Karlsruher Institut für Technologie

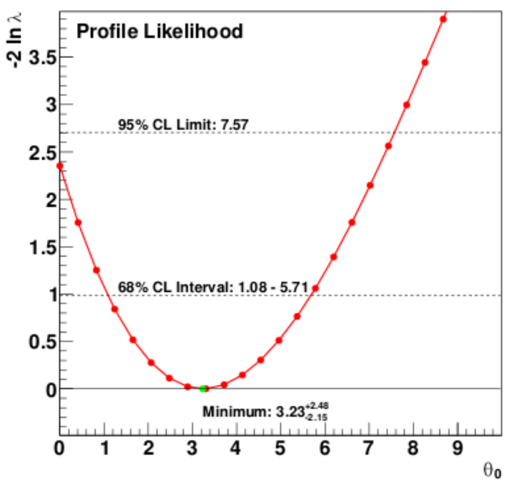
Moderne Methoden der Datenanalyse

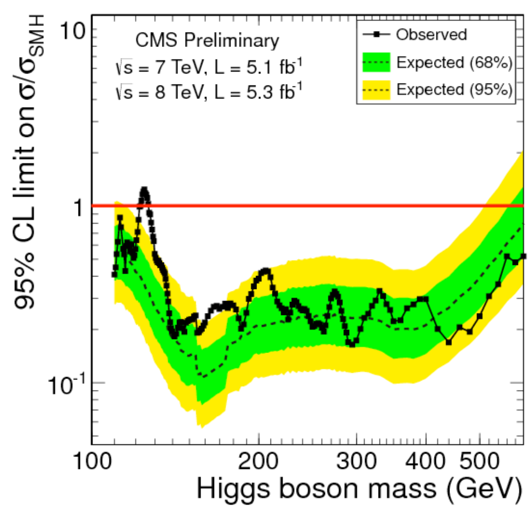
Konfidenz-Intervalle (2)

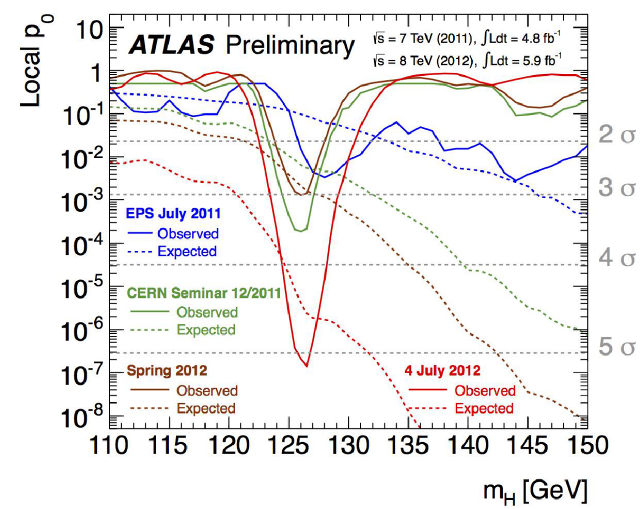
Andreas B.Meyer

20. Juni 2017

Fakultät für Physik
Institut für Experimentelle Kernphysik - IEKP







Vorlesungsprogramm

- Einführung: Überblick und grundlegende Konzepte (1)
- Einführung: Überblick und grundlegende Konzepte (2)
- Zufallszahlen und Monte-Carlo Methoden

A.MEYER

- Hypothesentests
- Parameterschätzung
- Parameterschätzung (Goodness-of-fit)
- Optimierungs- und Parametrisierungsmethoden

R.ULRICH

- Konfidenzintervalle (1)
- Konfidenzintervalle (2)
- Klassifikation (lineare Methoden)
- Klassifikation (ANN, BDT, SVM)
- Klassifikation (Deep Learning)
- Messen und Entfalten
- Systematische Unsicherheiten

A.MEYER

6. Konfidenz-Intervalle (1)

- Neyman-Konstruktion
- Obere und untere Schranken
- Feldman-Cousins Unified Approach

6.1
FREQUENTISTISCHER
ANSATZ

- Einfluss des Priors
- Limits an physikalischen Grenzen

6.2
BAYES'SCHE STATISTIK

- Die CL_S -Methode

6.3
"MODIFIZIERTER FREQUENTISTISCHER" ANSATZ

- Literatur:
  - G.Cowan: *Statistical Data Analysis (Kap. 9)*
  - G.Cowan: *The Review of Particle Physics (Kap. 39)* http://pdg.lbl.gov
  - R.Cousins: *Why isn't every physicist a Bayesian?* Am.J.Phys. 65 (1995) 398.
  - T.Junk: *Confidence Level Computation for combining searches with small statistics*, NIM, A 434 (1999) 435-443.
  - A.Read: *Presentation of Search Results: the CL_S Technique*, J.Phys.G: 28 (2002) 2693-2704.

