

# Auswirkungen des Look-Elsewhere-Effekts auf die Interpretation von simulierten Messergebnissen des ATLAS-Experiments

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

der Universität Siegen



Department Physik

vorgelegt von  
**Fabian Dünkel**

September 2016



# Inhaltsverzeichnis

<b>1</b>	<b>Motivation</b>	<b>2</b>
<b>2</b>	<b>LHC und ATLAS-Detektor</b>	<b>4</b>
2.1	Der LHC . . . . .	4
2.2	Das ATLAS-Experiment . . . . .	6
<b>3</b>	<b>Statistische Methoden</b>	<b>9</b>
3.1	Signifikanzbestimmung eines Zähl-experiments . . . . .	9
3.2	Konfidenzintervalle eines Zähl-experiments . . . . .	11
3.2.1	Einfache Methode . . . . .	11
3.2.2	Likelihood-Quotienten-Test mit Teststatistik . . . . .	12
3.2.3	$CL_s$ -Methode . . . . .	13
3.2.4	Importance Sampling-Methode . . . . .	17
<b>4</b>	<b>Der Look-Elsewhere-Effekt</b>	<b>20</b>
4.1	Allgemeine Beschreibung . . . . .	20
4.2	Das Diphoton-Massenspektrum als Beispiel . . . . .	20
4.3	Untersuchung des LEE am Beispiel einer Massenverteilung . . . . .	21
4.3.1	Lokaler p-Wert eines möglichen Signals einer Massenverteilung . . . . .	23
4.3.2	Globaler p-Wert eines möglichen Signals einer Massenverteilung . . . . .	26
4.3.3	Vergleich des globalen und lokalen p-Werts und Interpretation . . . . .	26
<b>5</b>	<b>Zusammenfassung und Ausblick</b>	<b>28</b>
	<b>Liste der verwendeten Begriffe</b>	<b>i</b>
	<b>Literaturverzeichnis</b>	<b>ii</b>
	<b>Danksagung</b>	<b>iv</b>
	<b>Erklärung</b>	<b>v</b>

# 1 Motivation

Das Standardmodell der Teilchenphysik wurde entwickelt, um den Aufbau der Materie zu beschreiben. Es beinhaltet Materieteilchen und die Wechselwirkungen zwischen ihnen, die über Wechselwirkungsteilchen vermittelt werden. Das Standardmodell kann jedoch nicht alle beobachteten Phänomene beschreiben. So sind die Materie-Antimaterie-Asymmetrie des Universums, die Existenz von Dunkler Materie oder die Gravitation nicht mit dem Standardmodell erklärbar. Aus diesem Grund gibt es verschiedene vorgeschlagene Erweiterungen des Standardmodells, die diese Phänomene erklären sollen. Diese Änderungen des Modells haben auch Auswirkungen auf die Vorhersage der Zerfallsverhalten von Teilchen. Besonders seltene Zerfälle, die im Standardmodell weitestgehend unterdrückt sind, durch Erweiterungen jedoch mit einer höheren Wahrscheinlichkeit vorkommen, sind hier von Interesse [1].

Mit Detektorexperimenten wie dem ATLAS-Experiment am Large Hadron Collider (LHC) am CERN in Genf wird die Häufigkeit dieser seltenen Zerfälle untersucht, um aus dem Vergleich von Messwerten mit den verschiedenen Theorien Rückschlüsse auf die Richtigkeit dieser Theorien ziehen zu können. Eine häufig genutzte Größe in der Datenanalyse ist die invariante Masse des Mutterteilchens eines Zerfalls. Dazu wird aus den Zerfallsprodukten eines (vermuteten) Zerfalls die invariante Masse des Ausgangszustandes bestimmt. Das Standardmodell sowie dessen Erweiterungen geben hier Vorhersagen, welche Form die Verteilung der invarianten Masse hat. Zu berücksichtigende Faktoren sind nicht nur die physikalischen Grundlagen, sondern auch die Auflösung und Akzeptanz des genutzten Detektors und Untergründereignisse aus anderen Zerfallskanälen mit denselben Teilchen als Endzustand. Besonders bei seltenen Zerfällen treten viele solcher Untergründereignisse auf, die typischerweise weit häufiger sind als Ereignisse aus dem zu untersuchenden Zerfall (Signalereignisse). Für sehr geringe Anteile an Signalereignissen ist es von großer Wichtigkeit, abzuschätzen, ob wirklich Signalereignisse registriert wurden, oder nur eine statistische Fluktuation von Untergründereignissen in den Daten vorliegt. Bei der Analyse der Daten werden hier häufig Signifikanztests genutzt, um festzustellen, ob Signalereignisse registriert wurden. Um bei kleiner verfügbarer Statistik Grenzen auf Parameter, beispielsweise die Signalstärke, zu setzen, können einseitige Konfidenzintervalle verwendet werden.

Der Look-Elsewhere-Effekt (LEE) ist zu berücksichtigen, wenn z. B. bei der Untersuchung eines Massenspektrums eine Abweichung von der Untergrundverteilung gefunden wurde, jedoch vorher nicht bekannt ist, bei welchem Massewert diese zu erwarten ist. So wurde im Diphotonenspektrum eine Abweichung festgestellt, die u. a. aus einer Erweiterung des Standardmodells mit weiteren Higgs-Teilchen stammen könnte [2]. Die Signifikanz des möglichen Signals war jedoch in den Daten von 2015 nicht hoch genug für eine Entdeckung. Eine neue Analyse vom August 2016 ergab, dass es sich bei der Abweichung nur um eine statistische Fluktuation handelte [3]. Die Unkenntnis der Stelle der Abweichung erhöht die Wahrscheinlichkeit, eine statistische Fluktua-

tion im Massenspektrum fälschlicherweise als Signal zu interpretieren (Fehler 1. Art/ $\alpha$ -Fehler). Die Vernachlässigung des LEE kann somit dazu führen, dass die Signifikanz einer Abweichung überbewertet wird [4]. Dies wiederum kann zu voreiligen Schlüssen und möglicherweise fehlerinterpretierten Forschungsergebnissen führen.

Ziel dieser Arbeit ist es, den Einfluss des LEE auf die Signifikanz einer Messung zu untersuchen. Hierzu werden Monte-Carlo (MC)-Simulationen genutzt, mit welchen der Einfluss der (Un-)Kenntnis der Position des Signals im Spektrum der invarianten Masse auf die Aussagekraft einer Analyse abgeschätzt werden soll.

## 2 LHC und ATLAS-Detektor

Der LHC ist ein Teilchenbeschleuniger am Europäischen Forschungszentrum für Teilchenphysik (CERN) [5]. Er befindet sich in einem 27 km langen kreisförmigen Tunnel, der in 80 bis 100 Meter Tiefe unter der Grenze zwischen Frankreich und der Schweiz liegt. In diesem werden in zwei getrennten Strahlröhren Protonenpakete auf Schwerpunktsenergien von zur Zeit maximal 13 TeV beschleunigt. Die Protonenstrahlen durchdringen sich an vier verschiedenen Stellen entlang des Rings, an denen sich die Detektoren ATLAS, CMS, ALICE und LHCb sowie mehrere kleine Experimente befinden.

### 2.1 Der LHC

Bevor die Protonenpakete in den LHC geführt werden, durchlaufen sie eine Reihe von Vorbeschleunigern, in denen sie zunehmend an Energie gewinnen. Durch die Nutzung mehrerer hintereinander geschalteter Beschleuniger können so höhere Energien erreicht werden. Die Protonenquelle ist eine einfache Wasserstoffgasflasche. Die Wasserstoffatome werden mithilfe eines elektrischen Feldes ionisiert und dann mit Linac 2, dem ersten Beschleuniger, auf eine Strahlenergie von 50 MeV beschleunigt. Der Strahl wird dann im Proton Synchrotron Booster (PSB) auf 1.4 GeV beschleunigt, gefolgt vom Proton Synchrotron (PS) mit 25 GeV. Im Super Proton Synchrotron (SPS) werden die Protonenpakete auf 450 GeV beschleunigt, bevor sie schließlich in die beiden gegenläufigen Strahlröhren des LHC geführt werden [6]. In diesem sollen die Protonen auf bis zu 7 TeV beschleunigt werden, was Proton-Proton-Kollisionen mit einer Schwerpunktsenergie von bis zu  $\sqrt{s} = 14$  TeV ermöglicht. Nach einer zweijährigen Betriebspause, in der der LHC aufgerüstet wurde, sind Anfang 2015 Schwerpunktsenergien von 13 TeV erreicht worden. Um die Protonen bei Energien dieser Größe auf der Kreisbahn zu halten, werden 1232 supraleitende Dipolmagnete mit einer Länge von 15 Metern und Feldstärken von bis zu 8,36 T benötigt. Die Supraleitung der Magnete wird erreicht, indem die Magnete mit supraflüssigem Helium auf  $-271.3^\circ$  C gekühlt werden. Der Strahl wird durch 392 Quadrupolmagnete fokussiert. Bei maximaler Füllung besteht jeder der beiden Strahlen des LHC aus 2808 Paketen mit im Mittel  $1,1 \cdot 10^{11}$  Protonen im Abstand von 25 ns. In den Mittelpunkten der vier großen Experimente durchdringen die fokussierten Protonenstrahlen einander und es kommt zu Kollisionen der Protonen. Am Kollisionsspunkt sind die Protonenpakete komprimiert, was bei einer maximalen Luminosität von  $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$  zu 600 Millionen Kollisionen pro Sekunde führt. Dies entspricht im Mittel 20 Proton-Proton-Kollisionen bei jedem Aufeinandertreffen zweier Teilchenpakete mit einer Frequenz von 30 MHz [7].

Die vier großen Experimente am LHC (Abb. 2.1) sind:

**ALICE - A Large Ion Collider Experiment:** ALICE [8] ist ein Experiment zur Untersuchung der Eigenschaften des Quark-Gluon-Plasmas. In diesem Zustand sind Quarks und Gluonen nicht länger in Hadronen gebunden. Es wird angenommen, dass ein solcher Materiezustand kurz nach dem Urknall, vor der Bildung der Hadronen, existierte.

**ATLAS - A Torodial LHC Apparatus:** ATLAS [9] ist ein Vielzweck-Detektor, der das gesamte Spektrum der teilchenphysikalischen Fragestellungen am LHC abdecken soll. Dies reicht von der Suche nach dem Higgs-Boson bis zur Suche nach Anzeichen neuer Physik, also Phänomenen, die mit dem Standardmodell der Teilchenphysik nicht erklärbar sind. Hierzu gehören z. B. die Supersymmetrie und die Existenz zusätzlicher Dimensionen. Ein weiteres Ziel ist die detaillierte Untersuchung der Elementarteilchen, die aus Proton-Proton-Kollisionen bei hohen Schwerpunktsenergieen von bis zu  $\sqrt{s} = 14$  TeV entstehen.

**CMS - Compact Muon Solenoid:** CMS [10] ist wie ATLAS ein Vielzweck-Detektor und verfolgt dieselben physikalischen Ziele. Der Aufbau sowie die genutzten Teildetektoren unterscheiden sich jedoch.

**LHCb - Large Hadron Collider b:** LHCb [11] untersucht vor allem die Wechselwirkung von Teilchen, die b-Quarks enthalten.

