}$ . *Abbildung 14* zeigt die so erhaltenen Ausschlussgrenzen für unterschiedliche Werte von  $\bar{b}$  und  $n_{obs}$ .

Es lässt sich erkennen, dass die oberen Grenzen  $s_{up}$  umso höher sind, je weniger Untergrunder-

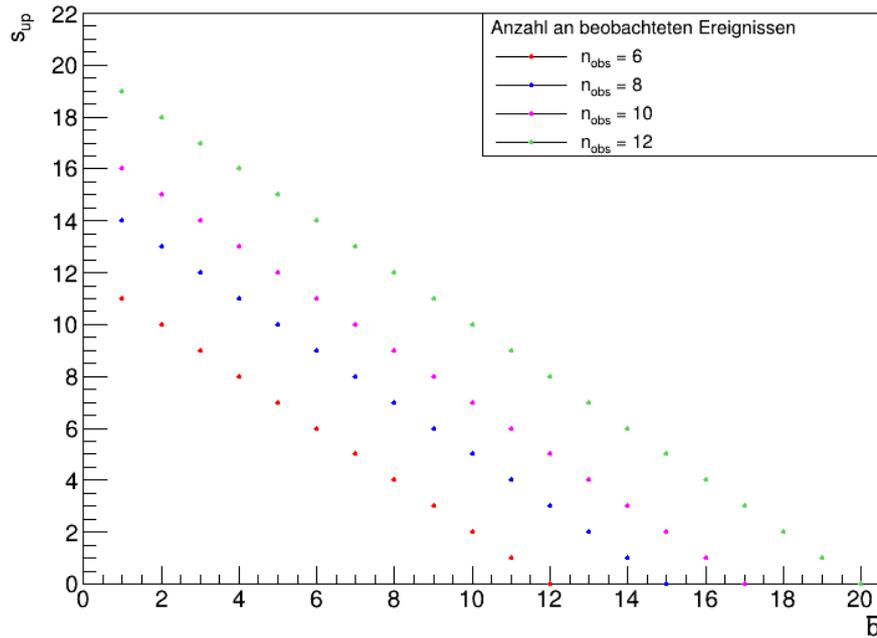


Abbildung 14: Mit der direkten Methode und der Teststatistik  $n$  bei dem Konfidenzniveau  $CL = 95\%$  ermittelte Ausschlussgrenzen  $s_{up}$  für verschiedene Anzahlen an erwarteten Untergrundereignissen  $\bar{b}$  und beobachteten Ereignissen  $n_{obs}$ . Ist  $\bar{b}$  groß im Vergleich zu  $n_{obs}$ , liefert die direkte Methode negative, d. h. unphysikalische, Ausschlussgrenzen.

eignisse erwartet werden. Da in diesen Fällen deutlich mehr Ereignisse  $n_{obs}$  beobachtet als Untergrundprozesse  $\bar{b}$  erwartet werden, ist auch ein starkes Signal noch mit der Beobachtung verträglich und das Konfidenzintervall mit oberer Grenze  $s_{up}$  entsprechend größer.

Eine weitere Beobachtung ist, dass die direkte Methode zu negativen Ausschlussgrenzen führen kann, wenn der erwartete Untergrund  $\bar{b}$  groß ist im Vergleich zu der Anzahl an beobachteten Ereignissen  $n_{obs}$ . Negative Grenzen  $s_{up}$  sind physikalisch nicht sinnvoll, da nicht weniger als null Ereignisse beobachtet werden können. Die direkte Methode ist zwar auch in diesen Fällen korrekt, liefert aber ein Konfidenzintervall, das, im Fall eines 95 %-Konfidenzniveaus, gerade zu den 5 % gehört, die den wahren Wert der Signalstärke nicht enthalten.

Werden die oberen Ausschlussgrenzen wie in *Abbildung 15* für eine bestimmte Anzahl an erwarteten Untergrundereignissen  $\bar{b}$  gegen unterschiedliche  $n_{obs}$  aufgetragen, fällt auf, dass die oberen Grenzen ab einem bestimmten  $n_{obs}$  null sind. Der Punkt, an dem das erreicht ist, entspricht dem 5%-Quantil der Untergrundhypothese, das in *Abbildung 16* dargestellt ist: Hier liegt die Wahrscheinlichkeit, bei der Suche nach einer Entdeckung einen kleineren als den beobachteten  $p_b^{obs}$ -Wert zu erhalten, per Definition bereits bei 95 %. Die obere Grenze entspricht daher für alle  $n \leq n_{obs}$  einem Signal der Stärke  $s = s_{up} = 0$ , womit die Teststatistiken  $H_0$  und  $H_1$  identisch sind.

Die direkte Methode lässt sich analog auf andere Teststatistiken anwenden. Im Falle des Likelihood-Quotienten  $q^{NP}$  nach Neyman und Pearson ist zu beachten, dass die Teststatistik-Verteilungen der Signal- und der Untergrundhypothese gemäß *Abbildung 2* die Seiten tauschen. Die Wahrscheinlichkeit,  $n \geq n_{obs}$  Ereignisse zu beobachten, entspricht nun dem Anteil der Werte  $q^{NP}$ , die gleich oder kleiner sind als  $q_{obs}^{NP}$ . Für die gesuchte Ausschlussgrenze  $s_{up}$  wird die erwartete Signalstärke  $\bar{s}$  wieder variiert, bis gilt:

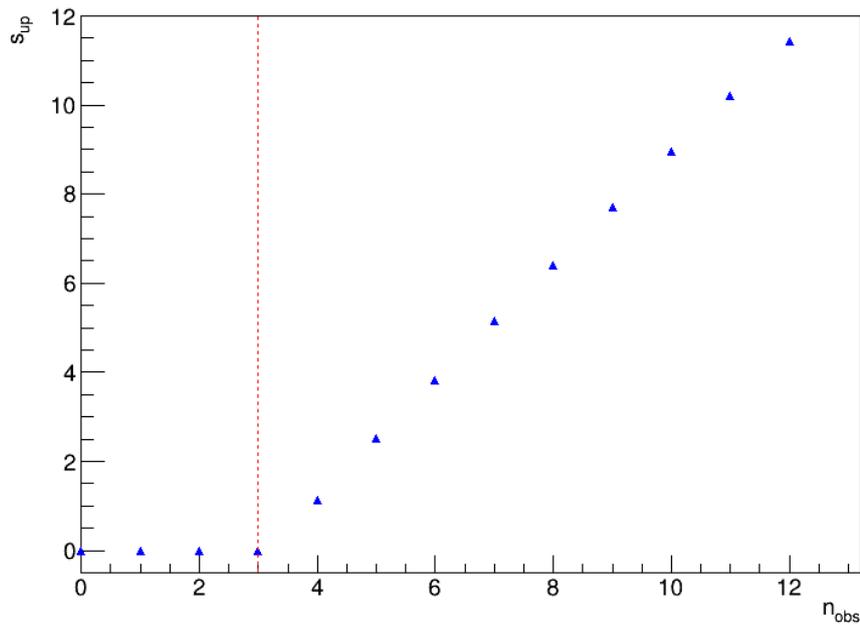


Abbildung 15: Für ein 95 %-Konfidenzintervall mit der Teststatistik  $n$  bestimmte Ausschlussgrenzen  $s_{\text{up}}$  für unterschiedliche Anzahlen an beobachteten Ereignissen  $n_{\text{obs}}$  und acht erwarteten Untergrundereignissen  $\bar{b}$ . Werden  $n \leq 3$  Ereignisse beobachtet, gilt für die oberen Grenzen der Konfidenzintervalle  $s_{\text{up}} = 0$  – in diesem Fall sind die Teststatistiken für  $H_0$  und  $H_1$  identisch. Die rote Linie markiert die obere Grenze des 5 %-Quantils der Teststatistik-Verteilung von  $H_1$ .

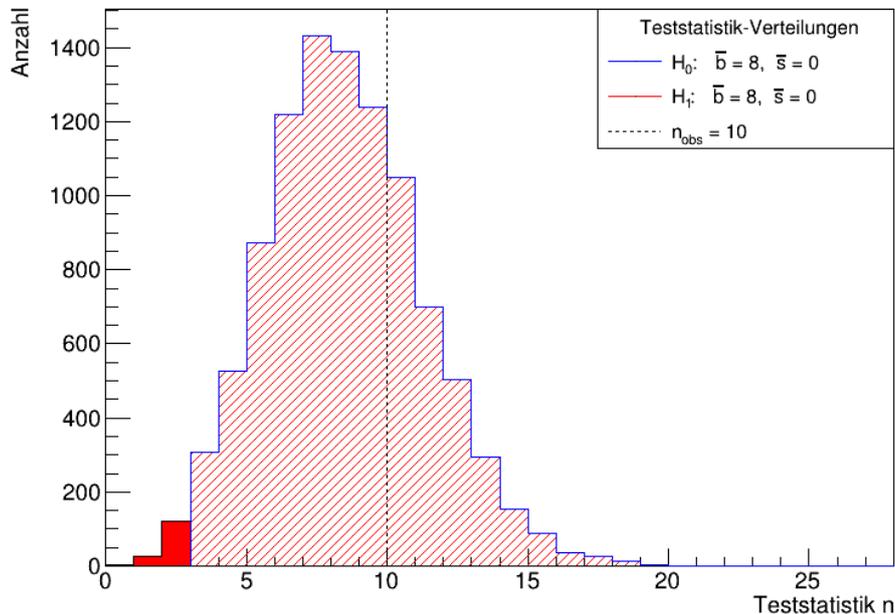


Abbildung 16: Die rot eingefärbte Fläche ist das 5 %-Quantil der Untergrundhypothese. Für die darin liegenden Ereigniszahlen  $n \leq 3$  gilt  $s_{\text{up}} = 0$ . Die rot schraffierte Fläche entspricht der Wahrscheinlichkeit (95 %), schon bei einer erwarteten Signalstärke von  $\bar{s} = 0$  mehr als drei Ereignisse zu beobachten. In diesem Fall stimmen die Signalthypothese  $H_0$  mit  $s_{\text{up}} = 0$  und die Untergrundhypothese  $H_1$  überein.

$$CL = P(q^{NP} \leq q_{obs}^{NP} | H_0) = 95 \% . \quad (32)$$

Mit  $\alpha_{s+b} = 1 - CL$  folgt äquivalent:

$$\alpha_{s+b} = P(q^{NP} > q_{obs} | H_0) = 5 \% . \quad (33)$$

Da  $q^{NP}$  eine kontinuierliche Teststatistik ist, lassen sich die Ausschlussgrenzen genauer bestimmen, aber auch hier sind negative Grenzen  $s_{up}$  möglich. Die Ergebnisse sind in *Abbildung 17* und *Abbildung 18* dargestellt.