6. Konfidenz-Intervalle (2)

- Prüfgröße, p-Wert
- Neyman-Pearson Lemma
- Beispiele

6.4
ERINNERUNG:
HYPOTHESENTESTS

- Wilks Theorem
- Störparameter
- Profile-Likelihood-Quotient
- Higgs-Entdeckung
- Look-Elsewhere Effekt

6.5
PROFILE-LIKELIHOOD
QUOTIENT

- Literatur:
  - G.Cowan: *The Review of Particle Physics (Kap. 39)* http://pdg.lbl.gov
  - F.James: *Statistical Methods in Experimental Physics (Kap. 10)*
  - ATLAS+CMS *"Procedure for the LHC Higgs Boson Search combination in Summer 2011* https://cds.cern.ch/record/1379837
  - *G.*Cowan, K.Cranmer, E.Gross, O.Vitells *"Asymptotic Formulae for Likelihood-based tests of new physics,* https://arxiv.org/abs/1007.1727

6.

KONFIDENZ-INTERVALLE (2)

6.4

ERINNERUNG HYPOTHESENTESTS

Hypothesentests / Erinnerung

Karlsruher Institut für Technologie

- Hypothesentests: Anwendung statistischer Methoden zur Entscheidungsfindung, d.h. Ja/Nein oder Accept/Reject
  - Soll ich morgen einen Regenschirm mitnehmen ?
  - Wirkt ein Medikament wirklich?
  - Ist das entdeckte Boson das Higgs-Boson des Standard-Modells?

- Vergleich Daten/Theorie: Stimmen die Daten innerhalb einer vorgegebenen Signifikanz mit dem Untergrund überein?
  - Ausschluss von hypothetischen Signalen typischerweise mit 95% CL
  - Entdeckung von Signalen erfordert größere Abweichung von der Untergrundhypothese, im allgemeinen $> 5\sigma$, (also p-Wert = $5.7 \cdot 10^{-7}$)

“Extraordinary claims require extraordinary evidence”

Konfidenz-Intervalle und Hypothesentests

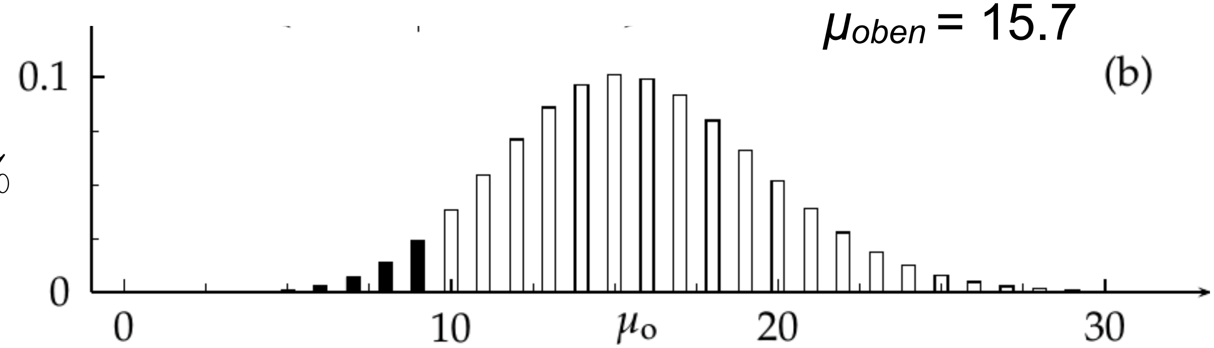
- Konfidenzintervall liefert Angabe der Häufigkeit bei wiederholten Experimenten den wahren Wert in den abgeschätzten Intervallen zu finden.

- Intervallbestimmung kann man als Serie von Hypothesentests auffassen: Variiere hypothetischen wahren Wert solange, bis Signal mit einem p-Wert von 5% ausgeschlossen werden kann. Genau an dem Punkt liegt die obere (einseitige) Grenze mit 95% CL.

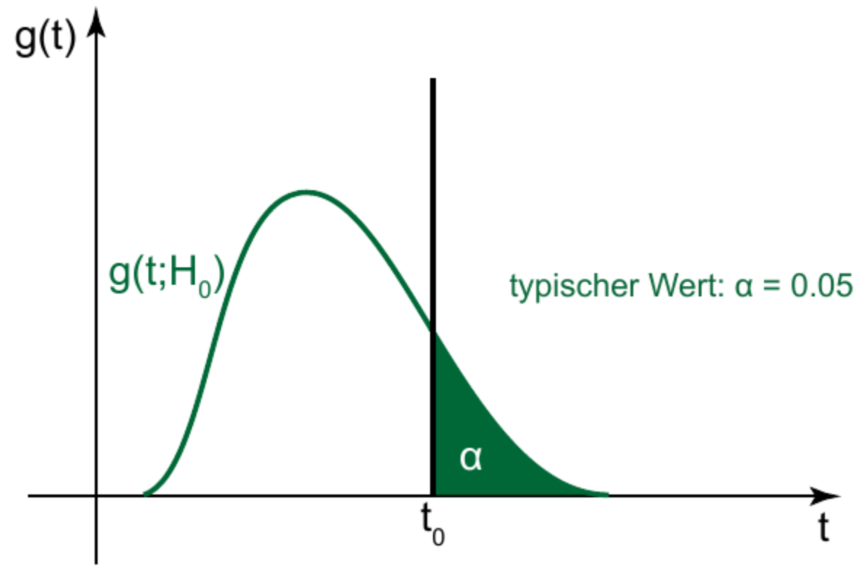
Beispiel Poisson-Verteilung: 9 Ereignisse gemessen