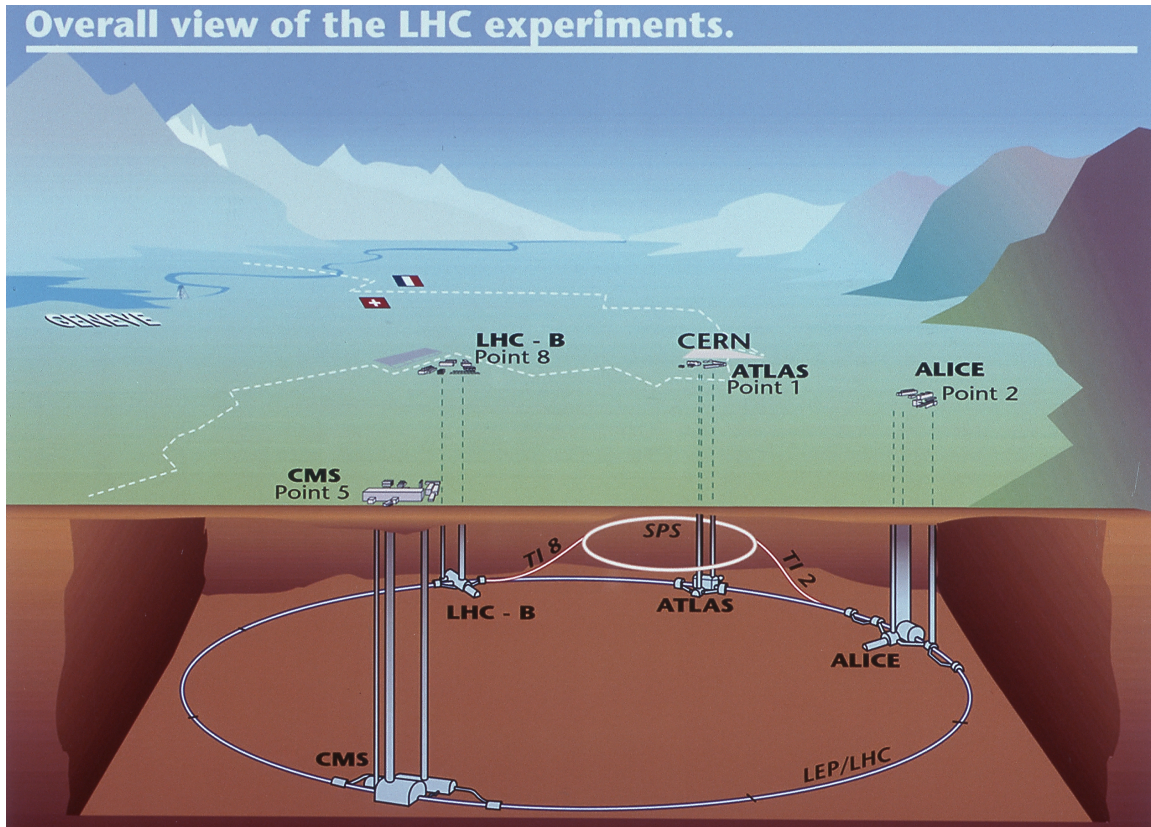


Abbildung 2.1: Die Hauptexperimente des LHC [12]

## 2.2 Das ATLAS-Experiment

Der A Torodial LHC ApparatuS (ATLAS)-Detektor [9] ist der größte der am LHC installierten Detektoren mit einem Gesamtgewicht von 7000 t, einer Höhe von 25 m und einer Länge von 44 m. Er ist aus mehreren Detektormodulen aufgebaut, welche symmetrisch um den Wechselwirkungspunkt angebracht sind. Der Detektor wird aufgrund seiner fassartigen Form in drei Bereiche aus Subdetektoren gegliedert: zwei Endkappen-Bereiche und den Fass-Bereich.

Am nächsten zum Wechselwirkungspunkt befindet sich der Innere Detektor, der von einer dünnen Magnetspule umschlossen wird, die ein 2 T starkes Magnetfeld erzeugt. Der Innere Detektor besteht wiederum aus drei Subdetektoren. Der innerste ist der Pixeldetektor mit 1744 Pixelsensoren mit je 47323 Pixeln. Diese ermöglichen eine hohe räumliche Auflösung der Spurpunkte zur Teilchenspurrekonstruktion von etwa 14  $\mu\text{m}$ . Den Pixeldetektor umschließt der Semi-Conductor-Tracker (SCT), der aus 4088 Modulen mit je 768 Siliziumstreifen-Detektoren besteht. Von diesem Detektor erfasste Spurpunkte besitzen eine Auflösung von 17  $\mu\text{m}$  in radialer Richtung und 580  $\mu\text{m}$  in Strahlrichtung. Der äußerste Teil des Inneren Detektors ist der Transition-Radiation-Tracker (TRT). Hier werden die Spurpunkte mithilfe von Driftrohren ermittelt. Die Auflösung der Spurpunkte dieses Subdetektors ist mit 130  $\mu\text{m}$  die geringste im Inneren Detektor, jedoch werden vergleichsweise viele Spurpunkte geliefert. Aus der rekonstruierten Spur des Teilchens im Magnetfeld des Detektors können sein Impuls und Ladungsvorzeichen bestimmt werden.

Um den Inneren Detektor herum liegen die elektromagnetischen und hadronischen Kalorimeter. Der Zweck des elektromagnetischen Kalorimeter ist es, die Energie elektromagnetisch wechselwirkender Teilchen zu bestimmen. Im hadronischen Kalorimeter wird die Energie von Teilchen gemessen, die der starken Wechselwirkung unterliegen und das elektromagnetische Kalorimeter passiert haben. Beide Kalorimeter bestehen aus sich abwechselnden Schichten von Absorbern und Auslesematerial. Durch die Wechselwirkungen der passierenden Teilchen entstehen in den Absorbern Teilchenschauer, deren Energie im Auslesematerial bestimmt wird. Aus der in den Kalorimetern deponierten Energie lässt sich die Energie der durchdringenden Teilchen rekonstruieren.

Die Kalorimeter sind wiederum vom Myon-Spektrometer umschlossen. Dessen Zweck ist eine präzise Spurrekonstruktion der das Kalorimetersystem verlassenden Teilchen. Aufgrund ihrer hohen Lebensdauer und da sie minimal ionisierende Teilchen sind, sind Myonen die einzigen Teilchen, die diesen Detektor erreichen. Die Spurpunkte der Myonen werden hauptsächlich mit Driftkammern bestimmt, die in Schichten angeordnet sind. Um aus den Spurpunkten die Impulse der Myonen bestimmen zu können, ist das Myon-Spektrometer von einem Magnetfeld umschlossen. Dieses wird im Fassbereich von einem Toroidmagneten aus acht supraleitenden Luftspulen mit einer Stärke von 0,5 T erzeugt, im Endkappen-Bereich von einem weiteren Toroidmagneten, der ein Magnetfeld von 1 T erzeugt.

Die mit dem Detektor gewonnenen Daten werden aus Gründen der begrenzten Kapazität der Auslesegeräte in mehreren Stufen, teilweise schon in der Hardware, gefiltert. Die hierzu genutzten Filter, Trigger genannt, unterscheiden sich je nach Analyse. Beispiele für Trigger sind ein



biasfreier Trigger, der einen gewissen Anteil aller Ereignisse akzeptiert, und Trigger, bei denen nur Ereignisse akzeptiert werden, die vorgegebenen Topologien und Ereignisgrößen entsprechen. Mögliche Ereignisgrößen sind u. a. die invariante Masse und der Transversalimpuls von registrierten Teilchen. Durch die Triggerselektion wird die Datenmenge soweit reduziert, dass die Verarbeitung der Daten möglich wird.



## 3 Statistische Methoden

In diesem Kapitel werden verschiedene Methoden erläutert, die zur Bestimmung der Signifikanz eines vermuteten Signals und der Konfidenzintervalle einer Messung genutzt werden können. Hierzu wird ein einfaches Zählexperiment betrachtet, an welchem verschiedene Methoden erprobt werden. Die anhand dieses Beispiels erläuterten Methoden dienen als Grundlage für die Analyse komplexerer Experimente im folgenden Kapitel.

### 3.1 Signifikanzbestimmung eines Zählperiments

Die erste untersuchte Methode dient der Signifikanzbestimmung des Ergebnisses eines Zählperiments. Der Ausgangspunkt ist eine Messung, bei der  $n_{obs}$  Ereignisse registriert werden. Jeder dieser Messwerte entspricht entweder einem Signalereignis oder einem Untergrundereignis, diese können jedoch nicht unterschieden werden. Zudem sei der Erwartungswert der Anzahl an Untergrundereignissen  $\bar{b}$  bekannt. Zu bestimmen ist nun die Signifikanz der Messung. Hierzu werden folgende Annahmen gemacht: Sowohl die Anzahl  $n_b$  der Untergrundereignisse, als auch die Anzahl der Signalereignisse  $n_s$  folgen Poissonverteilungen mit den Parametern  $\bar{b}$  bzw.  $\bar{s}$ :

$$p_{\bar{b}}(n_b) = \frac{\bar{b}^{n_b}}{n_b!} e^{-\bar{b}}, \quad (3.1)$$

$$p_{\bar{s}}(n_s) = \frac{\bar{s}^{n_s}}{n_s!} e^{-\bar{s}}. \quad (3.2)$$

Somit folgt die Anzahl der Gesamtereignisse  $n_s + n_b$  ebenfalls einer Poissonverteilung. Diese besitzt als Parameter  $\bar{s} + \bar{b}$ :

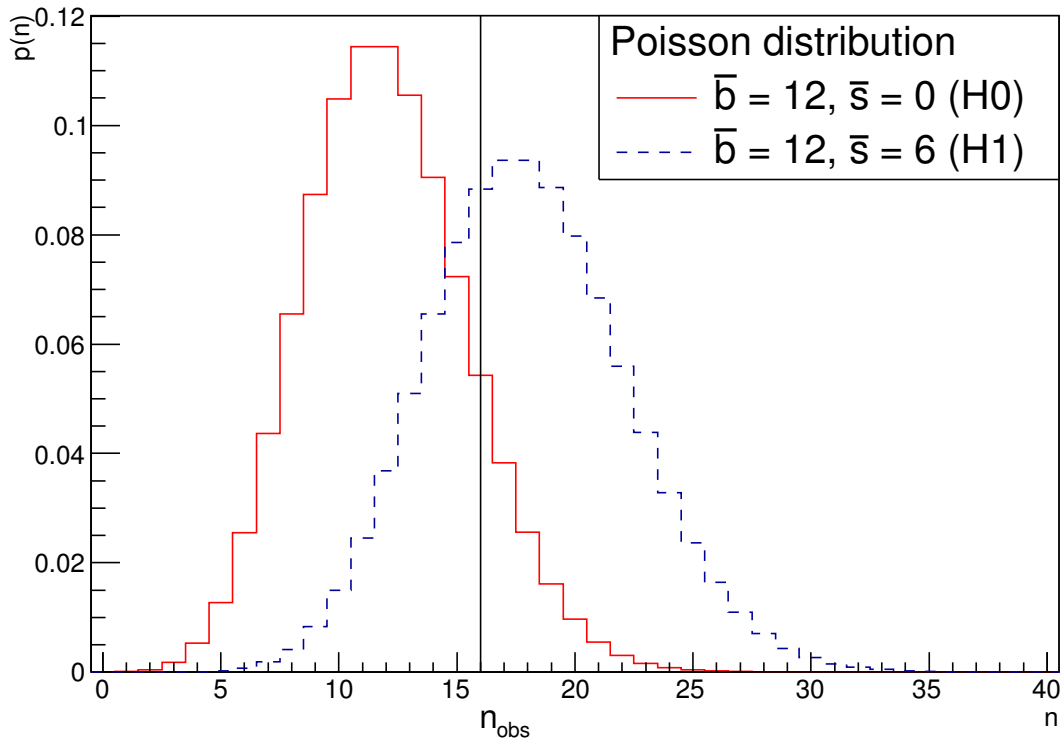
$$p_{\bar{s}+\bar{b}}(n_s + n_b) = \frac{(\bar{s} + \bar{b})^{n_s+n_b}}{(n_s + n_b)!} e^{-(\bar{s}+\bar{b})}. \quad (3.3)$$

Aus diesen Annahmen soll nun die Signifikanz einer Messung von  $n_{obs}$  Ereignissen hergeleitet werden. Hierzu werden sogenannte  $p$ -Werte genutzt. Der  $p$ -Wert ist die Wahrscheinlichkeit, unter der Nullhypothese ein genauso wahrscheinliches oder weniger wahrscheinliches Ergebnis wie das beobachtete zu erhalten. In diesem Fall entspricht die Nullhypothese  $H_0$  einer Messung, die nur Untergrundereignisse beinhaltet ( $\bar{s} = 0$ ). Bei der Signalthypothese  $H_1$  sind Signalereignisse möglich ( $\bar{s} > 0$ ). Würden weitere Zählexperimente unter der Nullhypothese mit Ergebnissen  $n$

durchgeführt, so würde  $n$  der in (3.1) beschriebenen Poissonverteilung folgen. Der Anteil der möglichen Ergebnisse  $n$  dieser Zähl-experimente unter  $H_0$  mit  $n \geq n_{obs}$  ergibt den  $p$ -Wert:

$$p = p(n \geq n_{obs}) = \sum_{n=n_{obs}}^{\infty} p_{\bar{b}}(n) = \sum_{n=n_{obs}}^{\infty} \frac{\bar{b}^n}{n!} e^{-\bar{b}}. \quad (3.4)$$

Der  $p$ -Wert ist somit der Anteil der Ergebnisse  $n$ , die mit  $n \geq n_{obs}$  höchstens genauso wahrscheinlich sind wie  $n_{obs}$ .



**Abbildung 3.1:** Auf 1 normierte Poissonverteilungen für  $\bar{b} = 12$  unter der Nullhypothese ( $\bar{s} = 0$ ) sowie unter der Signalthypothese für  $\bar{s} = 6$  mit dem beobachteten Wert  $n_{obs} = 16$ . Der  $p$ -Wert dieses Experiments beträgt 0.1556 ( $1.013\sigma$ ), was der Fläche rechts von  $n_{obs}$  unter der linken Verteilung  $H_0$  entspricht. Die rechte Verteilung veranschaulicht, wie sich die Poissonverteilung ändert, wenn zu den Untergrundereignissen Signalereignisse hinzukommen. Der Wert  $\bar{s} = 6$  ist willkürlich gewählt und hat keinen Einfluss auf die Bestimmung des  $p$ -Werts.

In Abb. 3.1 sind sowohl ein Beispiel für den  $p$ -Wert eines Experiments mit  $n_{obs} = 16$  gegeben, als auch die Poissonverteilungen unter  $H_0$  und  $H_1$  bei  $\bar{b} = 12$  und  $\bar{s} = 6$  dargestellt.