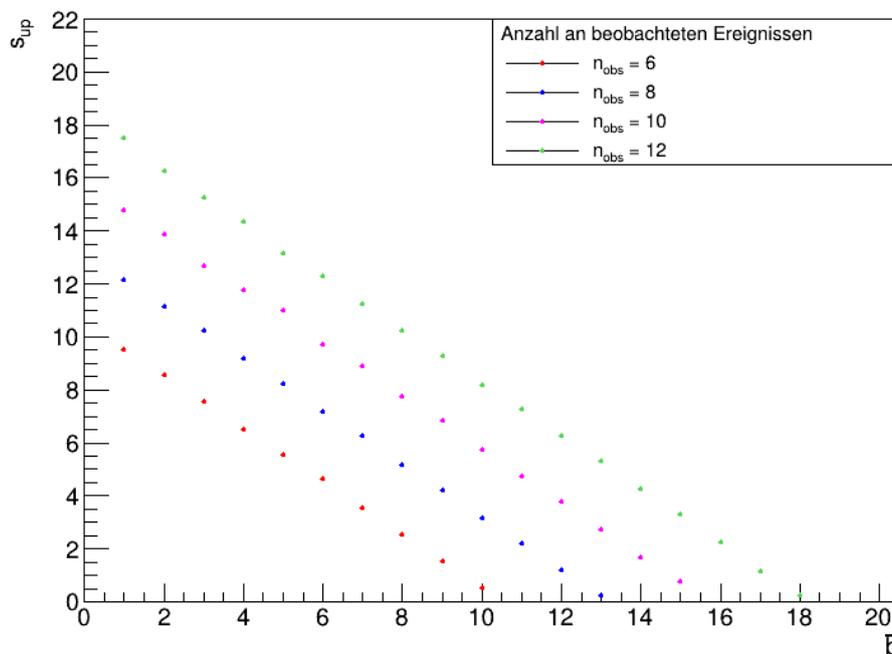


Abbildung 17: Mit der direkten Methode und der Teststatistik  $q^{NP}$  bei einem Konfidenzniveau  $CL = 95 \%$  ermittelte Ausschlussgrenzen  $s_{up}$  für verschiedene Anzahlen an erwarteten Untergrundereignissen  $\bar{b}$  und beobachteten Ereignissen  $n_{obs}$ . Auch hier liefert die direkte Methode negative Ausschlussgrenzen, wenn  $\bar{b}$  groß ist im Vergleich zu  $n_{obs}$ .

### 3.3.2 CLs-Methode

Hat das untersuchte Signal nur eine kleine Signalstärke  $s$ , unterscheiden sich Untergrund- und Signalthypothese kaum voneinander. In diesem Fall kann es passieren, dass allein aufgrund statistischer Fluktuationen Werte für den Parameter  $s$  ausgeschlossen werden, für die das Experiment nicht mehr sensitiv ist. Eine Möglichkeit, dennoch aussagekräftige Ausschlussgrenzen zu bestimmen, bietet die CLs-Methode [14].

Die CLs-Methode modifiziert die im vorangegangenen Kapitel beschriebene direkte Methode, indem sie einen zweiten  $p$ -Wert berücksichtigt: Neben dem  $p_{s+b}$ -Wert unter der Signalthypothese,

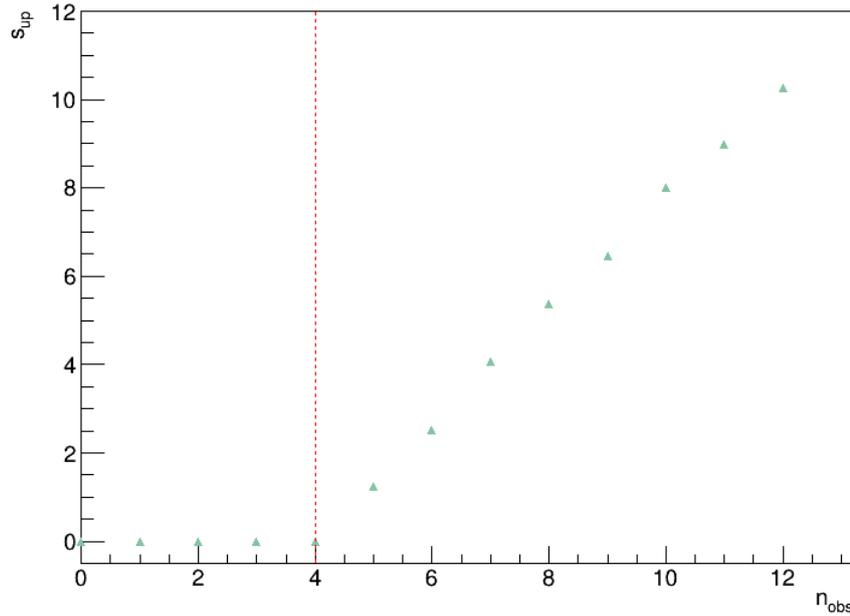


Abbildung 18: Für ein 95 %-Konfidenzintervall mit der Teststatistik  $q^{NP}$  bestimmte Ausschlussgrenzen  $s_{\text{up}}$  für unterschiedliche Anzahlen an beobachteten Ereignissen  $n_{\text{obs}}$  und acht erwarteten Untergrundereignissen  $\bar{b}$ . Werden  $n \leq 4$  Ereignisse beobachtet, gilt für die oberen Grenzen der Konfidenzintervalle  $s_{\text{up}} = 0$  – in diesem Fall sind die Teststatistiken für  $H_0$  und  $H_1$  identisch. Die rote Linie markiert die obere Grenze des 5 %-Quantils der Teststatistik-Verteilung von  $H_1$ .

wird nun auch der  $p_b$ -Wert unter der Untergrundhypothese in die Bestimmung der Ausschlussgrenzen einbezogen. Mithilfe dieser beiden  $p$ -Werte wird eine neue Größe definiert, die den bisher für die  $s_{\text{up}}$ -Bestimmung genutzten  $p_s$ -Wert um den Faktor  $1 - p_b$  korrigiert [8]:

$$\text{CLs} = \frac{p_{s+b}}{1 - p_b}. \quad (34)$$

Die Signalstärke  $s$ , für die  $\text{CLs} = \alpha_{s+b}$  gilt, beschreibt ein Signal mit  $s_{\text{up}}$ . Sie wird analog zu dem in Kapitel 3.3.1 beschriebenen Vorgehen bestimmt, indem  $\bar{s}$  solange variiert wird, bis die Bedingung  $\text{CLs} = \alpha_{s+b}$  erfüllt ist. Alle Signale der Stärken  $s > s_{\text{up}}$  werden ausgeschlossen, da für sie  $\text{CLs} < \alpha_{s+b}$  gilt.

Die CLs-Methode bietet den Vorteil, dass die Größe CLs für sich stark überlappende Hypothesen aufgrund des entsprechend großen  $p_b$ -Werts im Nenner  $1 - p_b$  nicht zu klein wird. Dadurch werden auch bei geringer Sensitivität eines Hypothesentests Signalstärken noch zugelassen, die sonst allein wegen statistischer Fluktuationen ausgeschlossen werden könnten.

Es ist zu beachten, dass die Größe CLs das Verhältnis zweier Wahrscheinlichkeiten ist und daher selbst keine Wahrscheinlichkeit darstellt. Da  $p_b \leq 1$  gilt, und folglich  $\text{CLs} \leq p_s$ , erhöht die CLs-Methode die Konfidenzniveaus im Vergleich zur direkten Methode und führt zu größeren Konfidenzintervallen. Dies kann *Abbildung 19* entnommen werden, die noch einmal zeigt, dass die mit der CLs-Methode erhaltenen Ausschlussgrenzen weniger stringent sind als diejenigen, die mit der direkten Methode bestimmt wurden. Eine weitere Konsequenz daraus lässt sich

in *Abbildung 20* erkennen: Selbst für den Fall, dass keine Ereignisse beobachtet werden, d. h.  $n_{obs} = 0$  gilt, folgt für die oberen Grenzen auf das mögliche Signal  $s_{up} > 0$ .

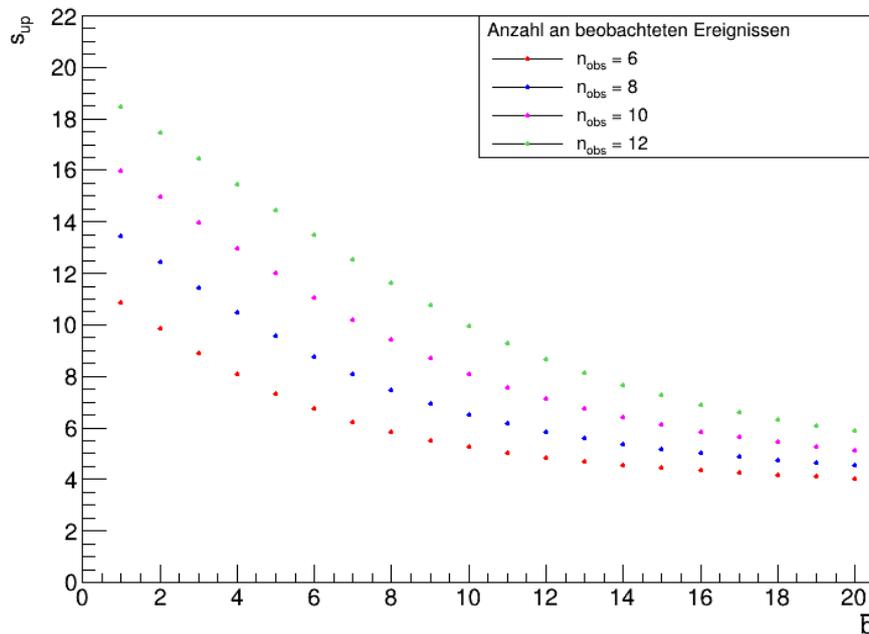


Abbildung 19: Mit der CLs-Methode und der Teststatistik  $n$  bei dem Konfidenzniveau  $CL = 95\%$  ermittelte Ausschlussgrenzen  $s_{up}$  für verschiedene Anzahlen an erwarteten Untergrundereignissen  $\bar{b}$  und beobachteten Ereignissen  $n_{obs}$ .