$\mu_{oben} = 15.7$

$$\beta = \sum_{k=0}^{n} P(k|\mu_o) = 5\%$$

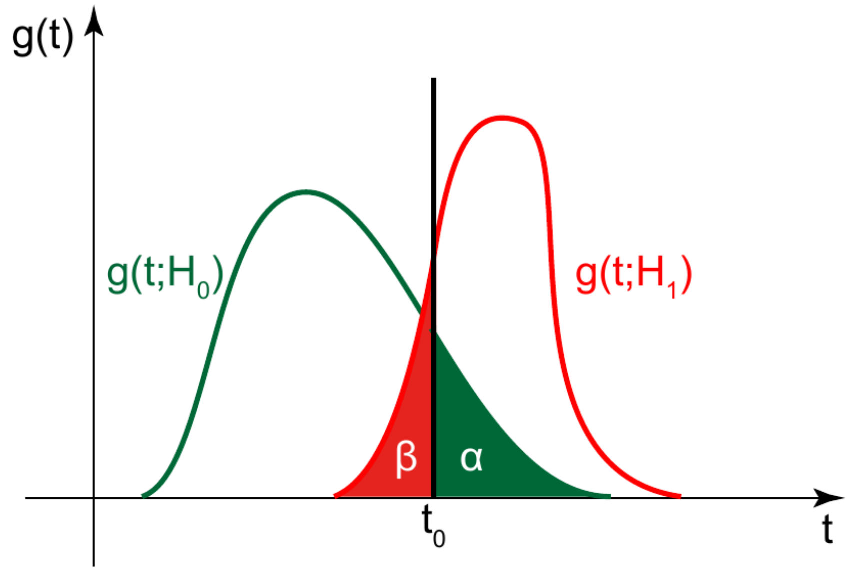


Hypothesentests / Erinnerung



- Hypothesen formuliert als PDF
- Vergleich einer Stichprobe (Daten) mit einer oder mehreren Hypothesen H_i
- Einzelne zu testende Hypothese: Null-Hypothese H_0:
  - Beispiel: Übereinstimmung der Daten mit dem Standard-Modell
  - Häufig: Goodness-of-fit tests
  - Null-Hypothese kann nie bewiesen werden, nur widerlegt - es kann immer eine (geringfügig) bessere Alternative geben.

Hypothesentests / Erinnerung



- Hypothesen formuliert als PDF

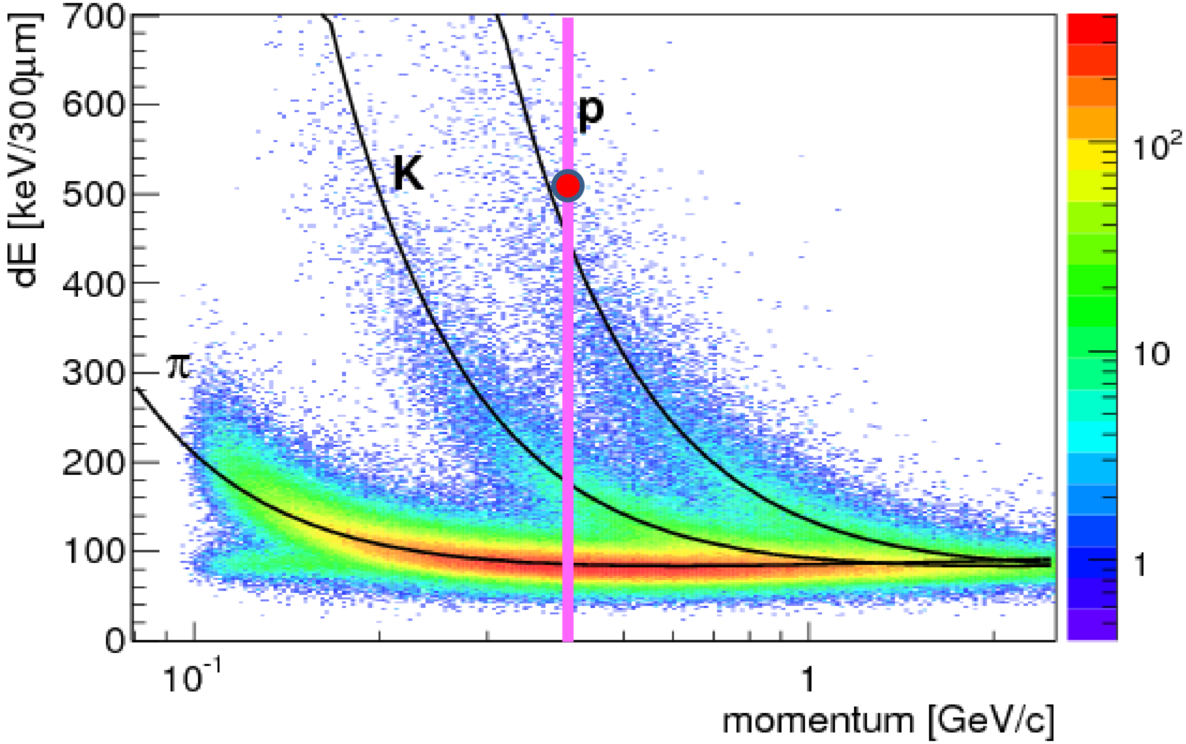
- Vergleich einer Stichprobe (Daten) mit einer oder mehreren Hypothesen H_i