Üblicherweise werden  $p$ -Werte genutzt, um eine mögliche Entdeckung zu untersuchen. Hierzu wird zunächst ein Signifikanzniveau festgelegt. Liegt der  $p$ -Wert eines Ergebnisses unter diesem Signifikanzniveau, so gilt es als statistisch signifikant. Es ist hierbei wichtig, das Signifikanzniveau festzulegen, bevor der  $p$ -Wert bekannt ist, um eine Beeinflussung durch diesen zu vermeiden.

Typische Signifikanzniveaus, abhängig von der Art der Beobachtungen, sind 5%, 1% sowie der Standard für Entdeckungen in der Teilchenphysik  $5\sigma$ . Dabei steht  $5\sigma$  bei einem beidseitigen Test für den Anteil einer Normalverteilung, der außerhalb des Bereiches von  $-5\sigma$  bis  $5\sigma$  liegt. Bei einem einseitigen Test, werden also nur Abweichungen in eine Richtung untersucht, steht  $5\sigma$  für den Anteil der Normalverteilung über  $5\sigma$ .

## 3.2 Konfidenzintervalle eines Zähl-experiments

Hier werden Methoden untersucht, mit denen Konfidenzintervalle einfacher Zähl-experimente bestimmt werden können. Der Ausgangspunkt ist wie in Abschnitt 3.1 eine Messung, in der  $n_{obs}$  Ereignisse registriert wurden. Jedes dieser Ereignisse entspricht entweder einem Signalereignis oder einem Untergründereignis. Zudem sei der Erwartungswert der Anzahl an Untergründereignissen  $\bar{b}$  bekannt. Zu bestimmen ist die obere Grenze  $s_{up}$  auf die Anzahl an Signalereignissen in der Messung. Wie zuvor gelten folgende Annahmen: Die Anzahl  $n_b$  der Untergründereignisse, die Anzahl der Signalereignisse  $n_s$  sowie die Gesamtzahl der Ereignisse  $n_s + n_b$  folgen den in Abschnitt 3.1 beschriebenen Gleichungen (3.1), (3.2) und (3.3). Ziel ist es nun, aus der Messung von  $n_{obs}$  Ereignissen eine obere Grenze  $s_{up}$  auf den Parameter  $\bar{s}$  zu setzen. Hierzu wird die Signifikanz  $\beta$  der gesuchten Grenze festgelegt. Ein häufig genutzter Wert, der auch hier durchgehend Verwendung findet ist  $\beta = 5\%$ . Dies bedeutet in diesem Fall, dass, falls  $\bar{s} = s_{up}$  gilt, in 5% aller Fälle ein Wert  $n < n_{obs}$  erhalten wird. Diese Stelle  $s_{up}$  wird in sämtlichen nachfolgend erläuterten Methoden bestimmt, indem  $\bar{s}$  variiert wird, bis  $\beta = 5\%$  erreicht ist. Das Vertrauensniveau  $CL = 1 - \beta$  beträgt dann 95%. Ein Vertrauensniveau von 95% mit oberer Grenze  $s_{up}$  bedeutet, dass bei unter gleichen Bedingungen durchgeführten Experimenten mit Ergebnissen  $n_i$  in 95% der Fälle der wahre Wert  $\bar{s}$  in den bestimmten Intervallen unterhalb von  $s_{up,i}$  liegen würde, also  $\bar{s} \leq s_{up,i}$  gelten würde.

### 3.2.1 Einfache Methode

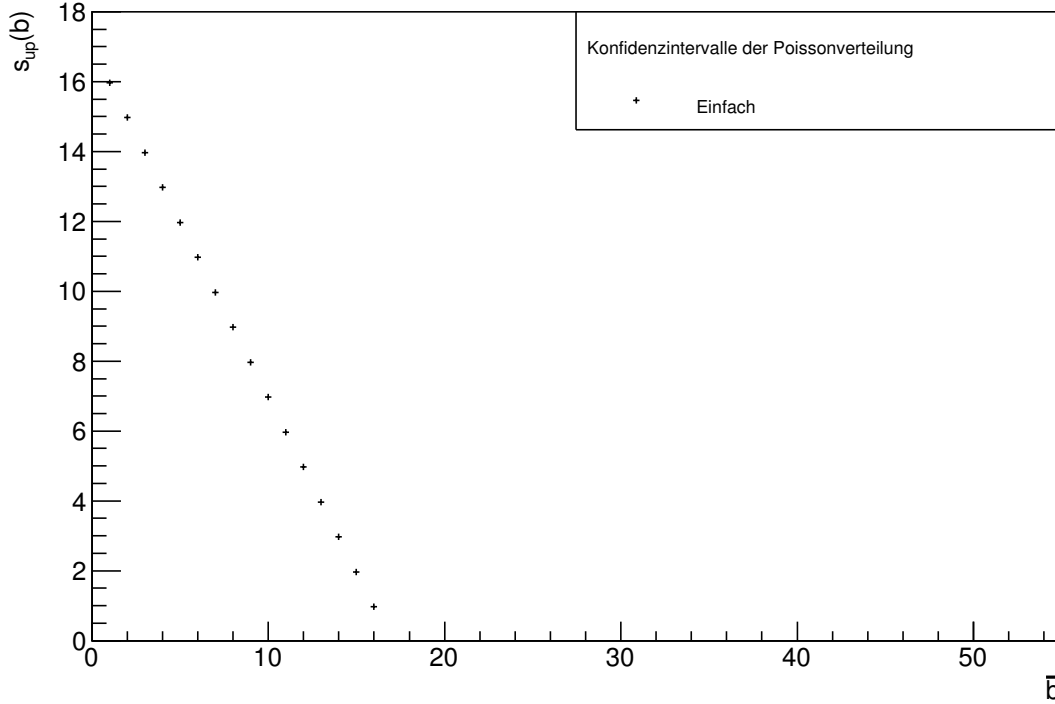
Die einfachste hier betrachtete Methode folgt direkt aus dieser Definition. Dabei wird  $\bar{s} = s_{up}$  so gewählt, dass gilt:

$$\beta = p_{s_{up}+\bar{b}}(n \leq n_{obs}) = \sum_{n=0}^{n_{obs}} \left( \frac{(s_{up} + \bar{b})^n}{n!} e^{-(s_{up} + \bar{b})} \right) = 5\% . \quad (3.5)$$

In der Praxis wird dies dadurch erreicht, dass  $\bar{s}$  solange variiert wird, bis  $\beta = 5\%$  erreicht ist. Der zuletzt erreichte Wert von  $\bar{s}$  entspricht dann dem gesuchten oberen Limit von  $s_{up}$ .

Diese Methode weist einige Schwächen auf. Für eine komplexere Analyse, die nicht nur von einem Messwert abhängt, ist es oft nicht möglich, analytisch die Wahrscheinlichkeit zu bestimmen. Außerdem ist es möglich, eine obere Grenze  $s_{up} < 0$  zu erhalten (Abb. 3.2). Ein wahrer Wert von  $\bar{s} < 0$  ist jedoch in den hier untersuchten Zähl-experimenten nicht möglich, da schon aus den

zugrundeliegenden Verteilungen folgt, dass es nicht weniger als 0 beobachtete Signalereignisse geben kann.



**Abbildung 3.2:** Obere Grenzen  $s_{up}$  für  $n_{obs} = 10$  und  $\beta = 5\%$  bei verschiedenen Werten von  $\bar{b}$ , bestimmt mit der einfachen Methode.

### 3.2.2 Likelihood-Quotienten-Test mit Teststatistik

Diese Methode unterscheidet sich von der vorherigen im Wesentlichen dadurch, dass die Wahrscheinlichkeit  $p_s = p_{\bar{s}+\bar{b}}(n \leq n_{obs})$  nicht mehr analytisch bestimmt, sondern mit einer Teststatistik approximiert wird. Um diese Teststatistik zu erzeugen, werden Zufallszahlen  $n$  aus folgender Poissonverteilung gezogen:

$$p_{\bar{s}+\bar{b}}(n) = \frac{(\bar{s} + \bar{b})^n}{n!} e^{-(\bar{s}+\bar{b})} \bar{s}. \quad (3.6)$$

Zwei Hypothesen sind zu beachten: Die Nullhypothese  $H_0$  ist das Fehlen eines Signals ( $\bar{s} = 0$ ), die zweite Hypothese  $H_1$  ist die Existenz eines Signals ( $\bar{s} > 0$ ). Die Wahrscheinlichkeit für eine Hypothese unter einem gegebenen Messwert  $n_{obs}$  ist durch die entsprechende Likelihood-Funktion gegeben:

$$L(n_{obs}|H_0) = \frac{\bar{b}^{n_{obs}}}{n_{obs}!} e^{-\bar{b}}, \quad (3.7)$$

$$L(n_{obs}|H_1) = \frac{(\bar{s} + \bar{b})^{n_{obs}}}{n_{obs}!} e^{-(\bar{s}+\bar{b})}. \quad (3.8)$$

Für einen Hypothesentest ist eine möglichst gute Trennung dieser beiden Hypothesen wünschenswert. Das Neyman-Pearson-Lemma (NPL) besagt, dass dies durch die Transformation von  $n$  zum Likelihood-Quotienten (LQ)  $t(n)$  erreicht wird [13]:

$$t(n) = -2 \cdot \ln \left( \frac{L(n|H_1)}{L(n|H_0)} \right) . \quad (3.9)$$

Das Einsetzen der Likelihood-Funktionen des obigen Beispiels ergibt:

$$t(n) = -2 \cdot n \cdot \ln \left( 1 + \frac{s}{b} \right) + 2 \cdot s . \quad (3.10)$$

Der Faktor (-2) sowie die Nutzung des Logarithmus des Likelihood-Quotient (LQ) sind Konvention und haben keinen Einfluss auf das endgültige Ergebnis. Um eine Teststatistik für eine Hypothese bei einem bestimmten Wert von  $\bar{s}$  zu erzeugen, wird zunächst ein Satz von Werten  $n_i$  aus der entsprechenden Wahrscheinlichkeitsdichtefunktion (WDF) ((3.6) mit  $\bar{s} = 0$  für  $H_0$  bzw.  $\bar{s} > 0$  für  $H_1$ ) gezogen. Mit (3.9) transformiert, ergeben sich aus diesen Pseudoexperimenten die Einträge  $t_i$  der Teststatistiken für  $H_0$  und  $H_1$ . Solchermaßen erzeugte Teststatistiken sind in Abbildung 3.3 zu sehen. Der beobachtete Wert

$$t_{obs}(n_{obs}) = -2 \cdot \ln \left( \frac{L(n_{obs}|H_1)}{L(n_{obs}|H_0)} \right) \quad (3.11)$$

ist ebenso in der Grafik markiert. Die Wahrscheinlichkeit, ein  $n \leq n_{obs}$  zu ziehen, entspricht nun dem Anteil der Werte  $t \geq t_{obs}$  in den Pseudoexperimenten:

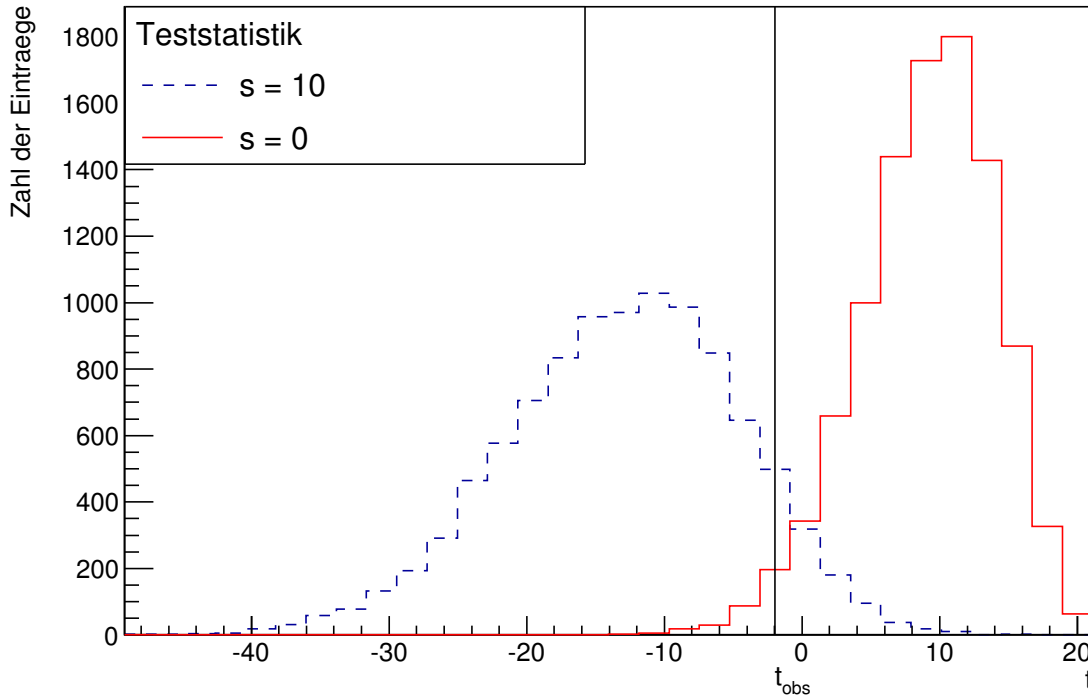
$$p(n \leq n_{obs}) = p(t \geq t_{obs}) . \quad (3.12)$$

Im Beispiel wird der Wert  $s_{up}$  erreicht, wenn  $p(t \geq t_{obs}) = \beta$ . In der praktischen Umsetzung wird dies wieder dadurch erreicht, dass  $\bar{s}$  variiert wird, bis  $p(t \geq t_{obs}) = \beta$  erreicht ist. Der damit erreichte Wert entspricht der oberen Grenze  $s_{up}$ .