### 3.3.3 Bayesianischer Ansatz

Mit Ausnahme des Bayesianischen Ansatzes stützen sich alle in der vorliegenden Arbeit durchgeführten Analysen auf den frequentistischen Wahrscheinlichkeitsbegriff – der in der Physik am häufigsten genutzten Möglichkeit, Wahrscheinlichkeit zu interpretieren<sup>3</sup>. Wahrscheinlichkeit wird demnach als relative Häufigkeit aufgefasst, mit der ein Ereignis in einer unendlich langen Serie von Zufallsexperimenten auftritt, die unabhängig voneinander unter gleichen Bedingungen durchgeführt werden [15]:

$$P(A) = \lim_{n \rightarrow \infty} \frac{\text{absolute Wahrscheinlichkeit, dass Ereignis } A \text{ eintritt}}{\text{Gesamtzahl } n \text{ der Versuche}}. \quad (35)$$

Trotz einer in der Praxis endlichen Serie von Zufallsexperimenten kann die ermittelte relative Häufigkeit in vielen Fällen als gute Näherung für die Wahrscheinlichkeit eines Ereignisses angesehen werden. Da eine frequentistisch interpretierte Wahrscheinlichkeit ausschließlich von den Ergebnissen beliebig oft wiederholbarer Messungen abhängt, die unter identischen Bedingungen durchgeführt werden, wird ihr ein objektiver Charakter zugeschrieben. Sie wird daher als objektive Wahrscheinlichkeit bezeichnet.

Der Bayesianische Ansatz interpretiert Wahrscheinlichkeiten anders: Er nutzt bereits vorhandene Kenntnisse, Erfahrungen und Intuition, um zu einer besser begründeten Wahrscheinlich-

<sup>3</sup> Da die Größe CLs selbst keine Wahrscheinlichkeit darstellt, ist sie streng genommen weder dem frequentistischen noch dem bayesianischen Wahrscheinlichkeitsbegriff zuzuordnen [8].

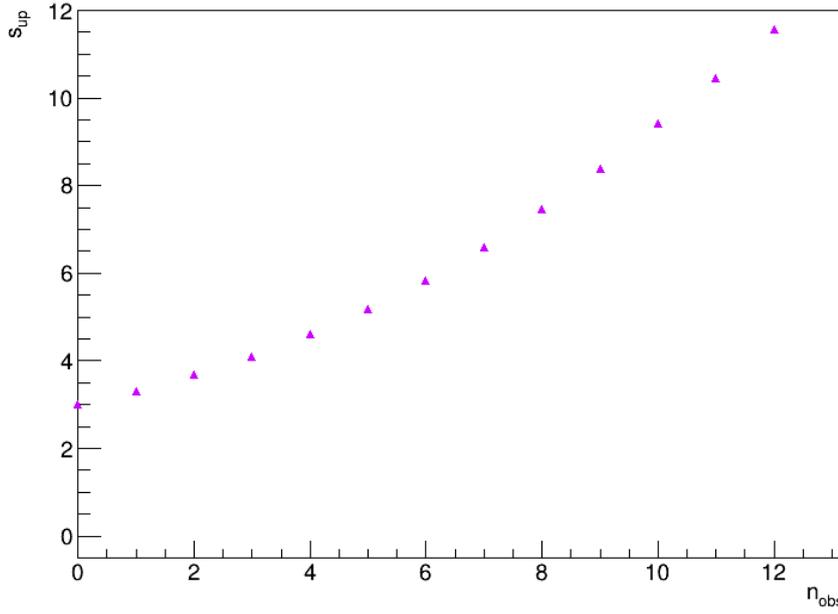


Abbildung 20: Für ein 95 %-Konfidenzintervall mit der CLs-Methode bestimmte Ausschlussgrenzen  $s_{up}$  für unterschiedliche Anzahlen an beobachteten Ereignissen  $n_{obs}$  und acht erwarteten Untergrundereignissen  $\bar{b}$ .

keitsaussage zu kommen. Das hat zur Folge, dass es sich bei einer bayesianisch interpretierten Wahrscheinlichkeit immer um eine subjektive und bedingte Wahrscheinlichkeit handelt, die sich ändert, wenn zusätzliche Informationen gewonnen werden [8]:

$$P(A | B) = \frac{P(B | A) P(A)}{P(B)}. \quad (36)$$

Mithilfe von Gleichung 36, die als Bayes-Theorem bekannt ist, wird einem Einzelereignis A eine Wahrscheinlichkeit  $P(A | B)$  zugeschrieben.  $P(A)$  entspricht dabei dem Grad an Überzeugung, dass Ereignis A eintritt, bevor bekannt ist, dass Ereignis B eintritt;  $P(A)$  wird daher als a-priori-Wahrscheinlichkeit bezeichnet. Demgegenüber steht die a-posteriori-Wahrscheinlichkeit  $P(A | B)$ : Sie entspricht dem Grad an Überzeugung, dass Ereignis A eintritt, nachdem bekannt ist, dass Ereignis B eingetreten ist [8].

Im Kontext der zu bestimmenden Ausschlussgrenzen entspricht  $P(A)$ , jetzt als  $\pi(s)$  bezeichnet, dem Grad an Überzeugung, dass das gesuchte Signal die Stärke  $s$  hat, bevor bekannt ist, wie viele Ereignisse bei einer Messung beobachtet werden.  $P(A | B)$  entspricht dem Grad an Überzeugung, dass das Signal die Stärke  $s$  hat, nachdem eine bestimmte Anzahl an Ereignissen  $n_{obs}$  beobachtet wurde.  $P(B)$  schließlich ist die Wahrscheinlichkeit, eine bestimmte Anzahl an Ereignissen  $n_{obs}$  bei einer beliebigen Signalstärke zu beobachten. Der Faktor  $\frac{1}{P(B)}$  normiert daher die a-posteriori-Wahrscheinlichkeit. Ausgedrückt durch die entsprechenden Likelihoods, nimmt Gleichung 36 folgende Form an:

$$P(s | n_{obs}) = \frac{L(n_{obs} | s)\pi(s)}{\int_0^\infty L(n_{obs} | s')\pi(s')ds'}. \quad (37)$$

Um die obere Ausschlussgrenze  $s_{up}$  zu erhalten, muss gelten:

$$CL = \int_0^{s_{up}} P(s | n_{obs}) ds. \quad (38)$$

Aus  $\alpha = 1 - CL$  folgt die äquivalente Beziehung:

$$\alpha = \int_{s_{up}}^{\infty} P(s | n_{obs}) ds = \frac{\int_{s_{up}}^{\infty} L(n_{obs} | s) \pi(s) ds}{\int_0^{\infty} L(n_{obs} | s) \pi(s) ds}. \quad (39)$$

Für den hier betrachteten Fall eines Zähl-experiments mit Poisson-verteilter Zufallsvariable  $n$  und unter der Annahme einer konstanten a-priori-Wahrscheinlichkeit der Form

$$\pi(s) = \begin{cases} 1, & s \geq 0 \\ 0, & s < 0 \end{cases} \quad (40)$$

wird Gleichung 39 zu:

$$\alpha = \sum_{s=s_{up}}^{\infty} P(s | n_{obs}) = \frac{\sum_{s=s_{up}}^{\infty} L(n_{obs} | s)}{\sum_{s=0}^{\infty} L(n_{obs} | s)} = \frac{e^{-(s_{up}+\bar{b})} \sum_{n=0}^{n_{obs}} \frac{(s_{up}+\bar{b})^n}{n!}}{e^{-\bar{b}} \sum_{n=0}^{n_{obs}} \frac{\bar{b}^n}{n!}} \quad (41)$$

$$\implies \alpha = e^{-s_{up}} \frac{\sum_{n=0}^{n_{obs}} \frac{(s_{up}+\bar{b})^n}{n!}}{\sum_{n=0}^{n_{obs}} \frac{\bar{b}^n}{n!}}. \quad (42)$$

Auch hier führt numerisches Lösen, bei dem  $\bar{s}$  solange variiert wird, bis Gleichung 42 für ein festgelegtes  $\alpha$  (hier:  $\alpha = 5\%$ ) erfüllt ist, zu den gesuchten Ausschlussgrenzen  $s_{up}$ . Die Ergebnisse, dargestellt in *Abbildung 21* und *Abbildung 22*, stimmen mit den Ergebnissen überein, die mithilfe der CLs-Methode erhalten wurden. Im Falle des Likelihood-Quotienten nach Neyman und Pearson lässt sich diese Übereinstimmung für ein Zähl-experiment mit Poisson-verteilter Zufallsvariable auch analytisch nachweisen. Das spiegelt den allgemeinen Fall wider, dass in vielen realistischen Anwendungen und unter Verwendung einer konstanten a-priori-Wahrscheinlichkeit die CLs-Methode und der Bayesianische Ansatz ähnliche Ausschlussgrenzen  $s_{up}$  liefern [8].