- Einzelne zu testende Hypothese: Null-Hypothese H_0:
  - Beispiel: Übereinstimmung der Daten mit dem Standard-Modell
  - Häufig: Goodness-of-fit tests
  - Null-Hypothese kann nie bewiesen werden, nur widerlegt - es kann immer eine (geringfügig) bessere Alternative geben.

- Mehrere Hypothesen: H_0 und Alternativ-Hypothese(n) H_i
  - Beispiel: Standard-Modell (H_0) vs SM + Neue Physik (H_1)

Mehr als zwei Hypothesen

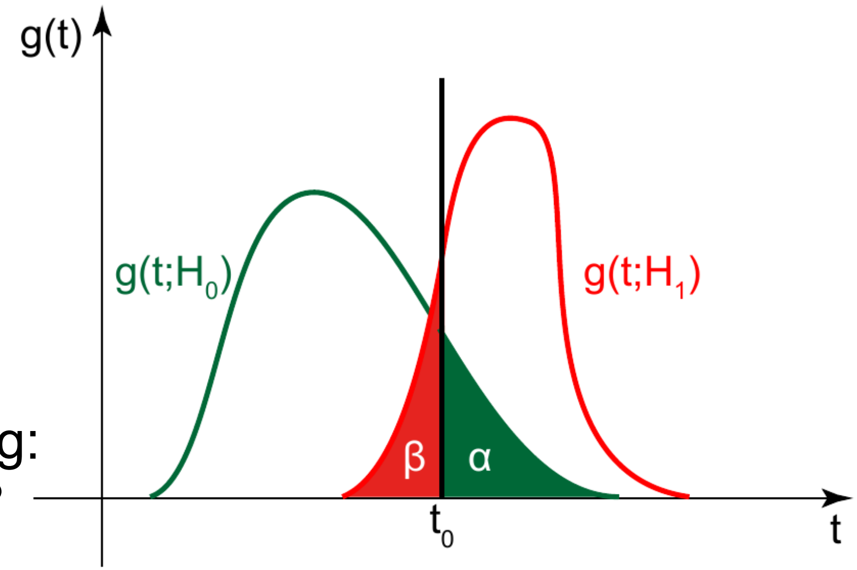
- Beispiel: Energieverlust dE/dx vs. Impuls eines Teilchens in einem Spurdetektor
  - Pion-Hypothese falsifiziert
  - Kaon-Hypothese falsifiziert



- Hypothesen können nie streng verifiziert werden. Es könnte immer noch eine bessere Hypothese geben. Konsistenz mit Hypothesen kann getestet werden.