Diese Methode ist numerisch und damit universeller anwendbar. Es ist allerdings immer noch möglich, im Bereich von  $\bar{b} > n_{obs}$  eine unphysikalische obere Grenze von  $s_{up} < 0$  zu erhalten ( Abb. 3.4).

### 3.2.3 $CL_s$ -Methode

Ähneln sich die beiden untersuchten Hypothesen, in diesem Beispiel bei einem kleinen Wert von  $\bar{s}$ , entsteht ein großer Überlapp der beiden Teststatistiken. Somit ist eine Trennung der Hypothesen unsauberer und die Analyse verliert an Aussagekraft. Ein Lösungsvorschlag zu diesem Problem ist die  $CL_s$ -Methode [14]. Diese Methode berücksichtigt den Verlust der Aussagekraft,



**Abbildung 3.3:** Teststatistiken für  $n_{obs} = 10$  und  $\bar{b} = 5$  unter den Hypothesen  $s = 10$  ( $H_1$ ) und  $s = 0$  ( $H_0$ ).

sodass auch bei kleinen Signalstärken ein aussagekräftiges Konfidenzlimit erhalten wird. Das so erhaltene Konfidenzlimit wird wie folgt konstruiert:

$$CL_s = \frac{p_s}{1 - p_b} = \frac{p(t \geq t_{obs}|H_1)}{1 - p(t \leq t_{obs}|H_0)}. \quad (3.13)$$

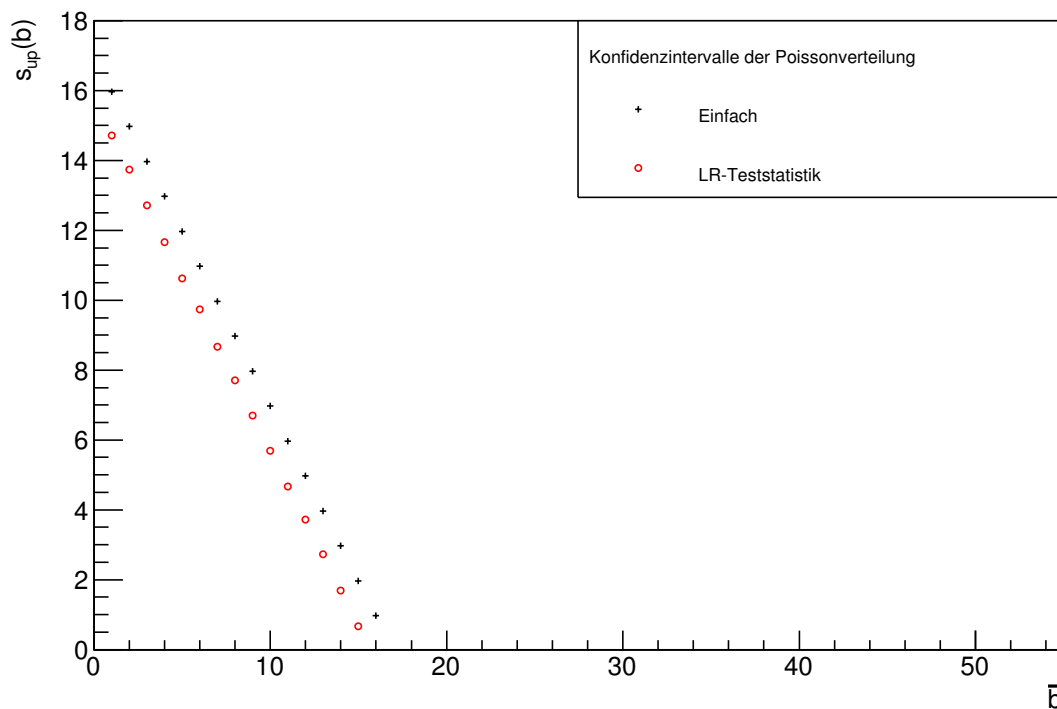
Dieses unterscheidet sich von dem in Abschnitt 3.2.2 genutzten Limit um den Faktor  $\frac{1}{1-p_b}$ . Dieser Faktor erhöht das Konfidenzlimit bei großen Überschneidungen von Signal und Untergrundverteilung. Bei kleinen Überschneidungen werden dieselben Werte wie mit der vorherigen Methode erhalten.

Die gesuchte obere Grenze  $s_{up}$  des Konfidenzintervalls ist mit dieser Methode erreicht, wenn gilt:

$$\beta = CL_s(s_{up}). \quad (3.14)$$

Das Konfidenzlimit wird aus einer Teststatistik bestimmt, die, wie in Abschnitt 3.2.2 beschrieben, erzeugt wurde. Durch die Variation von  $s_{up}$  bis zur Erfüllung der Bedingung (3.14) lässt sich numerisch der korrekte Wert  $s_{up}$  finden. Der große Vorteil dieser Methode ist, dass hier nicht nur die Teststatistik der Signalthypothese  $H_1$  betrachtet wird, sondern auch  $H_0$  bei der Bestimmung des Limits berücksichtigt wird. Jedoch entspricht ein Wert von  $\beta = 5\%$  in diesem Fall nicht mehr der Aussage, dass unter  $\bar{s} = s_{up}$  in 5% aller Fälle ein Wert  $n < n_{obs}$  beobachtet





**Abbildung 3.4:** Obere Grenzen  $s_{up}$  für  $n_{obs} = 10$  und  $\beta = 5\%$  bei verschiedenen Werten von  $\bar{b}$ , bestimmt mit der einfachen Methode sowie der Teststatistik.

wird. Diese Eigenschaft geht aufgrund des  $\frac{1}{1-p_b}$  Terms für Werte von  $p_b > 0$  verloren. Für ein mit dieser Methode bestimmtes  $s_{up}$  ist der Anteil an Werten in  $n < n_{obs}$  kleiner als bei den zuvor beschriebenen.

Die mit der  $CL_s$ -Methode bestimmten Werte für  $s_{up}$  bei konstantem  $n_{obs}$  und unterschiedlichen Erwartungswerten  $\bar{b}$  werden im Vergleich zu den zuvor besprochenen Methoden in Abb. 3.5 dargestellt. Hier wird deutlich, dass die Methode im Bereich  $b \gg n_{obs}$  anfällig für statistische Fluktuationen wird. Dies ergibt sich aus der Nutzung einer notwendigerweise endlichen Teststatistik zur Bestimmung des oberen Limits. Bei kleinen Werten von  $s_{up}$  geht der Anteil  $p_b$  gegen 1, sodass  $\frac{1}{1-p_b}$  groß wird. Hieraus folgt ein sehr kleiner Wert  $p_s = \beta \cdot (1 - p_b)$ . Aufgrund der endlichen Teststatistik führt dies dazu, dass die Anzahl  $N_{t > t_{obs}} = p_b N_{ges}$  der Werte im Bereich ( $t > t_{obs}$ ) klein wird (Abb. 3.6).

Mit einer geringen Anzahl an Ereignissen in diesem Bereich wird die Bestimmung von  $s_{up}$  somit ungenau und verliert im Extremfall ihre Aussagekraft. Dies lässt sich durch die Nutzung einer größeren Anzahl von Pseudoexperimenten beheben. Der Effekt ist jedoch gering und das Problem tritt für einen größeren Unterschied zwischen  $b$  und  $n_{obs}$  immer noch auf. Ein weiterer Schwachpunkt dieses Lösungsansatzes ist die linear mit der Anzahl der Ereignisse steigende benötigte Rechenleistung zur Erzeugung der Teststatistik-Verteilungen.

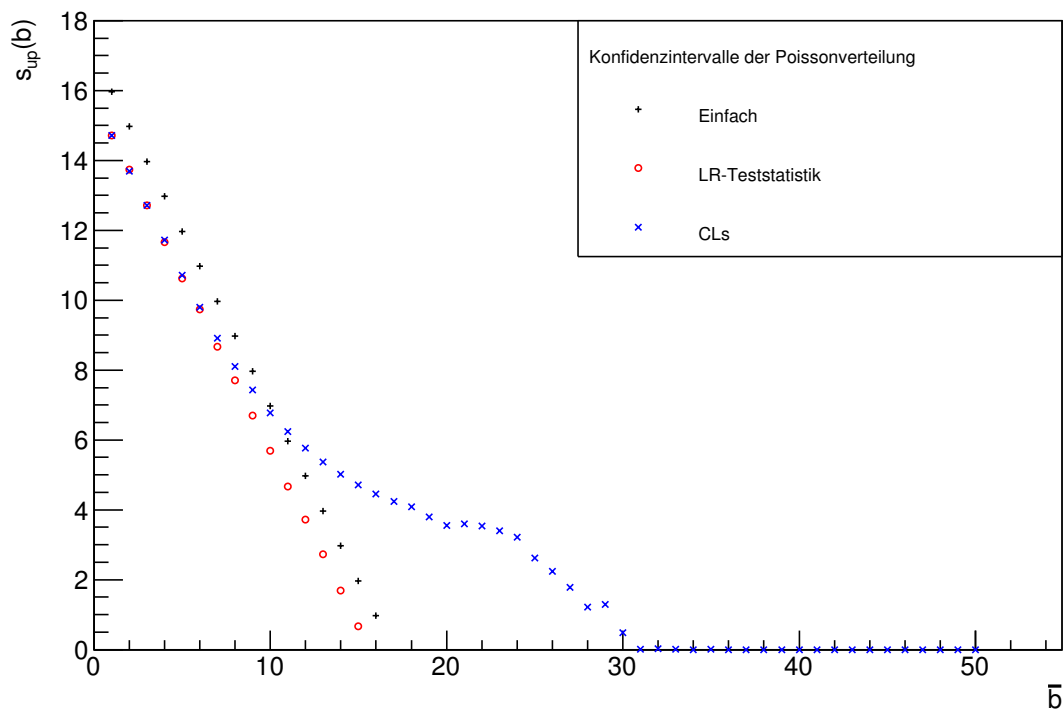


Abbildung 3.5: Obere Grenzen  $s_{up}$  für  $n_{obs} = 10$  und  $\beta = 5\%$  bei verschiedenen Werten von  $\bar{b}$ , bestimmt mit den drei beschriebenen Methoden.

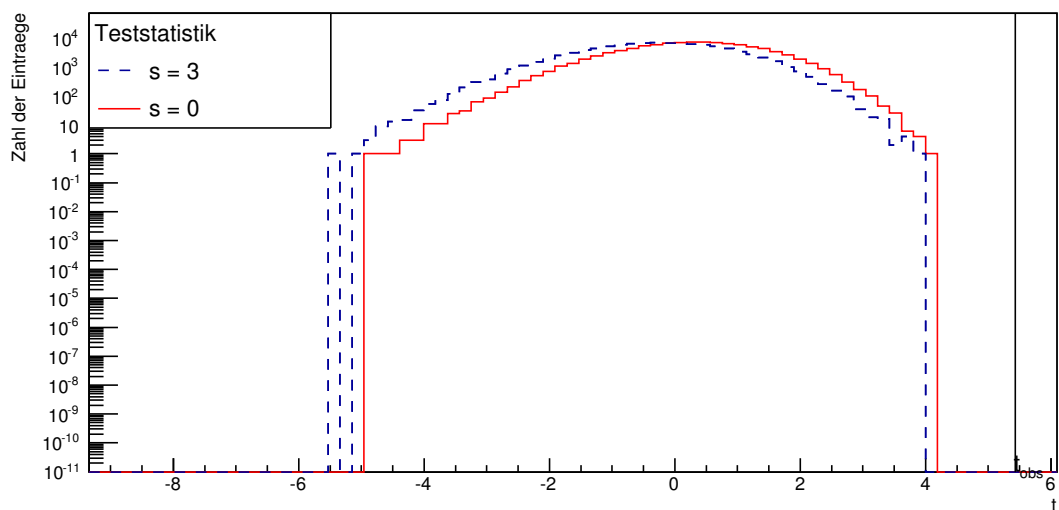


Abbildung 3.6: Teststatistiken für  $n_{obs} = 3$  und  $\bar{b} = 30$  unter den Hypothesen  $s = 3$  ( $H_1$ ) und  $s = 0$  ( $H_0$ ).