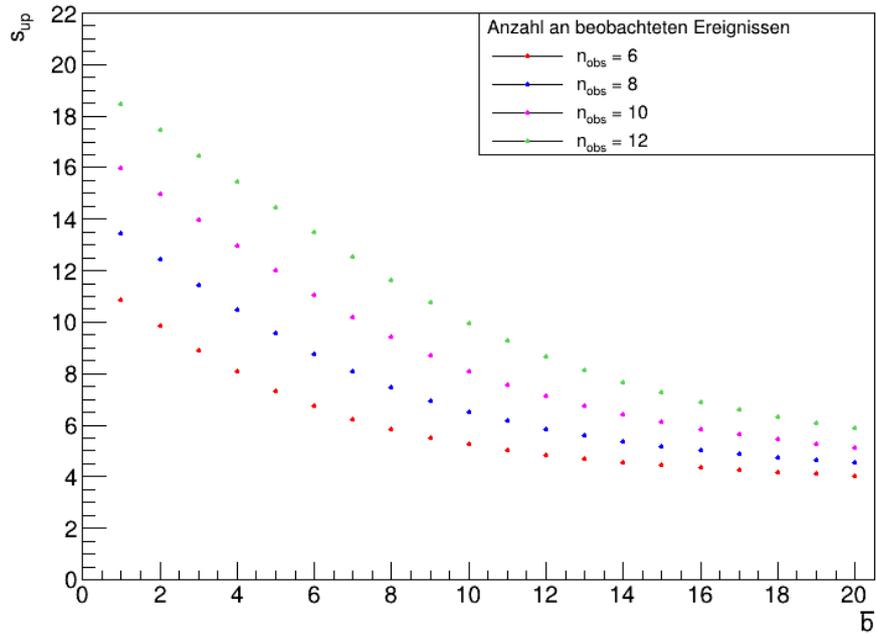


Abbildung 21: Mit dem Bayesianischen Ansatz und der Teststatistik  $n$  bei dem Konfidenzniveau  $CL = 95\%$  ermittelte Ausschlussgrenzen  $s_{up}$  für verschiedene Anzahlen an erwarteten Untergrundereignissen  $\bar{b}$  und beobachteten Ereignissen  $n_{obs}$ .

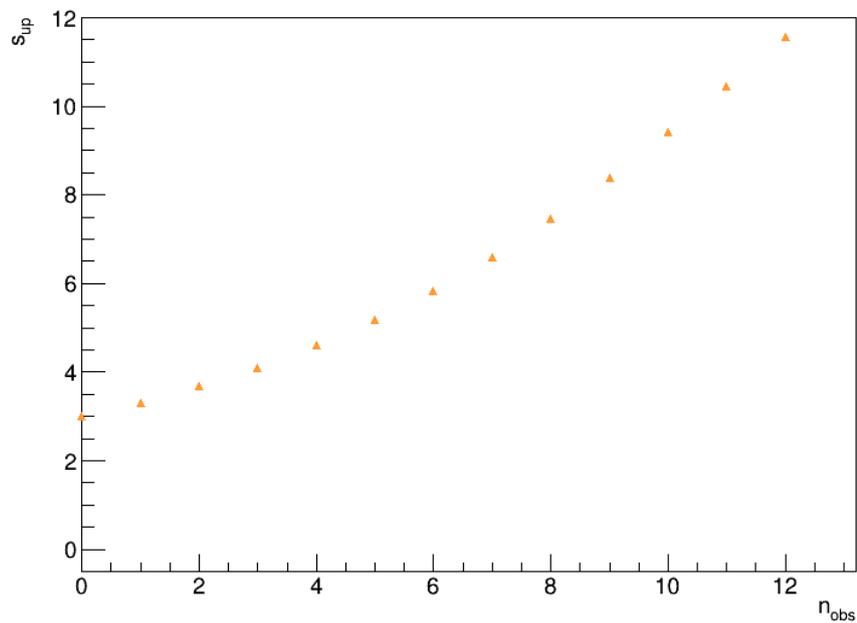


Abbildung 22: Für ein 95%-Konfidenzintervall mit dem Bayesianischen Ansatz bestimmte Ausschlussgrenzen  $s_{up}$  für unterschiedliche Anzahlen an beobachteten Ereignissen  $n_{obs}$  und acht erwarteten Untergrundereignissen  $\bar{b}$ .

## 4 Anwesenheit eines Störparameters

In diesem Kapitel werden der Look-Elsewhere-Effekt und die obere Grenze  $s_{up}$  auf die Stärke eines möglichen Signals für den Fall untersucht, dass der Parameter  $\bar{b}$  des Untergrunds nicht genau bekannt ist. Dazu wird das in Kapitel 2.4 vorgestellte Beispielspektrum einer invarianten Masse noch einmal betrachtet, wobei der Parameter  $\bar{b}$  als Störparameter behandelt wird; sein Wert wird durch eine Kontrollmessung in einem vorher definierten Seitenband eingegrenzt. Als Teststatistik wird der Profile-Likelihood-Quotient  $q^{PL}$  genutzt.

### 4.1 Kontrollmessung mit Seitenbändern

Sind die Vorhersagen der Untergrundhypothese mit Unsicherheiten behaftet, sind Störparameter dafür verantwortlich, d. h. Modellparameter, die nicht von Interesse sind oder nicht beobachtet werden können [16]. Im Folgenden wird angenommen, dass der Störparameter durch die unbekannte Anzahl  $\bar{b}$  an erwarteten Untergrundereignissen pro Bin gegeben ist. Um die Werte eines Störparameters sinnvoll einzugrenzen, werden häufig Messungen in einer zuvor festgelegten Kontrollregion durchgeführt. Dazu wird ein Teil des untersuchten Spektrums als Kontrollregion definiert, von dem angenommen werden kann, dass ausschließlich Untergrundereignisse dort auftreten. *Abbildung 23* zeigt eine solche Kontrollregion in dem hier betrachteten Beispielspektrum: Sie entspricht dem Seitenband, das sich jeweils aus den zwei äußeren Bins auf jeder Seite des Spektrums zusammensetzt. Der Bereich dazwischen, d. h. die inneren sechs Bins, stellt die Region dar, in der nach einem möglichen Signal gesucht wird. Ohne Beschränkung der Allgemeinheit befinde sich das tatsächliche Signal wieder in Bin 5 (Signalbin). Außerdem sei die mit  $m$  bezeichnete Summe der in den Bins der Kontrollregion beobachteten Einträge Poisson-verteilt mit Erwartungswert

$$E[m] = \tau \bar{b}. \quad (43)$$

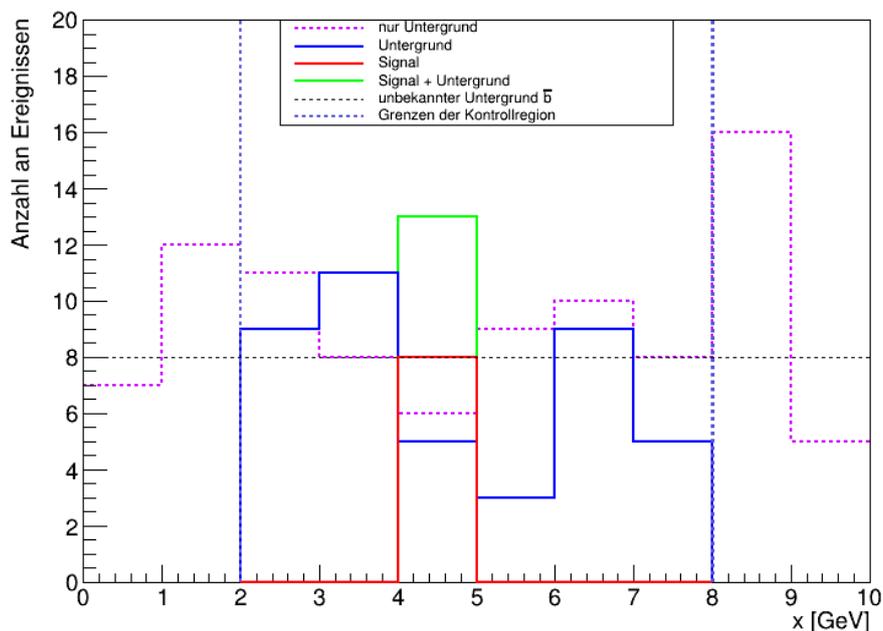


Abbildung 23: Die zwei äußeren Bins auf jeder Seite des Beispielspektrums der invarianten Masse  $x$  bilden ein Seitenband, in dem ausschließlich Untergrundereignisse auftreten. Sie werden im Rahmen einer Kontrollmessung gezählt, um daraus einen aussagekräftigen Schätzwert  $\hat{b}$  für die Untergrundereignisse im jeweils betrachteten Bin zu erhalten.

Die Größe  $\tau$  stellt das Verhältnis der Größe der Kontrollregion (hier: vier Bins) zu dem untersuchten Bereich des Massenspektrums dar: In diesem Beispiel wird – unter der Annahme voneinander unabhängiger Bins – jeweils ein Bin der inneren sechs Bins untersucht, sodass für den Skalierungsfaktor  $\tau$  für jede Messung  $\tau = \frac{4}{1} = 4$  gilt.<sup>4</sup>

Zunächst besteht die Aufgabe darin, die Teststatistik  $q^{PL}$  für die Berechnung der lokalen und globalen  $p$ -Werte sowie der oberen Ausschlussgrenzen  $s_{up}$  zu bestimmen:

$$q^{PL} = -2 \ln \frac{L(s | \hat{b})}{L(\hat{s} | \hat{b})} = -2 \ln \left( L(s | \hat{b}) \right) + 2 \ln \left( L(\hat{s} | \hat{b}) \right). \quad (44)$$

Dazu wird die Likelihood für den Parameter  $s$  der Signalstärke und die Likelihood für den unbekanntem Untergrund benötigt. Sie ist jeweils das Produkt zweier Poisson-Terme, von denen der erste die Suche nach einem Signal im untersuchten Massenbereich widerspiegelt und der zweite die Kontrollmessung im Seitenband:

$$L(s | b) = \frac{(s+b)^n}{n!} e^{-(s+b)} \frac{(\tau b)^m}{m!} e^{-(\tau b)}. \quad (45)$$

Die Variable  $n$  bezeichnet die beobachtete Anzahl an Ereignissen in dem jeweils untersuchten Bin des Massenspektrums. Aus Gleichung 45 folgt für die Teststatistik  $q^{PL}$ :

$$q^{PL} = -2 \ln \left( \frac{(s+\hat{b})^n}{n!} e^{-(s+\hat{b})} \frac{(\tau \hat{b})^m}{m!} e^{-(\tau \hat{b})} \right) + 2 \ln \left( \frac{(\hat{s}+\hat{b})^n}{n!} e^{-(\hat{s}+\hat{b})} \frac{(\tau \hat{b})^m}{m!} e^{-(\tau \hat{b})} \right). \quad (46)$$