Vorgehensweise

1. Definition einer Prüfgröße
   und Bestimmung der PDF $g(t;H_i)$



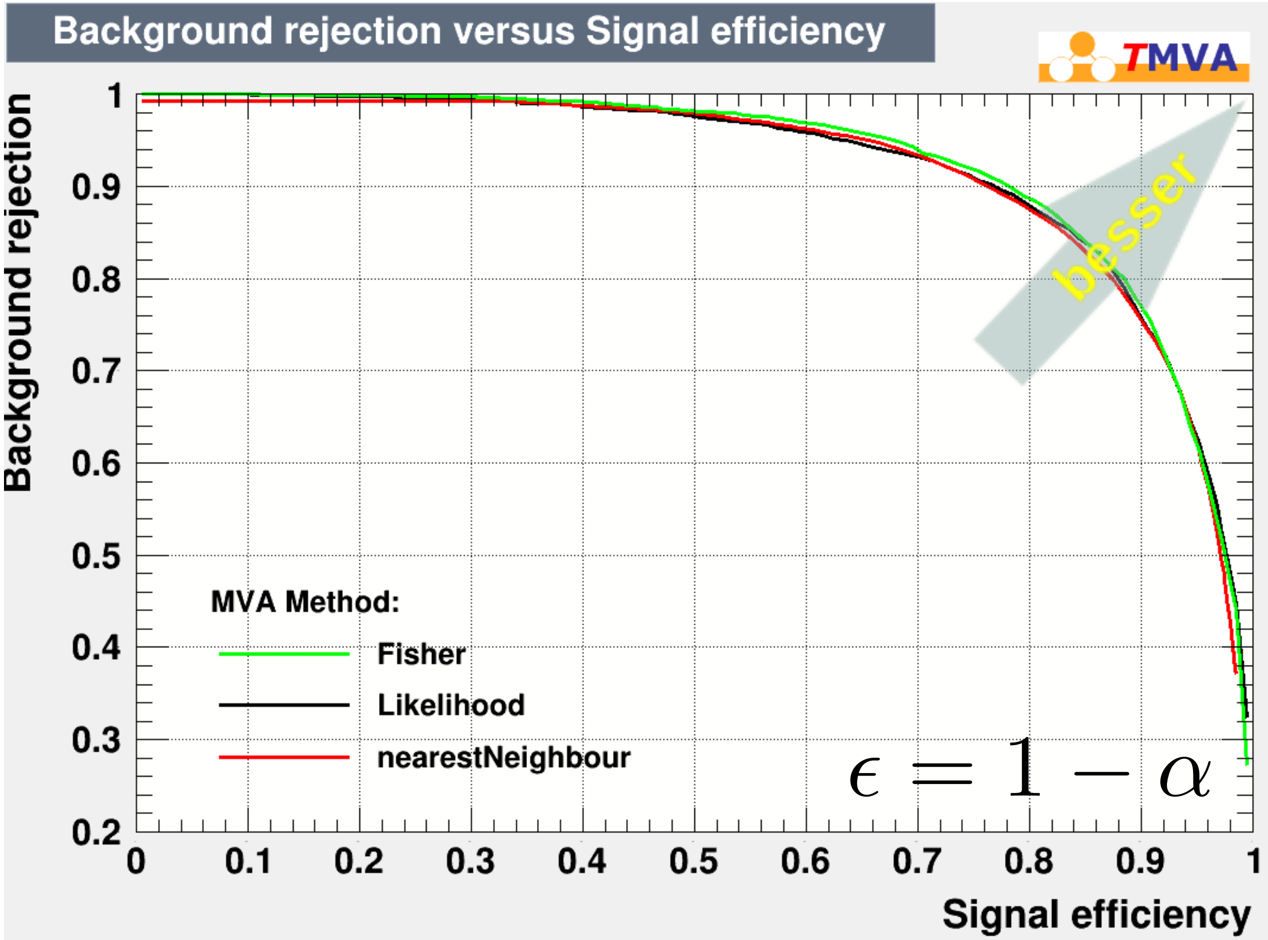
2. Festlegung des Signifikanz-Niveaus α
   → definiert Wert von t_0
   - Wahl von α abh. von Fragestellung: Hohe Reinheit *p* oder Effizienz *ε* ?

$$p = \frac{(1-\alpha)N_0}{(1-\alpha)N_0 + \beta N_1} \qquad \epsilon = 1 - \alpha$$

   - Aus der Wahl von α ergibt sich die Teststärke 1-β
   - Teststärke 1-β i.a. wichtiger als α: wenn zu klein, keine Trennung möglich.
3. Messung → Bestimmung des gemessenen p-Werts zur Entscheidung:
   - p-Wert liefert Wahrscheinlichkeit, dass Werte $t>t_m$ gemessen würden, falls H_0 wahr
   - p-Wert ist selbst eine Zufallsvariable

Effizienz-Reinheits-Kurve (ROC)

$$p = \frac{(1-\alpha)N_0}{(1-\alpha)N_0 + \beta N_1}$$



- Arbeitspunkt abhängig von Fragestellung
- Häufig wird daher die Area under Curve (“AUC”) als Vergleichskriterium für verschiedene Methoden verwendet

Neyman-Pearson Lemma

- Falls einfache Hypothesen, d.h. $f(x|H_i)$ vollständig bekannt, dann erreicht der Likelihood-Quotient

$$\lambda(x) = \frac{f(x|H_0)}{f(x|H_1)}$$

  optimale Trennung 1-β für vorgegebene Signifikanz α
- Äquivalent: Log-Likelihood Differenz:

$$q(x) = -2\ln\lambda(x) = 2(\ln f(x|H_1) - \ln f(x|H_0))$$

- Bemerkungen:
  - Finden der optimalen Prüfgröße: Klassifikation → folgende Vorlesungen
  - In der Praxis, verwende Monte Carlo Modelle, um die PDF für unterschiedliche Hypothesen zu generieren.
  - Neyman-Pearson Lemma gilt <u>nicht allgemein</u> für zusammengesetzte Hypothesen, d.h. Hypothesen mit freien Parametern, also z.B.: $f(x|H(\lambda_i,\mu_i))$ mit λ_i bekannt und μ_i frei)

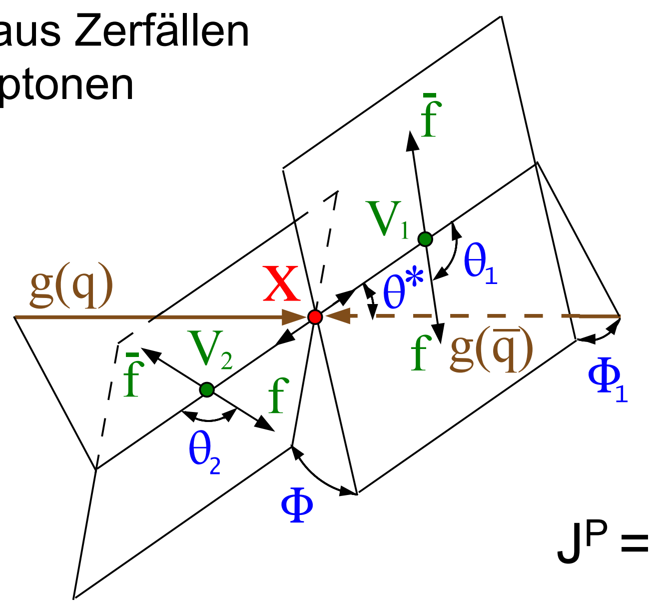
Beispiel: Eigenschaften des Higgs Bosons