### 3.2.4 Importance Sampling-Methode

Das Ziel der Importance Sampling-Methode ist es, die Präzision der Abschätzung des Anteils der Ereignisse einer MC-Simulation in Bereichen niedriger Ereignisstatistik zu erhöhen. In diesem Fall ist es wünschenswert, die Genauigkeit der Bestimmung des Wertes  $p_s$  zu erhöhen. Dies wird dadurch erreicht, dass mehr Ereignisse aus der Region  $t > t_{obs}$  gezogen werden. In der erzeugenden Poissonverteilung sind das die Werte  $n \rightarrow 0$ . Eine höhere Wahrscheinlichkeitsdichte bei  $n \rightarrow 0$  wird erreicht, indem man nicht aus einer Poissonverteilung  $P_\lambda(n)$  zieht, sondern aus einer Verteilung  $g_\mu(n)$  mit demselben Definitionsbereich, jedoch einer höheren Wahrscheinlichkeit für Werte in dem Bereich zuvor niedriger Statistik. Eine geeignete Funktion für das Beispiel der  $CL_s$ -Methode ist folgende Abwandlung der Exponentialfunktion, die als Definitionsbereich die natürlichen Zahlen hat:

$$g_\mu(n) = \int_n^{n+1} \frac{1}{\mu} e^{-\frac{x}{\mu}} dx. \quad (3.15)$$

Diese Funktion hat sowohl eine hohe Wahrscheinlichkeit für kleine Werte  $n$ , als auch denselben Definitionsbereich wie die Poissonverteilung. Wird die Teststatistik nun aus dieser Verteilung erzeugt, d. h. es werden Werte für  $n$  aus  $g_\mu(n)$  gezogen und diese wie zuvor mit Gleichung (3.10) transformiert, so hat die Teststatistik eine andere Form und nicht mehr die einer Poissonverteilung (Abb. 3.7). Um dies zu erreichen wird jedem  $t(n)$  ein Gewicht zugeordnet (Abb. 3.8):

$$w_i(t(n_i)) = w_i(n_i) = \frac{P_\lambda(n_i)}{g_\mu(n_i)}. \quad (3.16)$$

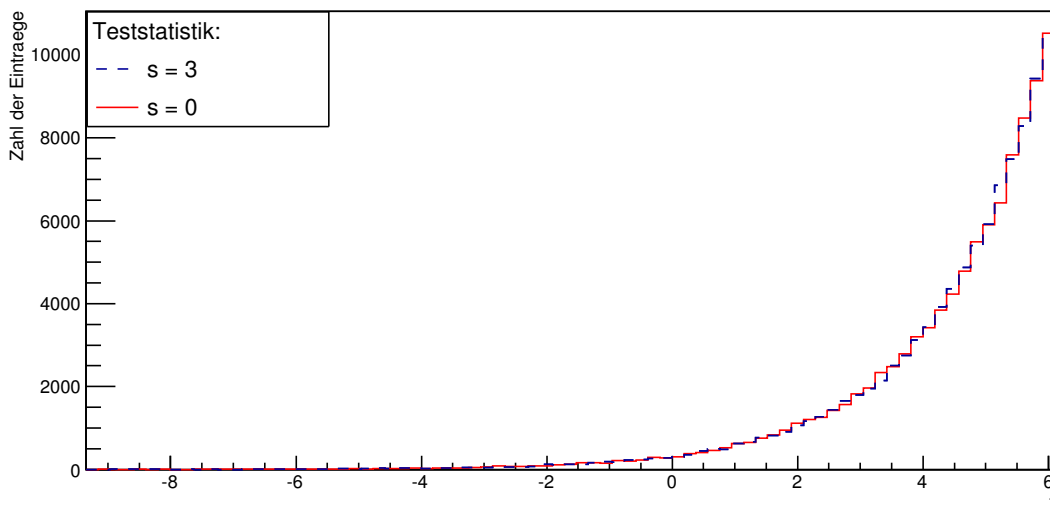
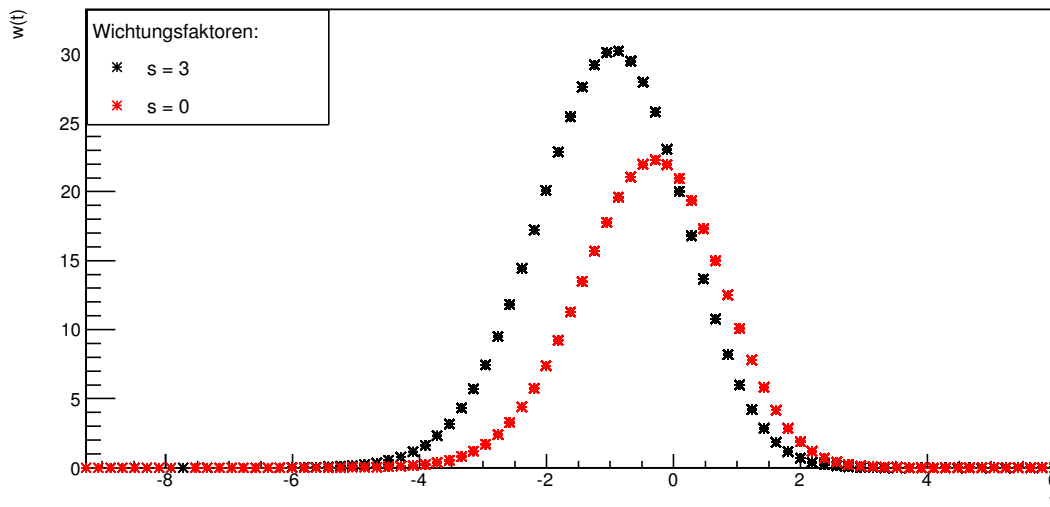


Abbildung 3.7: Aus  $g_\mu(n)$  gezogene Teststatistiken für  $\bar{b} = 30$ .

Eingetragen in ein Histogramm haben die gewichteten Werte  $w_i(n_i)$  die in Abb. 3.9 dargestellte Form. Die so erhaltene neue Teststatistik entspricht der ohne Importance Sampling erzeugten



**Abbildung 3.8:** *Wichtungsfaktoren  $w_i = \frac{P_\lambda(n_i)}{g_\mu(n_i)}$  für die Hypothesen  $s = 0$  ( $H_0$ ) und  $s = 3$  ( $H_1$ ) bei  $\bar{b} = 30$ .*

aus den vorherigen Abschnitten mit dem Unterschied, dass die Häufigkeit der Ereignisse an den Rändern der Statistik größer ist (mehr Werte mit kleinem Gewicht), während die Häufigkeit der Ereignisse in der Mitte abgenommen hat (weniger Werte mit großem Gewicht) (Abb. 3.9 und Abb. 3.6). Häufigeres Ziehen aus den zuvor unterdrückten Ausläufern der Teststatistik verringert die statistischen Fluktuationen in diesem Bereich. Somit können in dem betrachteten Beispiel mit der  $CL_s$ -Methode obere Grenzen  $s_{up}$  auch für große Werte von  $\bar{b}$  verlässlich bestimmt werden (Abb. 3.10).

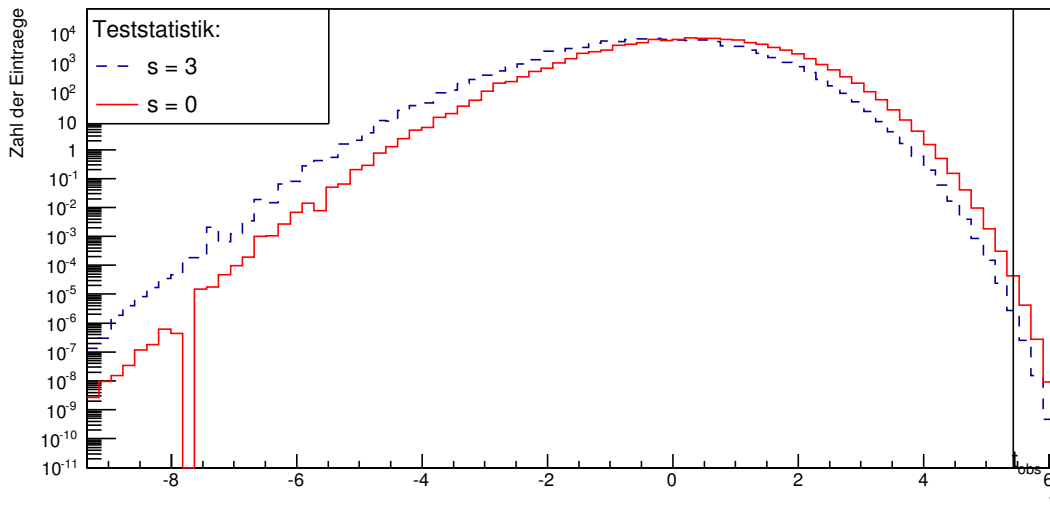


Abbildung 3.9: Mit Importance-Sampling erstellte Teststatistiken für die Hypothesen  $s = 0$  ( $H_0$ ) und  $s = 3$  ( $H_1$ ) bei  $\bar{b} = 30$  und  $n_{obs} = 3$ .

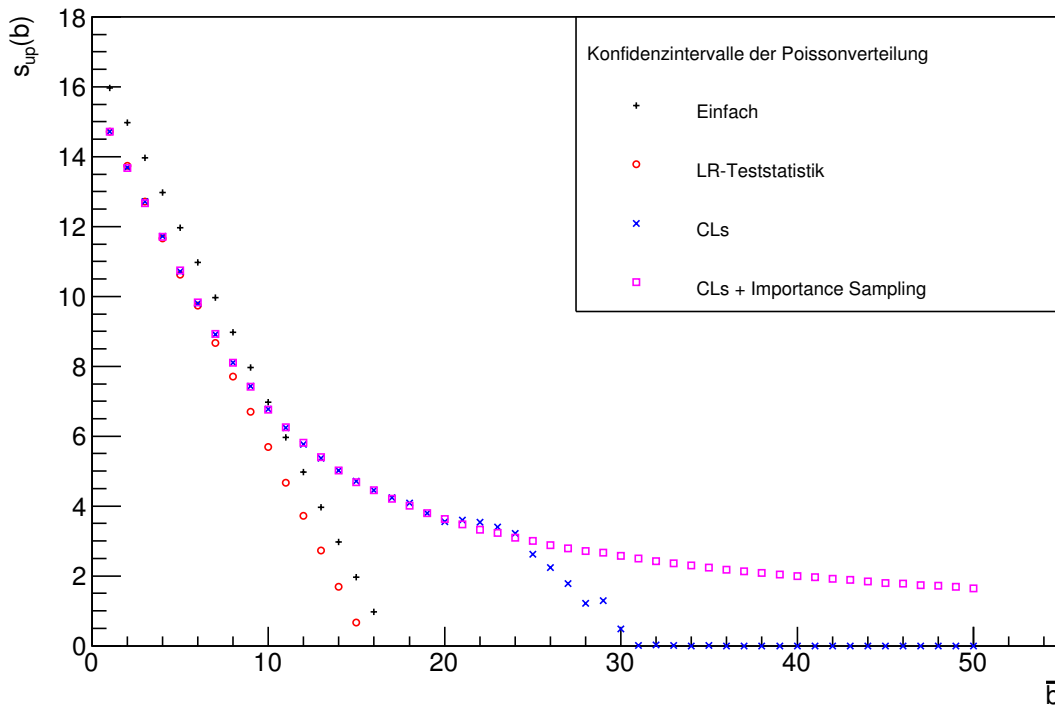


Abbildung 3.10: Obere Grenzen  $s_{up}$  für  $n_{obs} = 10$  und  $\beta = 5\%$  bei verschiedenen Werten von  $\bar{b}$ , mit allen vier Methoden bestimmt.

## 4 Der Look-Elsewhere-Effekt

In den folgenden Abschnitten wird der Look-Elsewhere-Effekt (LEE) beschrieben. Der erste Abschnitt ist eine allgemeine Beschreibung des LEE, im zweiten Teil wird die Relevanz des LEE am Beispiel der Analyse von simulierten Datensätzen eines Diphoton-Massenspektrums diskutiert.

### 4.1 Allgemeine Beschreibung

Der LEE ist ein Phänomen, das unter anderem in der statistischen Analyse wissenschaftlicher Experimente auftritt [15]. Dies ist beispielsweise bei Experimenten der Teilchenphysik zu beachten, wenn eine scheinbar statistisch signifikante Beobachtung aufgrund der Größe des Parameterraums auch nur eine statistische Fluktuation sein kann. Der LEE beschreibt den Anstieg der Wahrscheinlichkeit, eine Fluktuation fälschlicherweise als statistisch signifikant einzuschätzen (Fehler 1. Art, Alphafehler), wenn mehr als nur ein freier Parameter betrachtet wird, auch Alphafehler-Kumulierung genannt. Jeder freie Parameter, der nicht von Interesse ist, ist ein Störparameter, der zum LEE beiträgt. Dies ist in der Teilchenphysik beispielsweise dann der Fall, wenn die Ergebnisse eines Experiments in mehreren Parametern auf Abweichungen vom Standardmodell untersucht werden. Auch bei der Untersuchung von nur einer Messgröße muss der LEE berücksichtigt werden, so zum Beispiel bei einem Massenspektrum, wenn dieses nicht nur bei einem bekannten Massenwert auf Abweichungen untersucht wird, sondern an mehreren Stellen oder innerhalb eines Bereichs. In diesem Fall ist der unbekannte Massenwert des möglichen Signals ein weiterer Störparameter. Dies ist bei der Untersuchung von Diphoton-Ereignissen der Fall, auf die im folgenden Abschnitt eingegangen wird.