Da der zweite Poisson-Term in Gleichung 45 nicht von der Signalstärke  $s$  abhängt, stimmt die Herleitung für den besten Schätzer  $\hat{s}$  mit Gleichung 15 überein. Nun allerdings muss auch der Parameter für den Untergrund geschätzt werden, sodass  $\hat{s} = n - \hat{b}$  gilt. Der Schätzer  $\hat{b}$  wird ebenso wie  $\hat{s}$  mithilfe einer Maximum-Likelihood-Schätzung bestimmt, indem die logarithmierte Likelihood für eine untersuchte Signalstärke  $s$  nach dem Parameter  $b$  an der Stelle  $b = \hat{b}$  abgeleitet wird und verlangt wird, dass diese Ableitung gleich null ist:

$$\begin{aligned} \left. \frac{\partial \ln L(s | b)}{\partial b} \right|_{b=\hat{b}} &= \left. \frac{\partial \ln \left( \frac{(s+b)^n}{n!} e^{-(s+b)} \frac{(\tau b)^m}{m!} e^{-(\tau b)} \right)}{\partial b} \right|_{b=\hat{b}} \\ &= \left. \frac{\partial \left( \ln \left( \frac{(s+b)^n}{n!} \right) + \ln \left( e^{-(s+b)} \right) + \ln \left( \frac{(\tau b)^m}{m!} \right) + \ln \left( e^{-(\tau b)} \right) \right)}{\partial b} \right|_{b=\hat{b}} \\ &= \left. \frac{\partial \left( n \ln(s+b) - \ln(n!) - s - b + m \ln(\tau b) - \ln(m!) - \tau b \right)}{\partial b} \right|_{b=\hat{b}} \\ &= \frac{n}{s+\hat{b}} - 1 + m \frac{\tau}{\tau \hat{b}} - \tau \stackrel{!}{=} 0 \end{aligned}$$

<sup>4</sup> Unter realistischen Bedingungen kann auch der Skalierungsfaktor  $\tau$  mit Unsicherheiten behaftet sein und muss dann selbst als Störparameter betrachtet werden [10]. Im hier vorgestellten Beispiel sei der Wert für  $\tau$  als bekannt angenommen.

$$\iff \frac{n}{s + \hat{b}} + \frac{m}{\hat{b}} = 1 + \tau. \quad (47)$$

Der Schätzer  $\hat{b}$  für das globale Maximum der Likelihood kann nun durch Einsetzen von  $\hat{s} = n - \hat{b}$  berechnet werden:

$$\frac{n}{\hat{s} + \hat{b}} + \frac{m}{\hat{b}} = 1 + \tau \iff \hat{b} = \frac{m}{\tau}. \quad (48)$$

Die maximierte Likelihood  $L(\hat{s} | \hat{b})$  ist für die Berechnung der lokalen und globalen  $p$ -Werte sowie der oberen Grenzen für die Signalstärke gleich. Allerdings unterscheidet sich der Schätzer  $\hat{b}$ , da unterschiedliche Hypothesen – d. h. unterschiedliche Werte für  $s$  – getestet werden. Im Falle der lokalen und globalen  $p$ -Werte nimmt die Untergrundhypothese die Rolle der getesteten Nullhypothese  $H_0$  ein und es gilt  $s = 0$ . Wird dieser Wert für  $s$  in Gleichung 47 eingesetzt, die für  $\hat{b}$  und  $\hat{b}$  gleichermaßen gültig ist, folgt:

$$\hat{b}(s = 0) = \frac{n + m}{1 + \tau}. \quad (49)$$

Für die Bestimmung der oberen Ausschlussgrenzen  $s_{up}$  auf das Signal wird eine Signalstärke getestet, für die  $s > 0$  gilt – die Größe  $s$  verschwindet daher in Gleichung 47 nicht. Die Auflösung nach dem Parameter  $\hat{b}$  ergibt:

$$\begin{aligned} & \frac{n}{s + \hat{b}} + \frac{m}{\hat{b}} = 1 + \tau \quad | \cdot \hat{b} \quad | \cdot (s + \hat{b}) \\ \iff & \hat{b} n + m(s + \hat{b}) = (1 + \tau) \hat{b}(s + \hat{b}) \\ \iff & \hat{b}(n + m) + m s = (1 + \tau) (\hat{b} s + \hat{b}^2) \\ \iff & \hat{b}^2 (1 + \tau) - \hat{b}(n + m - s(1 + \tau)) - m s = 0 \quad | \cdot \frac{1}{1 + \tau} \\ \iff & \hat{b}^2 - \hat{b} \left( \frac{n + m - s(1 + \tau)}{1 + \tau} \right) - \frac{m s}{1 + \tau} = 0 \quad | \text{quadratische Ergänzung} \\ \iff & \left( \hat{b} - \frac{n + m - s(1 + \tau)}{2(1 + \tau)} \right)^2 - \left( \frac{n + m - s(1 + \tau)}{2(1 + \tau)} \right)^2 - \frac{m s}{1 + \tau} = 0 \\ \implies & \hat{b}(s \geq 0) = \frac{n + m - s(1 + \tau)}{2(1 + \tau)} \pm \sqrt{\left( \frac{n + m - s(1 + \tau)}{2(1 + \tau)} \right)^2 + \frac{m s}{1 + \tau}}. \quad (50) \end{aligned}$$

Das physikalisch relevante Ergebnis entspricht der positiven Lösung dieser Gleichung. Sie stimmt für  $s = 0$  mit Gleichung 49 überein. Werden die Schätzer  $\hat{s}$ ,  $\hat{b}$  und  $\hat{b}$  in Gleichung 46 eingesetzt<sup>5</sup>,

<sup>5</sup> Für eine bessere Übersicht wurde auf das Einsetzen des Schätzers  $\hat{b}$  in die Teststatistik  $q_{s>0}^{PL}$  verzichtet.

liefert das die gesuchten Teststatistiken  $q^{PL}$ :

$$q_{s=0}^{PL} = \begin{cases} 2 \left( n \ln \left( \frac{n(1+\tau)}{n+m} \right) + m \ln \left( \frac{m(1+\tau)}{\tau(n+m)} \right) \right), & \hat{s} = n - \frac{m}{\tau} \geq 0 \\ 0, & \hat{s} = n - \frac{m}{\tau} < 0 \end{cases} \quad (51)$$

$$q_{s>0}^{PL} = \begin{cases} 2 \left( n \ln \left( \frac{n}{s+\hat{b}} \right) + m \ln \left( \frac{m}{\tau \hat{b}} \right) + \hat{b}(1+\tau) + s - n - m \right), & \hat{s} = n - \frac{m}{\tau} \leq s \\ 0, & \hat{s} = n - \frac{m}{\tau} > s \end{cases} \quad (52)$$

Die Bedingung für  $\hat{s}$  in Gleichung 51 stellt sicher, dass nur positive Signale betrachtet werden: Fluktuationen in Richtung geringerer Ereigniszahlen ( $n < \frac{m}{\tau} = \bar{b}$ ) werden bei einem Test der Untergrundhypothese nicht gegen die Untergrundhypothese verwendet. Die Bedingung für  $\hat{s}$  in Gleichung 52 stellt entsprechend sicher, dass bei der Bestimmung der Ausschlussgrenzen  $s_{up}$  Fluktuationen in Richtung höherer Ereigniszahlen ( $n > s + \frac{m}{\tau} = s + \bar{b}$ ) nicht als Beleg gegen die getestete Signalthypothese gewertet werden. Die Signalstärke  $s$  im Nenner des Logarithmus in Gleichung 52 führt dazu, dass die Werte  $q_{s>0}^{PL}$  im Mittel kleiner sind als die Werte  $q_{s=0}^{PL}$ . Die Hypothese  $H_0$  liegt daher links von der Hypothese  $H_1$ . *Abbildung 24* zeigt ein Beispiel für die  $H_0$ - und  $H_1$ -Verteilung der normierten Teststatistik  $q_{s=0}^{PL}$ , *Abbildung 25* zeigt sie für die Teststatistik  $q_{s>0}^{PL}$ .

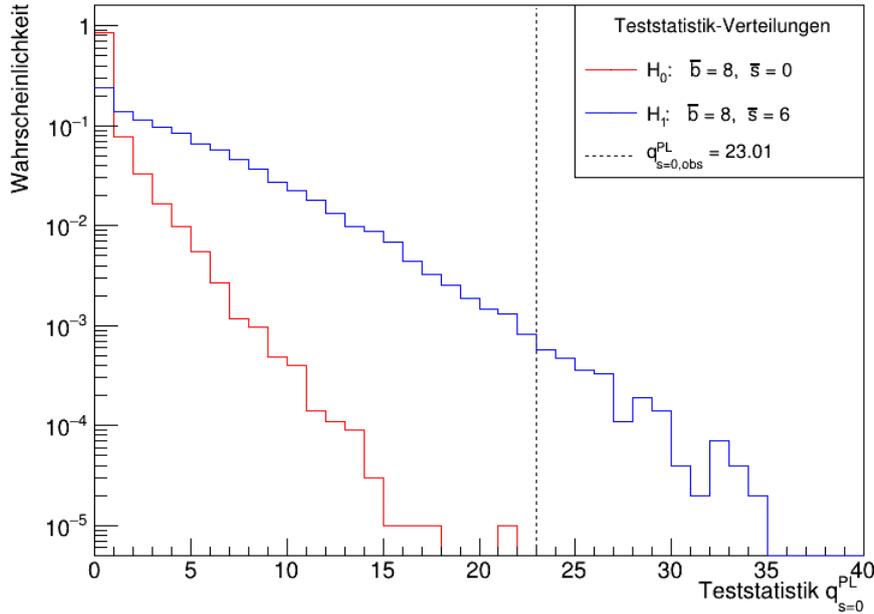


Abbildung 24: Auf 1 normierte Verteilungen der Teststatistik  $q_{s=0}^{PL}$  für den Fall, dass acht Untergrundereignisse erwartet werden, wenn die getestete Untergrundhypothese wahr ist (rot, linke Verteilung) und zusätzlich sechs Signalereignisse erwartet werden, wenn die Signalthypothese wahr ist (blau, rechte Verteilung). Die gestrichelte Linie markiert den Wert  $q_{s=0,obs}^{PL}$ , der zehn beobachteten Ereignissen entspricht.