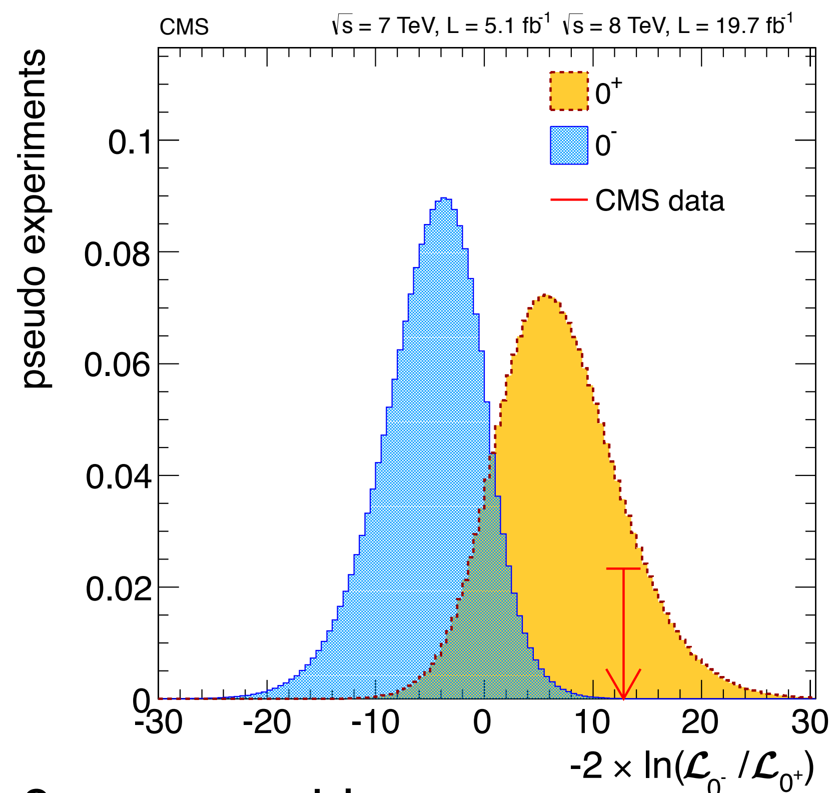
CMS HIG-13-002

- Ist das Higgs Boson ein skalares Teilchen?
  - Null-Hypothese: $J^P = 0^+$
  - Alternativ-Hypothese: z.B. $J^P = 0^-$
- Konstruktion einer skalaren Prüfgröße (Sample Likelihood aus sog. Matrixelement-Methode):