### 4.2 Das Diphoton-Massenspektrum als Beispiel

Viele Erweiterungen des Standardmodells sagen Zustände mit hohen Massen voraus, die in zwei Photonen zerfallen [16]. Der Diphoton-Endzustand bietet hier eine gute Auflösung der invarianten Masse, während gleichzeitig nur eine moderate Anzahl an Untergrundereignissen auftritt. Es wurden sowohl am ATLAS- als auch am Compact Muon Solenoid (CMS)-Experiment nach Resonanzen im Diphoton-Massenspektrum gesucht. Hierbei wurden zum Beispiel bei ATLAS Abweichungen vom Untergrundmodell mit lokalen Signifikanzen von bis zu 3.9 Standardabweichungen gefunden [2]. Da die Position der Abweichungen im Vorhinein nicht bekannt ist, muss

der LEE berücksichtigt werden, sodass die globale Signifikanz nur 2.1 Standardabweichungen beträgt. Diese Signifikanz ist zu niedrig, um daraus auf eine Entdeckung zu schließen. Aus diesem Grund wurde eine höhere Datenmenge benötigt, um damit die Signifikanz zu erhöhen. Seit August 2016 ist die Analyse der Abweichung mit zusätzlichen Daten abgeschlossen. Dabei hat sich die Abweichung im Massenspektrum als statistische Fluktuation herausgestellt, die kombinierten Daten des Diphoton-Massenspektrums sind mit dem Standardmodell verträglich [3].

### 4.3 Untersuchung des LEE am Beispiel einer Massenverteilung

In diesem Abschnitt werden nicht mehr Zählexperimente betrachtet, sondern Verteilungen, die ähnlich zu Massenverteilungen sind, wie sie beispielsweise im Diphoton-Massenspektrum beobachtet werden können. Da der Zweck dieser Arbeit nicht die Analyse konkreter Daten ist, werden keine realen Daten aus Experimenten verwendet, sondern die Datensätze selbst generiert. Somit ist sichergestellt, dass sämtliche Faktoren, die Einfluss auf die Erstellung der Daten haben, bekannt sind. Die Generierung der Datensätze wird im folgenden Abschnitt dargelegt.

Wie zuvor im Zählexperiment besteht ein Datensatz aus  $s$  Signalereignissen und  $b$  Untergrundeignissen. Sowohl  $s$  als auch  $b$  folgen Poissonverteilungen mit den Parametern  $\bar{s}$  bzw.  $\bar{b}$ :

$$p_{\bar{s}}(s) = \frac{\bar{s}^s}{s!} e^{-\bar{s}}, \quad (4.1)$$

$$p_{\bar{b}}(b) = \frac{\bar{b}^b}{b!} e^{-\bar{b}}. \quad (4.2)$$

Die Wahl der Poissonverteilung wird durch die Art der Datennahme mit einem Detektor gerechtfertigt. Dort besteht ein Datensatz üblicherweise aus sämtlichen Ereignissen, die in einer gewissen Laufzeit erfasst wurden. Im Mittel werden in der Laufzeit z. B.  $\bar{s}$  Signalereignisse und  $\bar{b}$  Untergrundeignisse erfasst. Dies ist äquivalent zu einem Zählexperiment. Hieraus ergibt sich, dass die Werte  $s$  und  $b$  eines Experiments den oben genannten Poissonverteilungen folgen.

Bei einem Detektorexperiment ist es unmöglich, festzustellen, ob ein individuelles Ereignis des Experiments ein Signal- oder ein Untergrundeignis des untersuchten Zerfalls ist. Allerdings sind verschiedene Parameter, wie die invariante Masse des Zerfalls oder die Energien der Zerfallsprodukte, bekannt. Im verwendeten Beispiel wird immer nur ein Parameter, die invariante Masse, in einem festgelegten Bereich betrachtet.

Für die Verteilungen der Signal- und Untergrundeignisse im Massenspektrum werden die folgenden Wahrscheinlichkeitsdichtefunktionen (WDFs) angenommen. Für das Signal ist dies eine

Normalverteilung  $f(m|\mu_s, \sigma_s)$  mit den Parametern  $\mu_s$  und  $\sigma_s$ , die auf den Bereich von  $m_{min}$  bis  $m_{max}$  normiert ist:

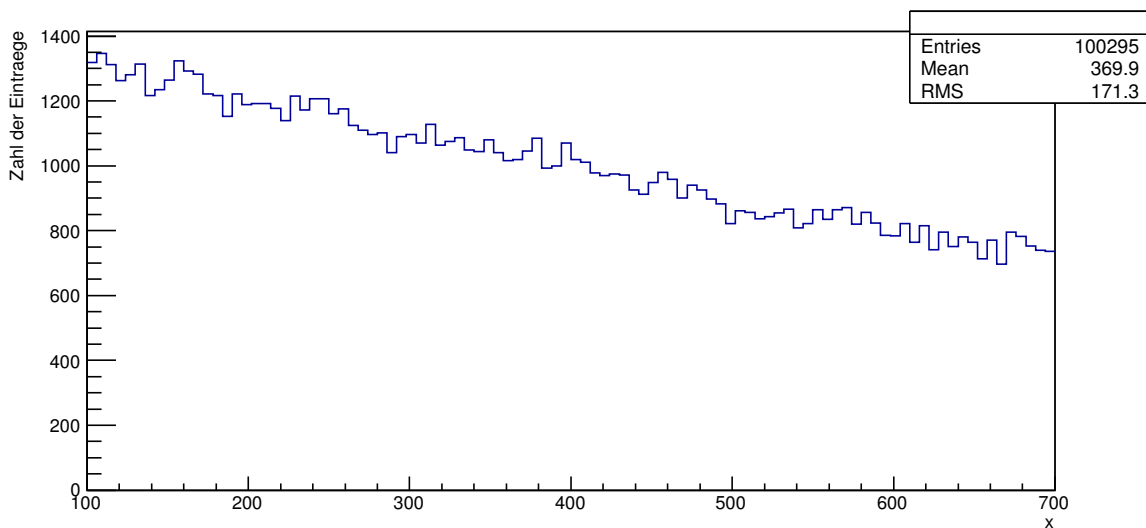
$$f(m|\mu_s, \sigma_s) = \begin{cases} 0 & \text{für } m < m_{min} , \\ \frac{1}{\sqrt{2\pi}\sigma_s} e^{-\frac{(m-\mu_s)^2}{2\sigma_s^2}} \frac{2}{\operatorname{erf}\left(\frac{\mu_s-m_{min}}{\sqrt{2}\sigma_s}\right) + \operatorname{erf}\left(\frac{m_{max}-\mu_s}{\sqrt{2}\sigma_s}\right)} & \text{für } m_{min} \leq m \leq m_{max} , \\ 0 & \text{für } m_{max} < m . \end{cases} \quad (4.3)$$

Der Faktor  $\frac{2}{\operatorname{erf}\left(\frac{\mu_s-m_{min}}{\sqrt{2}\sigma_s}\right) + \operatorname{erf}\left(\frac{m_{max}-\mu_s}{\sqrt{2}\sigma_s}\right)}$  dient der Normierung der Funktion im Bereich von  $m_{min}$  bis  $m_{max}$ , wobei  $\operatorname{erf}(x)$  die gaußsche Fehlerfunktion beschreibt. Die Verteilung der Untergrundeignisse ist auf den gleichen Bereich beschränkt. Für sie wird eine Exponentialverteilung  $g(m)$  mit dem Parameter  $\mu_b$  angenommen:

$$g(m|\mu_b) = \begin{cases} 0 & \text{für } m < m_{min} , \\ \frac{1}{\mu_b} e^{-\frac{m}{\mu_b}} \frac{1}{e^{-\frac{m_{min}}{\mu_b}} - e^{-\frac{m_{max}}{\mu_b}}} & \text{für } m_{min} \leq m \leq m_{max} , \\ 0 & \text{für } m_{max} < m . \end{cases} \quad (4.4)$$

Durch den Faktor  $\frac{1}{e^{-\frac{m_{min}}{\mu_b}} - e^{-\frac{m_{max}}{\mu_b}}}$  ist auch  $g(m|\mu_b)$  normiert.

Um einen Datensatz zu erzeugen, werden  $s$  Werte  $m_i$  aus  $f(m|\mu_s, \sigma_s)$  und  $b$  Werte  $m_i$  aus  $g(m|\mu_b)$  gezogen. Diese Werte, in beliebiger Reihenfolge aneinander gereiht, ergeben den Datensatz. Ein derart erzeugter Datensatz mit geringem Signalanteil ist in Abb. 4.1 dargestellt. Dieser Datensatz, das Beispielerperiment, besteht aus  $s = 198$  Signalereignissen und  $b = 100093$  Untergrundeignissen. Diese Ereigniszahlen sind in der folgenden Analyse nicht bekannt. Der



**Abbildung 4.1:** Ein mit folgenden Parametern erzeugter Datensatz:  $\bar{s} = 200$ ,  $\bar{b} = 100000$ ,  $\mu_s = 400$ ,  $\sigma_s = 20$ ,  $\mu_b = 1000$ ,  $m_{min} = 100$ ,  $m_{max} = 700$ .



Datensatz besteht somit aus  $n_s + n_b$  Werten  $m_i$  im Bereich von  $m_{min}$  bis  $m_{max}$ . Die Werte  $m_i$  folgen der aus (4.3) und (4.4) zusammengesetzten anzahlmäßig gewichteten WDF  $p(m)$ :

$$p(m|\mu_s, \sigma_s, \mu_b) = \frac{n_s}{n_s + n_b} f(m|\mu_s, \sigma_s) + \frac{n_b}{n_s + n_b} g(m|\mu_b). \quad (4.5)$$

Alternativ können gleichartige Datensätze erzeugt werden, indem nach dem Ziehen von  $n_s$  und  $n_b$  aus (4.1) bzw. (4.2) direkt die  $n_s + n_b$  Werte  $m_i$  des Experiments aus (4.5) gezogen werden. In den folgenden Abschnitten wird eine Methode erläutert, die die Aussagekraft dieser Datensätze, zunächst ohne, später mit Berücksichtigung des LEE bestimmt. Die diskutierte Methode nutzt hierzu sogenannte  $p$ -Werte. Der erste Schritt dieser Methode ist, aus dem Experiment mithilfe von Anpassungen Schätzer für die unbekannt Parameter zu erhalten. Das Ergebnis dieser Anpassungen wird interpretiert, indem der  $p$ -Wert des Schätzers für die Signalstärke bestimmt wird, also die Wahrscheinlichkeit, ein genauso wahrscheinliches oder weniger wahrscheinliches Ergebnis bei einem ohne Signal erzeugten Experiment zu erhalten. Es werden Pseudoexperimente genutzt, um die gesuchten Werte zu erhalten.