## 4.2 Simulation von Pseudoexperimenten

Für das Beispielexperiment mit unbekanntem Untergrund  $\bar{b}$  werden die sechs inneren Bins des Massenspektrums (d. h. Bin 3 bis Bin 8, vgl. *Abbildung 23*) zunächst auf ihre lokalen  $p$ -Werte

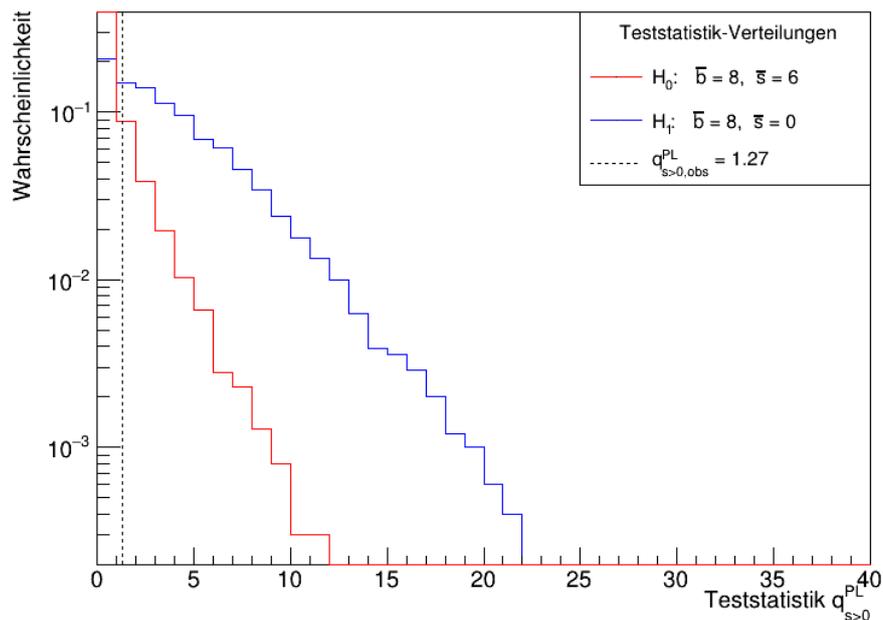


Abbildung 25: Auf 1 normierte Verteilungen der Teststatistik  $q_{s>0}^{PL}$  für den Fall, dass acht Untergrundereignisse erwartet werden, wenn die Untergrundhypothese wahr ist (blau, rechte Verteilung) und zusätzlich sechs Signaleereignisse erwartet werden, wenn die getestete Signalthypothese wahr ist (rot, linke Verteilung). Die gestrichelte Linie markiert den Wert  $q_{s>0,obs}^{PL}$ , der zehn beobachteten Ereignissen entspricht.

untersucht. Das Signal befinde sich unverändert in Bin 5, seine Position im Massenspektrum sei jedoch nicht bekannt. Ein einzelnes Pseudoexperiment wird wie folgt simuliert:

- die Anzahl an Untergrundereignissen im gesamten Simulationsbereich (Bin 1 bis Bin 10,  $n_{bins} = 10$ ) wird für acht erwartete Untergrundereignisse ( $\bar{b} = 8$ ) aus einer Poisson-Verteilung mit Erwartungswert  $\lambda = n_{bins} \cdot \bar{b}$  gezogen und den Untergrundereignissen gemäß der Gleichverteilung in Gleichung 22 zufällig ein Wert des Massenspektrums zugeordnet – das erzeugt die Untergrundverteilung für ein Zufallsexperiment
- in der Kontrollregion (Bin 1, Bin 2, Bin 9 und Bin 10) werden die Untergrundereignisse gezählt – die Summe liefert den Wert für die Variable  $m$  für ein Zufallsexperiment
- die Untergrundereignisse werden in jedem Bin des untersuchten Massenintervalls (innere sechs Bins) gezählt – das liefert für jedes Bin  $i$  einen Wert  $n_i$  für ein Zufallsexperiment
- in jedem der inneren sechs Bins wird mithilfe von Gleichung 51 und dem Skalierungsfaktor  $\tau = 4$  (vgl. Abschnitt 4.1) ein Wert  $q_{s=0}^{PL}$  der Teststatistik bestimmt.

Dieses Vorgehen wird wiederholt und auf diese Weise eine Teststatistik-Verteilung für jedes Bin der untersuchten Region erzeugt.

Die in einem realen Zählexperiment gemessenen Daten sollen wie bisher durch einen beobachteten Wert der Teststatistik repräsentiert werden, der ebenfalls aus der Simulation gewonnen wird. Für den im Bin  $i$  beobachteten Wert  $q_{s=0,obs,i}^{PL}$  eines Pseudoexperiments werden folgende Annahmen gemacht:

- für jedes Bin der untersuchten Region, das nicht dem Signalbin entspricht, lässt sich die Anzahl an beobachteten Ereignissen  $n_{obs,i}$  aus einer Poisson-Verteilung mit Erwartungswert  $\lambda = \bar{b}$  ziehen

- für das Signalbin lässt sich die Anzahl an beobachteten Ereignissen  $n_{obs,i=5}$  aus einer Poisson-Verteilung mit Erwartungswert  $\lambda = n_0$  ziehen, wobei ohne Beschränkung der Allgemeinheit  $n_0 = 12$  gelte
- für die Anzahl an Untergründereignissen in der Kontrollregion gelte  $m = \tau \bar{b}$
- die Einschränkung, dass die Anzahl an Ereignissen in jedem Bin größer oder gleich  $\frac{m}{\tau}$  sein muss, gelte nicht für die Werte  $n_{obs,i}$ ; dadurch soll die Menge an Werten, die von  $q_{obs,i}^{PL}$  angenommen werden kann, vergrößert werden, sodass die lokalen  $p$ -Werte besser aufgelöst werden können.

Die unter diesen Annahmen ebenfalls mit Gleichung 51 bestimmten Werte  $q_{obs,i}^{PL}$  erlauben nun die Berechnung der lokalen  $p$ -Werte in jedem Bin des untersuchten Massenintervalls. Der globale  $p$ -Wert wird analog zu dem in Abschnitt 2.4.2 vorgestellten Vorgehen bestimmt, indem die Anzahl  $n_i$  aus dem Bin  $i$  mit der höchsten Anzahl von Ereignissen gewählt wird.

Durch das Wiederholen des bis hierhin beschriebenen Vorgehens soll ein reales Experiment simuliert werden, das mehrfach durchgeführt wird, wobei jedes Mal die Anzahl an beobachteten Ereignissen  $q_{obs,i}^{PL}$  in jedem Bin zufällig bestimmt wird. Das wiederholte Durchführen führt in jedem Bin zu einer Verteilung der lokalen  $p$ -Werte und insgesamt zu genau einer Verteilung der globalen  $p$ -Werte.

Zur Berechnung der oberen Ausschlussgrenzen  $s_{up}$  wird die direkte Methode verwendet (vgl. Abschnitt 3.3.1). Die Simulation gliedert sich in folgende Schritte:

- in jedem Pseudoexperiment werden dieselben beobachteten Werte  $n_{obs,i}$  für das jeweils untersuchte Bin genutzt wie für die Berechnung der lokalen und globalen  $p$ -Werte – sie werden mithilfe von Gleichung 52 in die äquivalenten Werte  $q_{obs,i,s>0}^{PL}$  transformiert, wobei auch hier auf die entsprechende Einschränkung für die Werte  $n_{obs,i}$  verzichtet wird ( $n_{obs,i} > s + \frac{m}{\tau}$ ); für die Anzahl an Untergründereignissen in der Kontrollregion wird wieder  $m = \tau \bar{b}$  angenommen und für den Parameter  $s$  wird die jeweils getestete Signalstärke  $s = s_{up}$  verwendet
- analog zu dem am Anfang dieses Abschnitts beschriebenen Vorgehen wird eine Untergrundverteilung im gesamten Simulationsbereich generiert
- für die jeweils getestete Signalthypothese wird eine Anzahl an Signalereignissen aus einer Poisson-Verteilung mit Erwartungswert  $\lambda = s_{up}$  gezogen und zusätzlich in das Signalbin gefüllt
- Bin für Bin wird nun die Anzahl an Ereignissen gezählt, ein Wert  $q_{s>0}^{PL}$  bestimmt und durch mehrfaches Wiederholen eine Teststatistik-Verteilung generiert – das wird so lange unter Variation des Parameters  $s_{up}$  wiederholt, bis die gesuchte Ausschlussgrenze für ein Konfidenzniveau von  $CL = 95\%$  gefunden ist.

### 4.3 Ergebnisse

Die in einem Experiment beobachteten oberen Ausschlussgrenzen auf eine Signalstärke werden häufig zusammen mit den erwarteten Ausschlussgrenzen angegeben [8]. Ein reales Experiment wird hier durch ein beispielhaft ausgewähltes Pseudoexperiment vertreten und die erwarteten Werte werden jeweils durch den Median der in einem Bin generierten Verteilung der  $s_{up}$ -Werte repräsentiert. Zusätzlich werden die Umgebungen um den Erwartungswert angegeben, in der

68.3% ( $1\sigma$ -Umgebung) bzw. 95.4% ( $2\sigma$ -Umgebung) der  $s_{up}$ -Werte liegen. Die Mediane der lokalen  $p$ -Werte in jedem Bin werden für dieses Beispiel auf die gleiche Weise behandelt.

Die beobachteten Ereigniszahlen eines frei ausgewählten Pseudoexperiments sind in *Abbildung 26* dargestellt. *Abbildung 27* und *Abbildung 28* zeigen die mit ihnen erhaltenen lokalen  $p$ -Werte und oberen Ausschlussgrenzen  $s_{up}$  (durchgezogene Linie) sowie die entsprechenden Mediane. Für die Teststatistik-Verteilungen in jedem Bin wurden 1000 Pseudoexperimente durchgeführt, von denen jedes einen Wert der Teststatistik und einen globalen  $p$ -Wert geliefert hat. Insgesamt wurden 1000 Teststatistik-Verteilungen pro Bin generiert und dabei jeweils ein lokaler  $p$ -Wert und ein  $s_{up}$ -Wert erhalten<sup>6</sup>.