$$q = -2\ln(\mathcal{L}_{0-}/\mathcal{L}_{0+})$$

CP-Eigenschaften des Higgs aus Zerfällen in 4 Leptonen



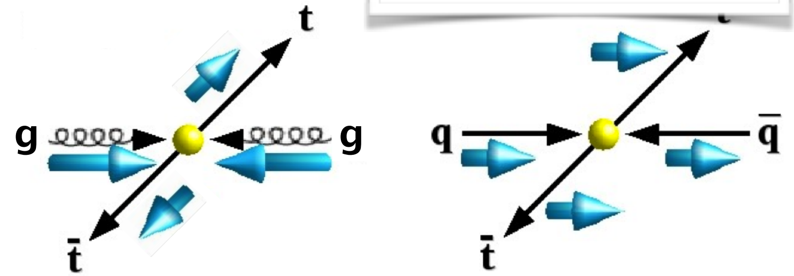


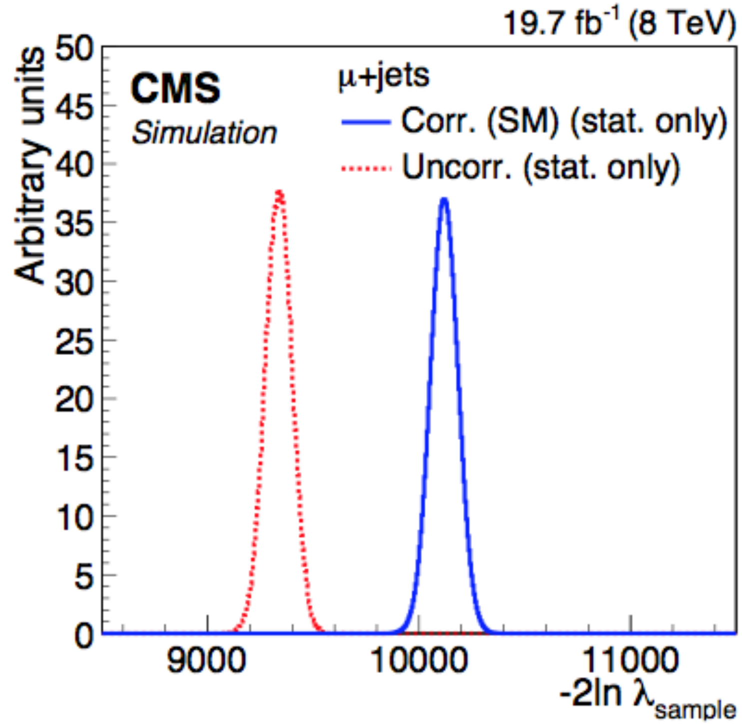
$J^P = 0^-$ mit 3.8σ ausgeschlossen

Beispiel: Spin-Korrelationen von Top-Quark Paaren

CMS TOP-13-015

- Neue Physik könnte relative Top-Quark Spins verändern → Winkelverteilungen
  - H_0: Spin-Korrelation gemäß Standard Modell (Spin 1/2 Teilchen)
  - H_1: keine Spin-Korrelation
- Konstruktion der Prüfgröße: Likelihood-Quotient → Log-Likelihood-Differenz
- Sample-Likelihoods aus Pseudo-Experimenten für H_0 und H_1
- Statistische Unsicherheiten: ~ Gaußsche Verteilungen → Zentraler Grenzwertsatz

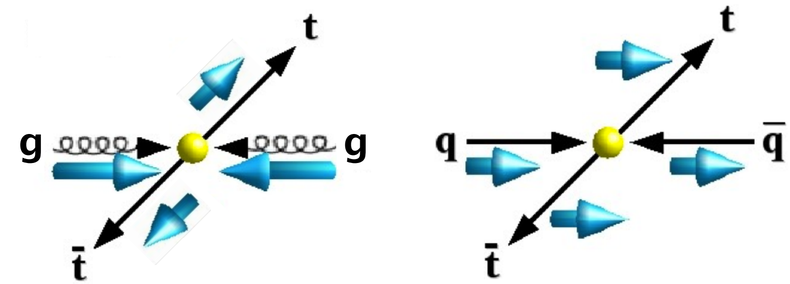


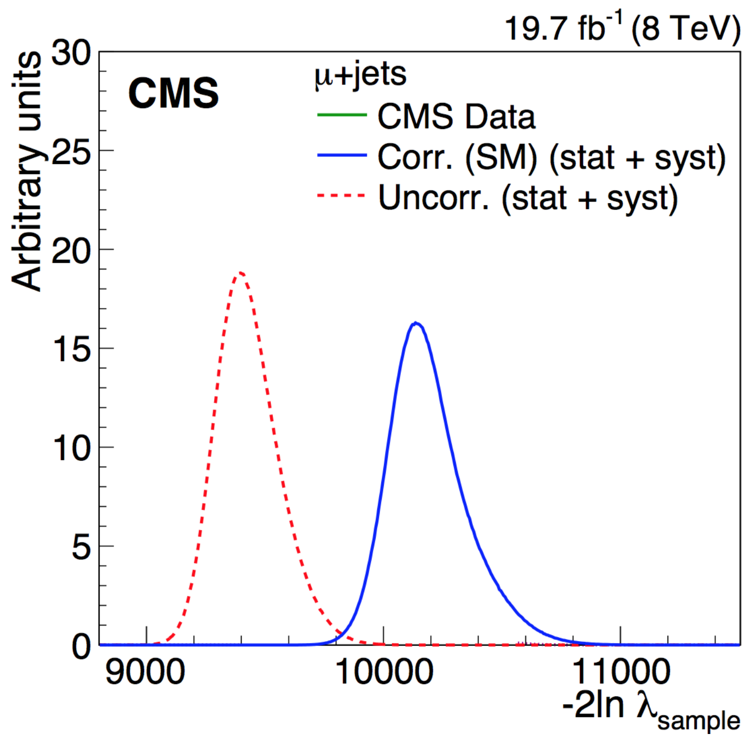


Abbildungen aus Doktor-Arbeit K.Beernaert, U Gent, 2015

Beispiel: Spin-Korrelationen von Top-Quark Paaren

- Neue Physik könnte relative Top-Quark Spins verändern → Winkelverteilungen
  - H_0: Spin-Korrelation gemäß Standard Modell (Spin 1/2 Teilchen)
  - H_1: keine Spin-Korrelation
- Konstruktion der Prüfgröße: Likelihood-Quotient → Log-Likelihood-Differenz
- Sample-Likelihoods aus Pseudo-Experimenten für H_0 und H_1
- Zusätzlich systematische Unsicherheiten: → Verbreiterung der PDF (nicht mehr Gauß-förmig)