### 4.3.1 Lokaler $p$ -Wert eines möglichen Signals einer Massenverteilung

In diesem Abschnitt wird erläutert, wie der  $p$ -Wert eines möglichen Signals in einer Massenverteilung bestimmt werden kann. Dazu müssen die Ausgangsbedingungen bekannt sein. In dem Beispielexperiment ist bekannt, dass sowohl  $n_s$  als auch  $n_b$  den Poissonverteilungen (4.1) und (4.2) folgen. Deren Parameter  $\bar{s}$  und  $\bar{b}$ , sowie die konkrete Anzahl der Signalereignisse  $s$  und die Zahl der Untergrundereignisse  $b$  des Beispielexperiments, sind jedoch unbekannt. Die in (4.3) und (4.4) dargestellten WDFs der Signal- und Untergrundverteilungen sind bekannt, ebenso die Parameter  $\sigma_s$  und  $\mu_b$  derselben. Der Ort des Signals  $\mu_s$  ist nicht bekannt, jedoch ist bekannt, dass  $\mu_s$  in einem Bereich zwischen  $\mu_{s,min}$  und  $\mu_{s,max}$  liegt. Außerdem sind die  $n$  Werte  $m_i$ , der Datensatz des Experiments, bekannt. Der hierzu benutzte Datensatz ist der des in Abb. 4.1 dargestellten Beispielexperiments. Anhand dieser Daten werden erst Schätzer für die unbekannt Parameter bestimmt. Danach werden diese Schätzer durch  $p$ -Werte interpretiert. Zunächst werden die wahrscheinlichsten Werte für die Unbekannten  $\mu_s$ ,  $s$  und  $b$  bestimmt. Dies wird durch Maximieren der erweiterten Likelihood-Funktion  $L(\mu_s, s, b)$  erreicht:

$$L(\mu_s, s, b) = \frac{(s+b)^n}{n!} e^{-(s+b)} \prod_{i=1}^n p(m_i|\mu_s, s, b). \quad (4.6)$$

Der Term  $\prod_{i=1}^n p(m_i|\mu_s, s, b)$  bezeichnet hier die normale Likelihood-Funktion. Durch die Erweiterung  $\frac{(s+b)^n}{n!} e^{-(s+b)}$  wird berücksichtigt, dass  $n$  aus einer Poissonverteilung mit unbekanntem Mittelwert  $s+b$  stammt. Die Faktoren  $p(m_i|\mu_s, s, b)$  sind ähnlich zu (4.5) gegeben durch:

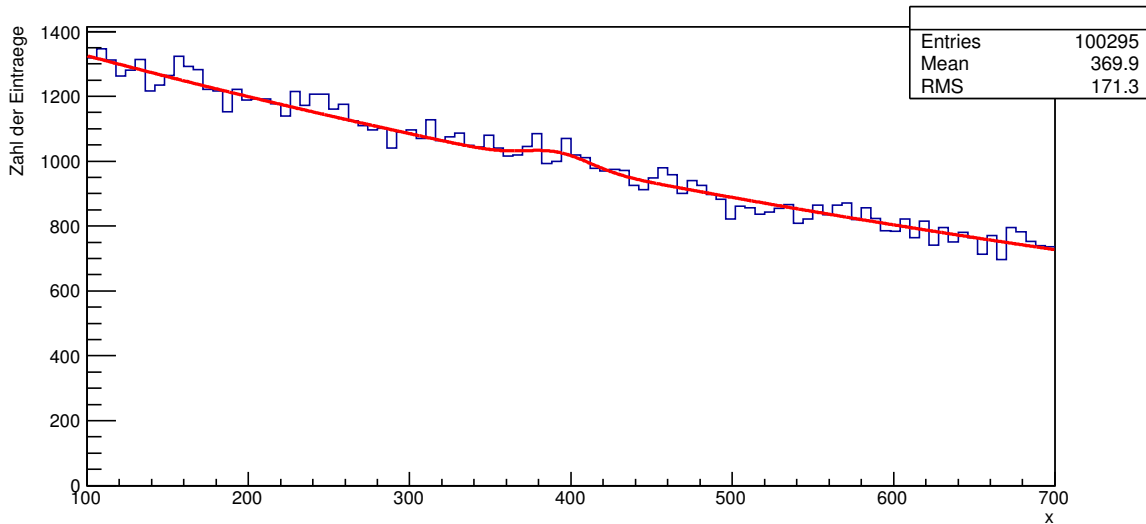
$$p(m_i|\mu_s, s, b) = \frac{s}{s+b} f(m_i|\mu_s, \sigma_s) + \frac{b}{s+b} g(m_i|\mu_b). \quad (4.7)$$

Die Verteilungen  $f(m_i|\mu_s, \sigma_s)$  sowie  $g(m_i|\mu_b)$  sind die im vorherigen Abschnitt beschriebenen WDFs (4.3) und (4.4).

Üblicherweise wird jedoch nicht die Likelihood-Funktion  $L(\mu_s, s, b)$  maximiert, sondern die Log-Likelihood-Funktion  $l(\mu_s, s, b) = -2 \ln(L(\mu_s, s, b))$  minimiert, da in dieser das Produkt durch eine Summe ersetzt wird, was die numerische Berechnung des Wertes vereinfacht:

$$l(\mu_s, s, b) = -2 \left( \ln \left( \frac{(s+b)^n}{n!} e^{-(s+b)} \right) + \sum_{i=1}^n \ln(p(m_i|\mu_s, s, b)) \right). \quad (4.8)$$

Die Werte  $\hat{\mu}_{s, obs}$ ,  $\hat{s}_{obs}$  und  $\hat{b}_{obs}$ , für die  $l(\mu_s, s, b)$  das Minimum erreicht, sind die besten Schätzer der Parameter  $\mu_s$ ,  $s$  und  $b$  des Experiments. Der Bereich, in dem  $\mu_s$  liegen kann, wird im Vorhinein auf Werte zwischen  $\mu_{s, min} = 200$  und  $\mu_{s, max} = 600$  festgelegt. In Abb. 4.2 ist das Ergebnis einer solchen Anpassung dargestellt. Die Histogrammeinträge sind Ereignisse des Experiments, die überlagerte Funktion ist  $p(m|\hat{\mu}_{s, obs}, \hat{s}_{obs}, \hat{b}_{obs})$ , skaliert auf die Gesamtzahl der Ereignisse. Von



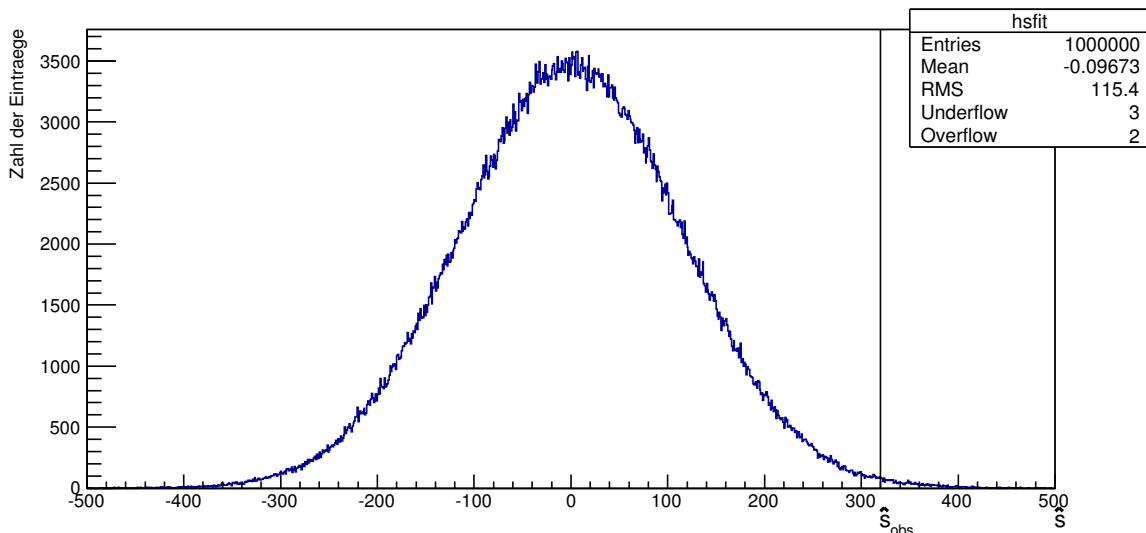
**Abbildung 4.2:** Das Ergebnis der Anpassung an das in Abb. 4.1 dargestellte Experiment. Die erhaltenen Schätzer sind  $\hat{\mu}_{s, obs} = 392 \pm 10$ ,  $\hat{s}_{obs} = 320 \pm 117$  und  $\hat{b}_{obs} = 99974 \pm 336$ .

den Ergebnissen der Anpassung ausgehend, kann nun der  $p$ -Wert bestimmt werden. Dieser ist definiert als die Wahrscheinlichkeit, unter der Nullhypothese ein Ergebnis zu erhalten, das einen Effekt impliziert, welcher weniger wahrscheinlich ist als der gemessene [4]. In diesem Beispiel ist die Nullhypothese  $H_0$  die Abwesenheit von Signalereignissen in dem Experiment ( $s = 0$ ). Ein Effekt, der weniger wahrscheinlich ist als der gemessene, entspricht einem Schätzer  $\hat{s}$  des Experiments, der größer ist als der beobachtete Wert  $\hat{s}_{obs}$ .

Um den  $p$ -Wert eines bestimmten Experiments zu bestimmen, wird ein Satz von Pseudoexperimenten unter  $H_0$  erzeugt. Diese werden wie in Abschnitt 4.3 beschrieben erzeugt, allerdings mit Parametern, die der Nullhypothese entsprechen. Diese sind der zuvor bekannte Parameter

$\mu_b$ , sowie die Parameter  $\bar{s}_0 = 0$  und  $\bar{b}_0$ . Der Wert des Parameters  $\bar{b}_0$  ist das Ergebnis einer Anpassung an das Massenspektrum des Beispielerperiments unter der Nullhypothese, also einer Anpassung, bei der der Parameter  $s = 0$  festgelegt und keine Variable ist. Weil  $\bar{s}_0 = 0$  gilt, gehen die Parameter  $\mu_s$  sowie  $\sigma_s$  nicht in die Erzeugung der Pseudoexperimente ein. Der Grund für die Wahl dieser Werte für  $\bar{s}_0$  und  $\bar{b}_0$  ist, dass sowohl  $\bar{s}_0 = 0$  unter  $H_0$  gelten muss, als auch die Gesamtzahl der Ereignisse der Pseudoexperimente im Mittel gleich der des Experiments sein muss.

Mithilfe dieser Pseudoexperimente kann nun der  $p$ -Wert bestimmt werden. Dies wird einmal ohne Berücksichtigung des LEE durchgeführt (lokal) und einmal mit LEE (global). Der lokale  $p$ -Wert wird bestimmt, indem bei jedem Pseudoexperiment die zuvor beschriebene Anpassung durchgeführt wird, wobei der Parameter  $\mu_s$  auf dem Wert  $\hat{\mu}_{s\text{obs}}$  festgehalten wird. Der Anteil der durch die Anpassung erhaltenen Schätzer  $\hat{s}$ , die grösser sind als  $\hat{s}_{\text{obs}}$  des Beispielerperiments, ist der  $p$ -Wert. Die Verteilung der Schätzer  $\hat{s}$  für unter der Nullhypothese erzeugte Pseudoexperimente kann Abb. 4.3 entnommen werden.

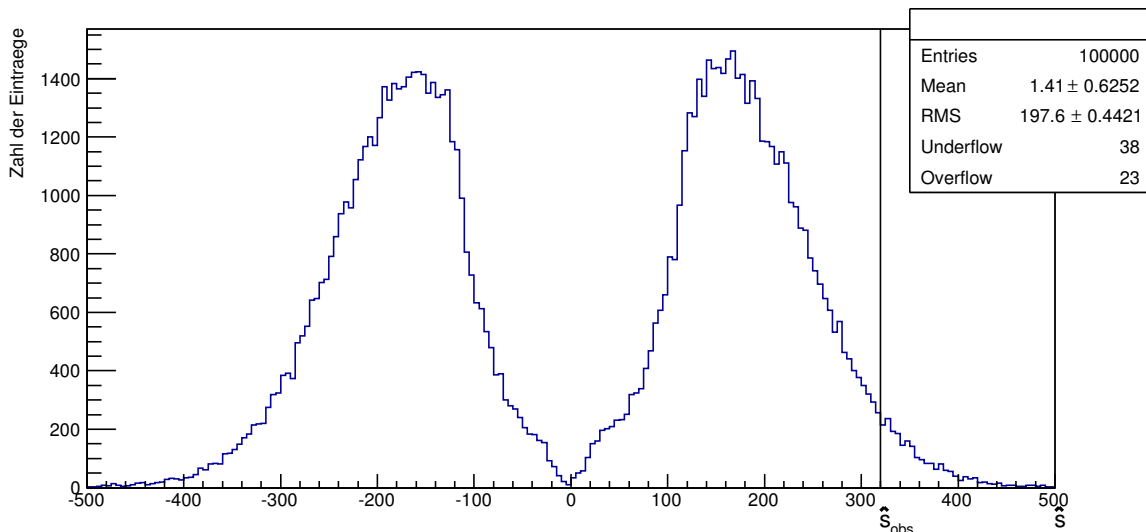


**Abbildung 4.3:** Die Verteilung der Schätzer  $\hat{s}$  bei 1000000 unter  $H_0$  erzeugten Pseudoexperimenten bei einer Anpassung mit festem Wert  $\mu_s = \hat{\mu}_{s\text{obs}}$ . Der Schätzer  $\hat{s}_{\text{obs}}$  des Beispielerperiments ist ebenso eingetragen. Der Anteil der Werte rechts von  $\hat{s}_{\text{obs}}$ , also der lokale  $p$ -Wert, liegt bei  $(0.306 \pm 0.006)\%$ .

Der Wert  $\hat{s}_{\text{obs}}$  ist ebenso eingetragen. Der Anteil aller Werte rechts von  $\hat{s}_{\text{obs}}$  entspricht dem lokalen  $p$ -Wert des Beispielerperiments. Der mit dieser Methode bestimmte lokale  $p$ -Wert des Beispielerperiments beträgt  $(0.306 \pm 0.006)\%$ .

### 4.3.2 Globaler $p$ -Wert eines möglichen Signals einer Massenverteilung

Der globale  $p$ -Wert ist die Wahrscheinlichkeit, irgendwo in dem Massenspektrum bei der Anpassung an die unter  $H_0$  erzeugten Pseudoexperimente einen Wert  $\hat{s} \geq \hat{s}_{obs}$  zu erhalten. Dem entsprechend wird die Anpassung an die Pseudoexperimente mit einem zusätzlichen variablen Parameter  $\mu_s$  für den Massenwert des Signals durchgeführt, der innerhalb festgelegter Grenzen  $\mu_{smin}$  und  $\mu_{smax}$  liegt. Durch das Einführen eines weiteren freien Parameters wird die Verteilung der Schätzer  $\hat{s}$  breiter als zuvor, da es sehr viel wahrscheinlicher ist, irgendwo entlang des Massenspektrums eine statistische Fluktuation zu erhalten als an einem festgelegten Punkt. Somit ist der globale  $p$ -Wert größer als der lokale  $p$ -Wert des gleichen Experiments (Abb. 4.4). Für das Beispielerperiment ergibt sich ein globaler  $p$ -Wert von  $(2.23 \pm 0.05)\%$ .