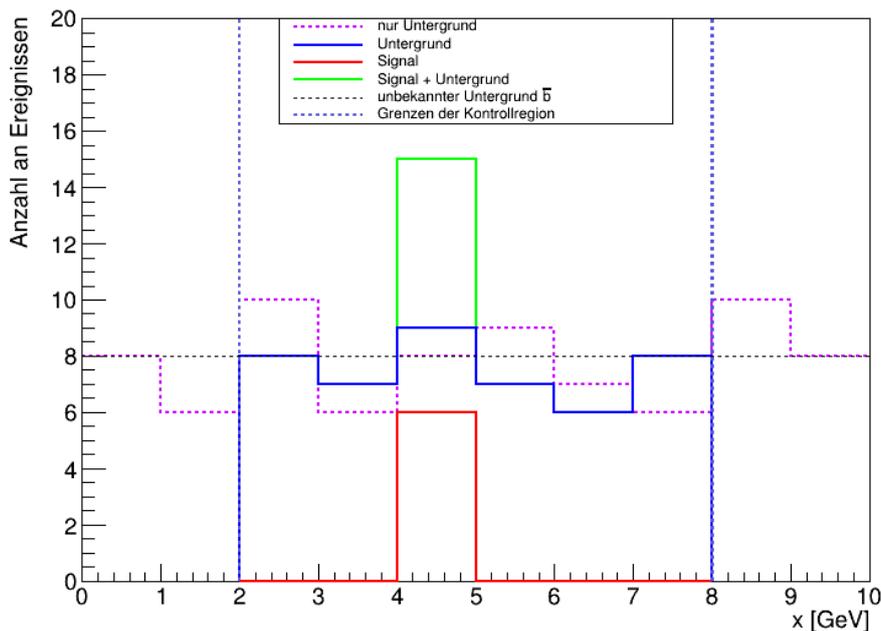


Abbildung 26: *Beispielspektrum eines frei ausgewählten Pseudoexperiments. Die beobachteten Ereigniszahlen haben zu den lokalen  $p$ -Werten und oberen Grenzen  $s_{up}$  geführt, die in den Abbildungen 27 und 28 durch die durchgezogene Linie veranschaulicht werden.*

Trotz der Annahme, dass die Position des Signals bei Durchführung der Experimente nicht bekannt ist, deuten die erwarteten lokalen  $p$ -Werte und  $s_{up}$ -Werte darauf hin, dass in Bin 5 eine höhere Signifikanz vorliegt. Dafür spricht zum einen, dass der Median der lokalen  $p$ -Werte hier geringer ist. Gleichzeitig spricht eine höhere obere Ausschlussgrenze  $s_{up}$  dafür, dass in Bin 5 mehr Signalereignisse aufgetreten sind als an anderer Stelle des untersuchten Beispielspektrums. Es muss allerdings beachtet werden, dass der lokale  $p$ -Wert ( $p_{\text{lokal}} \approx 0.13$ ) unter Berücksichtigung des Look-Elsewhere-Effekts etwa um den Faktor 4.85 erhöht wird ( $p_{\text{global}} \approx 0.63$ ). Die erwarteten Ausschlussgrenzen lassen außerdem einen möglichen Einfluss des Störparameters  $\bar{b}$  vermuten: Die Grenzen  $s_{up}$  sind im Vergleich zu den Grenzen, die mit bekanntem Untergrund bestimmt wurden, in Richtung höherer Werte verschoben (vgl. *Abbildung 14* und *Abbildung 17*). Das erscheint plausibel, da bei höheren Unsicherheiten (und gleichem Konfidenzniveau) weniger strenge Grenzen auf die Signalstärke gesetzt werden sollten.

Abschließend sei darauf hingewiesen, dass obere Ausschlussgrenzen jeweils nur für die betrachtete Position in einem Massenspektrum die Aussage treffen, dass bestimmte Signalstärken mit

<sup>6</sup> Für die Werte  $s_{up}$  wurde eine Intervallschachtelung genutzt, sodass die tatsächliche Anzahl an generierten Teststatistik-Verteilungen für die Bestimmung der oberen Ausschlussgrenzen höher ist.

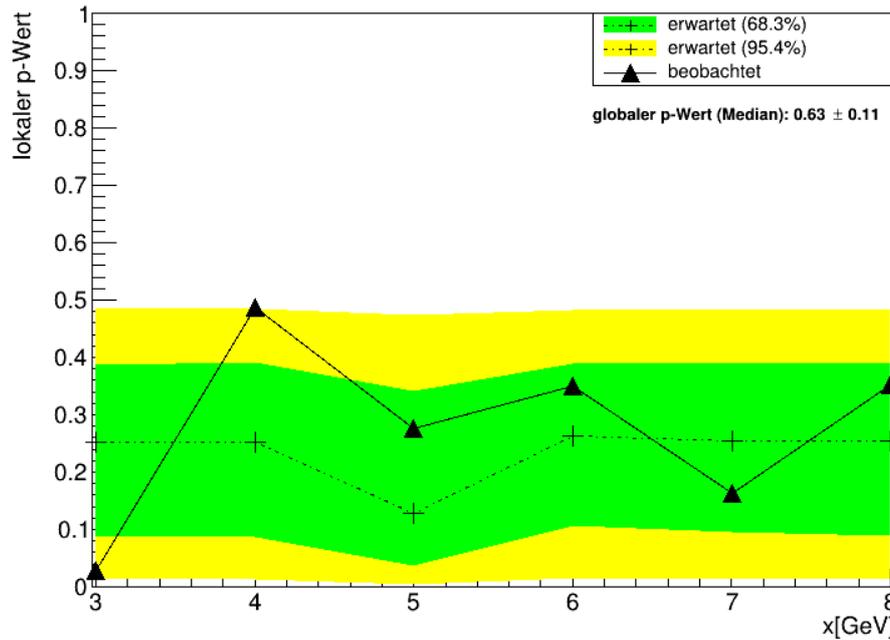


Abbildung 27: Die erwarteten lokalen  $p$ -Werte in der untersuchten Region (Bin 3 bis Bin 8) des Beispielspektrums, ausgedrückt durch ihre Mediane (+) und umgeben von ihrer  $1\sigma$ -Umgebung (grünes, inneres Band) und ihrer  $2\sigma$ -Umgebung (gelbes, äußeres Band). Die in einem einzelnen Pseudoexperiment beobachteten lokalen  $p$ -Werte werden durch die durchgezogene Linie dargestellt. Die Beobachtung weicht in diesem Fall deutlich von der Erwartung ab.

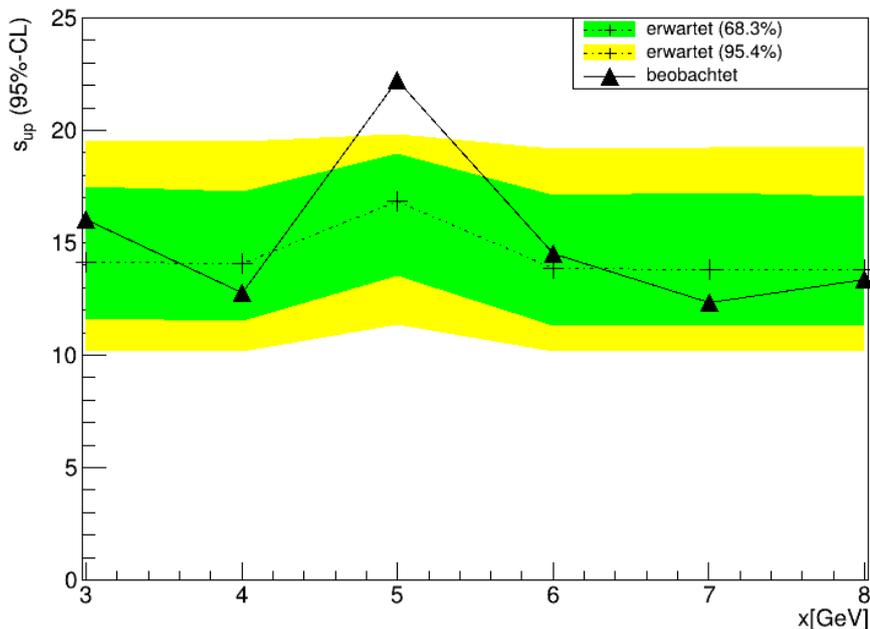


Abbildung 28: Die erwarteten  $s_{up}$ -Werte in der untersuchten Region (Bin 3 bis Bin 8) des Beispielspektrums, ausgedrückt durch ihre Mediane (+) und umgeben von ihrer  $1\sigma$ -Umgebung (grünes, inneres Band) und ihrer  $2\sigma$ -Umgebung (gelbes, äußeres Band). Die in einem einzelnen Pseudoexperiment beobachteten  $s_{up}$ -Werte werden durch die durchgezogene Linie dargestellt. Die Beobachtung einer höheren Ausschlussgrenze in Bin 5 stimmt mit der Erwartung überein – hier wurden die meisten Ereignisse gezählt.

einer bestimmten Wahrscheinlichkeit ausgeschlossen werden können. Für jede Position des Massenspektrums muss dies separat bestimmt werden, während eine Fluktuation, die als scheinbares Signal fehlinterpretiert werden kann, überall im Massenspektrum auftreten kann. Aus diesem Grund tritt der Look-Elsewhere-Effekt nur bei der Bestimmung der lokalen bzw. globalen  $p$ -Werte auf, nicht aber bei der Bestimmung der oberen Ausschlussgrenzen auf einen Modellparameter [17].

## 5 Zusammenfassung und Ausblick

Eines der wesentlichen Ziele der Hochenergiephysik ist es, festzustellen, ob die in Teilchenkollisionen gesammelten Messdaten signifikant von den Vorhersagen des Standardmodells (SM) abweichen. Mithilfe von Teststatistiken und  $p$ -Werten wird bestimmt, wie groß die Wahrscheinlichkeit eines gleich oder weniger wahrscheinlichen Messergebnisses ist: Ist sie  $2,87 \cdot 10^{-7}$  oder geringer, wird den beobachteten Daten eine Signifikanz von mindestens 5 Sigma zugeschrieben, ab der es in der Teilchenphysik gerechtfertigt ist, von einer Entdeckung eines neuen physikalischen Prozesses zu sprechen.

Die Untersuchung des Look-Elsewhere-Effekts mithilfe simulierter Zählexperimente hat gezeigt, dass Fluktuationen, die irrtümlich als Signal interpretiert werden, den Median einer Teststatistik-Verteilung erhöhen und so zu einem höheren  $p$ -Wert führen. Indem dieser Effekt berücksichtigt wird, kann das Risiko reduziert werden, einem Messergebnis fälschlicherweise den Status einer Entdeckung zu verleihen.

Ist der beobachtete  $p$ -Wert größer als  $2,87 \cdot 10^{-7}$ , kann nicht von einer Entdeckung gesprochen werden. Stattdessen können Grenzen auf Modellparameter gesetzt werden: In diesem Fall wird eine Signalthypothese getestet, während die Untergrundhypothese die Funktion der relevanten Alternativhypothese übernimmt. Dargestellt wurde dies in der vorliegenden Arbeit am Beispiel der oberen Ausschlussgrenzen auf die Stärke eines möglichen Signals. Die Berechnung der Ausschlussgrenzen hat gezeigt, dass die direkte Methode, die aus der Definition eines Konfidenzintervalls folgt, zu negativen und damit unphysikalischen Grenzwerten führen kann, wenn die Anzahl an beobachteten Ereignissen in einem Zählexperiment kleiner ist als die Anzahl an erwarteten Untergrundereignissen. Andere Methoden wie die CLs-Methode oder der Bayesianische Ansatz beheben diesen Nachteil, indem sie einen zweiten  $p$ -Wert berücksichtigen (CLs-Methode) oder eine subjektive a-priori-Wahrscheinlichkeit mit einbeziehen (Bayesianischer Ansatz). Beide Möglichkeiten haben in der durchgeführten Analyse zu identischen Ausschlussgrenzen geführt. Das stimmt mit der Erfahrung überein, dass in vielen praktischen Anwendungen, in denen beide Methoden unter der Annahme einer konstanten a-priori-Wahrscheinlichkeit angewendet werden, ähnliche Ausschlussgrenzen auf die Signalstärke beobachtet werden [8].

Auch Störparameter können in der statistischen Analyse der Messwerte berücksichtigt werden: Für eine unbekannte Anzahl an zu erwartenden Untergrundereignissen wurden in dieser Arbeit Kontrollmessungen in einer Region des untersuchten Massenspektrums simuliert, in der keine Signalereignisse zu erwarten sind. Mithilfe dieser Kontrollmessungen konnte der Störparameter geschätzt werden. Unter Verwendung des Profile-Likelihood-Quotienten wurden die Teststatistiken bestimmt, die für die Berechnung von  $p$ -Werten und das Setzen der oberen Ausschlussgrenzen bei Anwesenheit eines Störparameters genutzt werden können – sie sind nicht identisch, da unterschiedliche Hypothesen getestet werden. Die ermittelten Ausschlussgrenzen sind um etwa zwei Einheiten der Signalstärke höher als die, die bei bekanntem Untergrund ermittelt wurden. Das legt nahe, dass die Unsicherheiten eines Modellparameters die Konfidenzintervalle für ein gegebenes Konfidenzniveau vergrößern, da ein kleinerer Bereich für die Signalstärke ausgeschlossen werden kann. Es ist davon auszugehen, dass sich die oberen Grenzen auf die Signalstärke mit der Stärke des jeweils anwesenden Störparameters ändern.

Weitergehende Untersuchungen könnten die in dieser Arbeit genutzten Idealisierungen nach und nach durch realistischere Annahmen ersetzen. So könnte zum Beispiel eine Gauß-förmige Signalverteilung über einem exponentiell verteilten Untergrund betrachtet und Korrelationen zwischen den Bins berücksichtigt werden. Darüber hinaus könnte der Look-Elsewhere-Effekt in Abhängigkeit der Größe des untersuchten Parameterraums analysiert werden. Das könnte gerade

für zukünftige Experimente der Hochenergiephysik wichtig sein, denn sollte in immer größeren Massenbereichen nach neuen Prozessen gesucht werden, läge die Wahrscheinlichkeit beispielsweise bei 10000 untersuchten Bins bei etwa  $3 \cdot 10^{-3}$ , eine Fluktuation zu beobachten, deren Signifikanz die  $5\sigma$ -Grenze erreicht. Aus frequentistischer Sicht wäre dann eine  $6\sigma$ -Grenze für eine Entdeckung erforderlich, um das Vertrauensniveau der heutigen Experimente beizubehalten [18]. Interessant wäre daher auch die Untersuchung, inwieweit ein bayesianischer Beitrag, der z. B. das theoretische Vorwissen zu den jeweils in einem Experiment gesuchten neuen Prozessen nutzt, bei der Bestimmung der Signifikanz berücksichtigt werden könnte.

## Abkürzungsverzeichnis

<b>ATLAS</b>	A Toroidal LHC ApparatuS
<b>CERN</b>	Conseil Européen pour la Recherche Nucléaire
<b>CMS</b>	Compact Muon Solenoid
<b>LEE</b>	Look-Elsewhere-Effekt
<b>LHC</b>	Large Hadron Collider
<b>NP</b>	Neyman-Pearson
<b>PL</b>	Profile-Likelihood
<b>SM</b>	Standardmodell der Teilchenphysik
<b>WIMP</b>	Weakly interacting massive particles („schwach wechselwirkende massereiche Teilchen“)

## Literatur

- [1] vgl. Kyle Cranmer. Practical Statistics for the LHC. In *2011 European School of High-Energy Physics*, 267–308, 2014. URL: <https://doi.org/10.5170/CERN-2014-003.267>.
- [2] Stephane Fartoukh et al. LHC Configuration and Operational Scenario for Run 3. *CERN-ACC-2021-0007*, 2021. URL: <https://cds.cern.ch/record/2790409/files/CERN-ACC-2021-0007.pdf> (zuletzt besucht am 14.06.2023).
- [3] Operations group and Stephane Fartoukh. LHC Report: record luminosity, well done LHC. 2017. <https://home.cern/news/news/accelerators/lhc-report-record-luminosity-well-done-lhc> (zuletzt besucht am 14.06.2023).
- [4] E. Gross. Practical Statistics for High Energy Physics. *CERN Yellow Rep. School Proc.*, **3**, 199–221, 2018. URL: <https://doi.org/10.23730/CYRSP-2018-003.199>.
- [5] Center for Astrophysics | Harvard & Smithsonian. Mapping Dark Matter. 2017. URL: <https://www.cfa.harvard.edu/news/mapping-dark-matter> (zuletzt besucht am 14.06.2023).
- [6] H. Baer, V. Barger, D. Sengupta, S. Salam, and K. Sinha. Status of weak scale supersymmetry after LHC Run 2 and ton-scale noble liquid WIMP searches. *Eur. Phys. J. Spec. Top.*, **229**, 2020. URL: <https://doi.org/10.1140/epjst/e2020-000020-x>.
- [7] G. Cowan. Statistics for Searches at the LHC. In *69th Scottish Universities Summer School in Physics: LHC Physics*, 321–355, 2013. URL: [https://doi.org/10.1007/978-3-319-05362-2\\_9](https://doi.org/10.1007/978-3-319-05362-2_9).
- [8] L. Lista. *Statistical Methods for Data Analysis in Particle Physics*. 2. Auflage, Springer, 2017.
- [9] J. Neyman and E. Pearson. On the Problem of the Most Efficient Tests of Statistical Hypotheses, Containing Papers of a Mathematical or Physical Character. *Philos. Trans. R. Soc. A*, 289–337, 1933. URL: <https://doi.org/10.1098/rsta.1933.0009>.
- [10] Siehe zum Beispiel G. Cowan, K. Cranmer, E. Gross et al. Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J. C*, **71**, 1554, 2011. URL: <https://doi.org/10.1140/epjc/s10052-011-1554-0>.
- [11] E. Gross and O. Vitells. Trial factors for the look elsewhere effect in high energy physics. *Eur. Phys. J. C*, **70**, 525–530, 2010. URL: <https://doi.org/10.1140/epjc/s10052-010-1470-8>.
- [12] A. E. Bayer, U. Seljak, and J. Robnik. Self-calibrating the look-elsewhere effect: fast evaluation of the statistical significance using peak heights. *Mon. Not. R. Astron. Soc.*, **508**, 1346–1357, 2021. URL: <https://doi.org/10.1093/mnras/stab2331>.
- [13] C. Rover, C. Messenger, and R. Prix. Bayesian versus frequentist upper limits. In *PHYSTAT 2011*, 158–163, Geneva, 2011. CERN. URL: <https://doi.org/10.5170/CERN-2011-006.158>.
- [14] A. L. Read. Presentation of Search Results: the CLs technique. *J. Phys. G*, **28**, 2693, 2002. URL: <https://doi.org/10.1088/0954-3899/28/10/313>.
- [15] M. Krystek. *Berechnung der Messunsicherheit. Grundlagen und Anleitung für die praktische Umsetzung*. 3. Auflage, Springer, 2020.

- [16] K. Cranmer. Frequentist hypothesis testing with background uncertainty. *eConf*, C030908:WEMT004, 2003. URL: <https://doi.org/10.48550/arXiv.physics/0310108>.
- [17] L. Lyons. Comments on ‘Look Elsewhere Effect’. *University of Oxford*, 2010. URL: [https://users.physics.ox.ac.uk/~lyons/LEE\\_feb7\\_2010.pdf](https://users.physics.ox.ac.uk/~lyons/LEE_feb7_2010.pdf).
- [18] R. Dawid. Higgs Discovery and the Look Elsewhere Effect. *Philosophy of Science*, 82:76–96, (2015). URL: <https://doi.org/10.1086/679179>.

# Dank

Bedanken möchte ich mich zunächst bei Prof. Dr. Markus Cristinziani für die Möglichkeit, meine Bachelorarbeit in der Arbeitsgruppe Experimentelle Teilchenphysik der Universität Siegen schreiben zu dürfen.

Ein besonderer Dank gilt Dr. Wolfgang Walkowiak, der mich mit viel Zeit und Begeisterung an das Thema dieser Arbeit herangeführt hat. Von ihm habe ich viel über das wissenschaftliche Programmieren und die Arbeitsweise in der experimentellen Teilchenphysik gelernt.

Am meisten bedanken möchte ich mich bei denen, die mir besonders am Herzen liegen. Sie zeigen mir immer wieder, dass es keine Formel gibt, die das Wichtigste im Leben beschreiben kann...

## Eidesstattliche Erklärung

Hiermit erkläre ich, dass ich die vorliegende Bachelorarbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie Zitate und Ergebnisse Anderer kenntlich gemacht habe.

---

Ort, Datum

---

Unterschrift