**Abbildung 4.4:** Die Verteilung der Schätzer  $\hat{s}$  bei 100000 unter  $H_0$  erzeugten Pseudoexperimenten bei einer Anpassung mit variablem Wert  $\mu_s$ . Der Schätzer  $\hat{s}_{obs}$  des Beispielerperiments ist ebenso eingetragen. Der Anteil der Werte rechts von  $\hat{s}_{obs}$ , also der globale  $p$ -Wert, liegt bei  $(2.23 \pm 0.05)\%$ .

### 4.3.3 Vergleich des globalen und lokalen $p$ -Werts und Interpretation

In den vorangehenden Abschnitten wurden Methoden etabliert, den lokalen und globalen  $p$ -Wert eines möglichen Signals auf einer Massenverteilung zu bestimmen. Ebenso wurden der lokale und globale  $p$ -Wert einer Beispielveilung bestimmt. Hierbei ergibt sich ein lokaler  $p$ -Wert von  $(0.306 \pm 0.006)\%$  und ein globaler  $p$ -Wert von  $(2.23 \pm 0.05)\%$ .

Oft wird in der Teilchenphysik jedoch nicht der  $p$ -Wert als solcher angegeben, sondern Standardabweichungen der Normalverteilungen, die dem  $p$ -Wert entsprechen. Die Signifikanz  $\alpha$  eines

Experiments in Standardabweichungen der Normalverteilung lässt sich wie folgt aus dem  $p$ -Wert  $p$  berechnen, wobei  $\text{erf}^{-1}(x)$  die Inverse der gaußschen Fehlerfunktion ist:

$$\alpha = \sqrt{2} \text{erf}^{-1}(1 - 2p) . \quad (4.9)$$

Mit dieser Relation lassen sich der globale und lokale  $p$ -Wert des Experiments in die entsprechende Signifikanz umrechnen. Für den lokalen  $p$ -Wert ergibt sich eine Signifikanz von  $(2.741 \pm 0.006) \sigma$ , dem globalen  $p$ -Wert entspricht eine Signifikanz von  $(2.009 \pm 0.009) \sigma$ .

Die Ergebnisse beider  $p$ -Werte sind in Tab. 4.1 gegeben.

	p-Wert	Signifikanz
lokal	$(0.306 \pm 0.006)\%$	$(2.741 \pm 0.006) \sigma$
global	$(2.23 \pm 0.05)\%$	$(2.009 \pm 0.009) \sigma$

**Tabelle 4.1:** *Ergebnisse der Berechnung der lokalen und globalen  $p$ -Werte des Beispielsperiments.*

Der globale  $p$ -Wert ist höher als der lokale ( $\frac{p_{\text{glob}}}{p_{\text{lok}}} = 7.3$ ) und dementsprechend ist die globale Signifikanz des Experiments geringer. Bei einem echten Experiment, bei dem die Signifikanz eines möglichen Signals untersucht werden soll, hat die Vernachlässigung des LEE somit, in Abhängigkeit von dem zuvor gewählten Signifikanzniveau, die folgenden Auswirkungen.

Liegt das zuvor gewählte Signifikanzniveau über  $2.741 \sigma$  oder unterhalb von  $2.009 \sigma$ , so hat der LEE keinen Einfluss darauf, ob das Ergebnis des Experiments statistisch signifikant ist.

Liegt das Signifikanzniveau jedoch zwischen  $2.009 \sigma$  und  $2.741 \sigma$ , so ist die Signifikanz größer als das Signifikanzniveau, wenn der LEE vernachlässigt wird, und kleiner wenn man den LEE mit einbezieht. Da bei dem betrachteten Experiment der Ort des Signals nicht im voraus bekannt ist, ist hier der Vergleich mit der globalen Signifikanz zu nutzen, d.h. das Experiment ist als statistisch nicht signifikant zu beurteilen. Eine Vernachlässigung des LEE, also das Verwenden der lokalen Signifikanz, kann zu einer Falschaussage führen, da das Experiment als statistisch signifikant eingeschätzt wird, obwohl dies nicht der Fall ist.

In der Teilchenphysik ist ein Signifikanzniveau von mindestens  $3 \sigma$  der Standard für Entdeckungen. Das Ergebnis dieses Experiments wäre damit nicht statistisch signifikant, unabhängig von der Berücksichtigung des LEE. Bei einem Signal mit einer lokalen Signifikanz von etwas über  $3 \sigma$ , beispielsweise  $3.5 \sigma$ , müsste der LEE mit berücksichtigt werden, da die globale Signifikanz durch den LEE unterhalb des Signifikanzniveaus liegen kann, wodurch das Ergebnis statistisch nicht signifikant wird.

Somit besteht bei Analysen mit unbekanntem Variablen die Gefahr, die Signifikanz eines Ergebnisses zu hoch einzuschätzen, was zur verfrühten und möglicherweise falschen Bekanntgabe einer Entdeckung führen kann.

## 5 Zusammenfassung und Ausblick

Am ATLAS-Experiment am Europäischen Forschungszentrum für Elementarteilchenphysik CERN werden durch die Analyse von Proton-Proton-Kollisionen die Eigenschaften und Wechselwirkungen von Elementarteilchen untersucht. Dort wurde im Spektrum der invarianten Masse von Diphoton-Ereignissen ein mögliches Signal, das auf eine Abweichung vom Standardmodell hinweisen könnte, registriert, das inzwischen als statistisch insignifikant eingestuft wurde [3].

Gerade bei möglichen Entdeckungen, bei denen die Position der Signalereignisse in einem Spektrum nicht im Vorhinein bekannt ist, muss bei der Bestimmung der Signifikanz des aufgrund einer Überhöhung im Spektrum möglichen Signals der Look-Elsewhere-Effekt (LEE) mit beachtet werden. Durch diesen Effekt wird die Signifikanz des möglichen Signals bei einer Erweiterung des Parameterraums, also dem Zulassen von Störparametern, verringert. Ein Beispiel hierfür ist die Analyse einer Massenverteilung mit einem möglichen Signal, dessen Massenwert unbekannt ist. Der unbekannte Massenwert entspricht hier dem Störparameter. Um dies zu untersuchen, wurden in der vorliegenden Arbeit zunächst Methoden zur Bestimmung von Signifikanzen und Konfidenzintervallen bei Zählexperimenten erläutert.

Mit MC-Simulationen wurden Datensätze generiert, die Massenverteilungen entsprechen, die aus Untergrundereignissen sowie einer kleinen Anzahl von Signalereignissen an einem Massenwert, der bei der folgenden Analyse unbekannt ist, zusammengesetzt sind. Verschiedene Methoden zur Signifikanzbestimmung wurden genutzt, um die durch das Vernachlässigen des LEE verursachte Signifikanzänderung eines beispielhaft ausgewählten selbst erzeugten MC-Experiments zu untersuchen. Hierzu wurde die Signifikanz auf zwei Arten bestimmt: Zunächst wird der LEE bei einer lokalen Betrachtung, bei der angenommen wird, dass der gefundene Massenwert der korrekte ist, vernachlässigt. Bei der Analyse mit einem Störparameter, also unbekanntem Massenwert des Signals, wird die Signifikanz unter Berücksichtigung des LEE bestimmt. Aufgrund des LEE wird in diesem Fall erwartet, dass die Signifikanz der lokalen Betrachtung höher ist als die Signifikanz bei der Analyse mit Störparametern. Es wurde festgestellt, dass die Signifikanz des Signals durch den LEE von  $(2.741 \pm 0.006) \sigma$  auf  $(2.009 \pm 0.009) \sigma$  fällt. Hierdurch ändert sich die Interpretation bei einem Vergleich mit dem in der Teilchenphysik verwendeten Signifikanzniveau von  $3 \sigma$  für eine Entdeckung nicht. Liegt die Signifikanz des vermuteten Signals jedoch nahe des gewählten Signifikanzniveaus, so kann eine Vernachlässigung des LEE zu einer Überschätzung der Signifikanz führen, durch die das Signal fälschlicherweise als statistisch signifikant interpretiert wird. Es ist zu erwarten, dass der LEE auch einen Einfluss auf die Bestimmung von Konfidenzintervallen hat, wodurch sich eine Verschiebung der oberen Grenzen auf die Modellparameter ergeben kann. Aufgrund der größeren Komplexität der benötigten Methoden, die zur Bestimmung von Konfidenzintervallen benötigt werden, bleibt eine derartige Untersuchung des LEE einer zukünftigen Analyse vorbehalten.

## Liste der verwendeten Akronyme

<b>ALICE</b>	A Large Ion Collider Experiment
<b>ATLAS</b>	A Torodial LHC ApparatuS
<b>CMS</b>	Compact Muon Solenoid
<b>CERN</b>	Conseil Européenne pour la Recherche Nucléaire / Europäische Organisation für Kernforschung
<b>LEE</b>	Look-Elsewhere-Effekt
<b>LHC</b>	Large Hadron Collider
<b>LQ</b>	Likelihood-Quotient
<b>MC</b>	Monte-Carlo
<b>NPL</b>	Neyman-Pearson-Lemma
<b>WDF</b>	Wahrscheinlichkeitsdichtefunktion

## Literaturverzeichnis

- [1] Siehe z. B.: Bogdan Povh, Klaus Rith, Christoph Scholz, and Frank Zetsche, *Teilchen und Kerne. Eine Einführung in die physikalischen Konzepte*, Springer-Verlag GmbH, sixth edition, 2004.
- [2] Search for resonances in diphoton events with the ATLAS detector at  $\sqrt{s} = 13$  TeV, Technical Report ATLAS-CONF-2016-018, CERN, Geneva, Mar 2016.
- [3] Search for scalar diphoton resonances with  $15.4 \text{ fb}^{-1}$  of data collected at  $\sqrt{s}=13$  TeV in 2015 and 2016 with the ATLAS detector, Technical Report ATLAS-CONF-2016-059, CERN, Geneva, Aug 2016.
- [4] Siehe z. B.: G. Bohm and G. Zech, *Introduction to statistics and data analysis for physicists*, DESY, 2010.
- [5] Lyndon Evans and Philip Bryant, LHC Machine, *JINST*, 3:S08001, 2008.
- [6] Michael Benedikt, Paul Collier, V Mertens, John Poole, and Karlheinz Schindl, *LHC Design Report*, CERN, Geneva, 2004.
- [7] Lyndon Evans and Philip Bryant, LHC Machine, *Journal of Instrumentation*, 3(08):S08001, 2008.
- [8] K. Aamodt et al., The ALICE experiment at the CERN LHC, *JINST*, 3:S08002, 2008.
- [9] G. Aad et al., The ATLAS Experiment at the CERN Large Hadron Collider, *JINST*, 3:S08003, 2008.
- [10] S. Chatrchyan et al., The CMS experiment at the CERN LHC, *JINST*, 3:S08004, 2008.
- [11] Jr. A. Augusto Alves et al., The LHCb Detector at the LHC, *JINST*, 3:S08005, 2008.
- [12] CERN Document Server, The four main LHC experiments, CERN-AC-9906026, Jun 1999.
- [13] J. Neyman and E. S. Pearson, On the Problem of the Most Efficient Tests of Statistical Hypotheses, *Philosophical Transactions of the Royal Society of London Series A*, 231:289–337, 1933.
- [14] A L Read, Modified frequentist analysis of search results (the  $CL_s$  method), (CERN-OPEN-2000-205), 2000.
- [15] Eilam Gross and Ofer Vitells, Trial factors or the look elsewhere effect in high energy physics, *Eur. Phys. J.*, C70:525–530, 2010.



- [16] Siehe z. B.: Christoffer Petersson and Riccardo Torre, 750 gev diphoton excess from the goldstino superpartner, *Phys. Rev. Lett.*, 116:151804, Apr 2016.

## Danksagung

Am Ende dieser Arbeit möchte ich den Personen danken, die mich bei der Erstellung dieser Arbeit unterstützt haben.

Zunächst möchte ich mich bei Prof. Dr. Peter Buchholz bedanken, durch den es mir möglich war, in der Arbeitsgruppe Experimentelle Teilchenphysik diese Bachelorarbeit zu schreiben.

Mein besonderer Dank gilt Dr. Wolfgang Walkowiak für seine Anleitung und die vielen Gespräche, die bei der Erstellung dieser Arbeit geholfen haben.

Ich danke Dr. Angel Campoverde für das angenehme Arbeitsklima im Büro.

Ich danke ebenso meiner Familie und Freunden, die mich während der gesamten Arbeit durch Korrekturlesen und ermunternde Worte unterstützt haben.

# 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