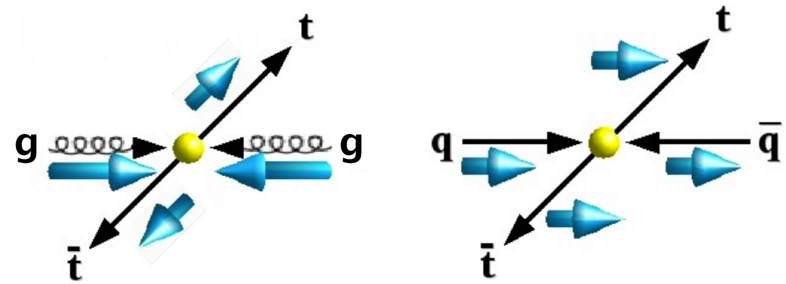


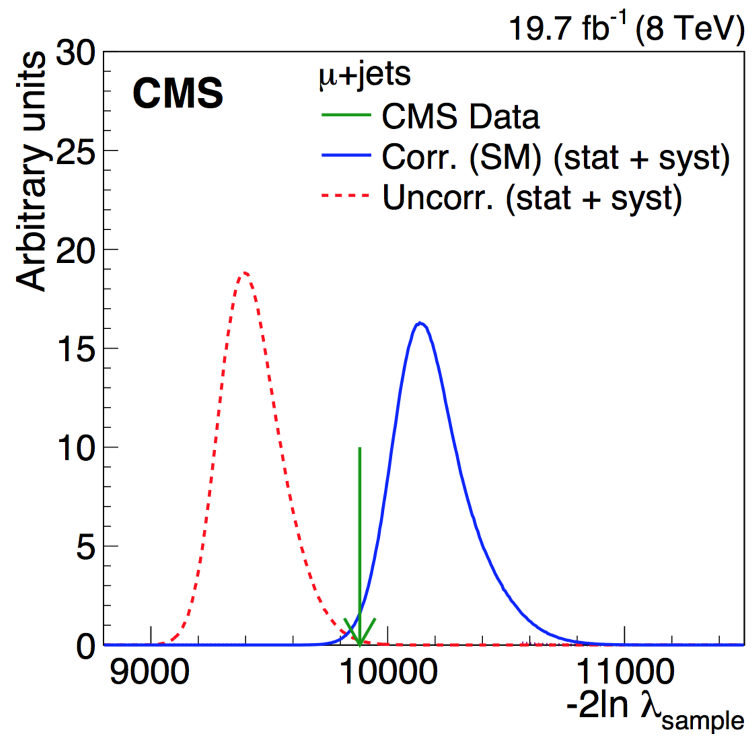


Abbildungen aus Doktor-Arbeit K.Beernaert, U Gent, 2015

Beispiel: Spin-Korrelationen von Top-Quark Paaren

- Neue Physik könnte relative Top-Quark Spins verändern → Winkelverteilungen
  - H_0: Spin-Korrelation gemäß Standard Modell (Spin 1/2 Teilchen)
  - H_1: keine Spin-Korrelation
- Konstruktion der Prüfgröße: Likelihood-Quotient → Log-Likelihood-Differenz
- Einzeichnen der Daten → Ergebnis:
  - 2.2σ (p-Wert 1.3%) konsistent mit Standard Model (H_0)
  - 2.9σ (p-Wert 0.2%) mit “unkorreliert” (H_1)





Abbildungen aus Doktor-Arbeit K.Beernaert, U Gent, 2015

Likelihood Quotient

- Verallgemeinerung auf nicht-einfache Tests:

$$\lambda(x) = \frac{\max_{\mu_j} \mathcal{L}(x_1, \ldots, x_n; \mu_1, \ldots, \mu_m)}{\max_{\mu_j \in \omega} \mathcal{L}(x_1, \ldots, x_n; \mu_1^{(\omega)}, \ldots, \mu_m^{(\omega)})} \quad \begin{matrix} H_1 \\ H_0 \end{matrix}$$

  - Zähler: Maximum im gesamten Parameter-Raum (Alternativ-Hypothese), d.h. μ_j freie Parameter
  - Nenner: Maximum im zulässigen Parameter-Raum ω für Null-Hypothese (z.B. einige Parameter μ_j festgelegt oder eingegrenzt), d.h. $\mu_j^{(\omega)}$ sind Parameter, so dass Null-Hypothese erfüllt

- Hypothese wird verworfen, falls $\lambda > \lambda_\alpha$ mit

$$\alpha = P(\lambda > \lambda_\alpha | H_0) = \int_{\lambda_\alpha}^{\infty} g(\lambda | H_0)\, d\lambda$$

z.B. Brandt, Abschnitt 8.6

Satz von Wilks

- Für sehr große Stichproben mit n Datenpunkten x_i, $n \rightarrow \infty$ gilt:

  Wird eine Grundgesamtheit durch die Wahrscheinlichkeitsdichte $f(x_i,\mu_i)$ beschrieben, und legt die Null-Hypothese $r=m-m^{(0)}$ der Parameter fest, so folgt die Testfunktion q = -2 ln λ einer χ^2 Verteilung mit r Freiheitsgraden.

  r = Unterschied in der Anzahl freier Parameter für H_1 und H_0

- Ausgangspunkt für sogenannte asymptotische Verfahren zur Berechnung des Likelihood-Quotienten λ(x). Aber: Anwendbarkeit muss im allgemeinen mit Monte Carlo Simulationen überprüft werden.

- Beispiel: Signal s über Untergrund b in Zählexperiment:

$$\Delta\chi^2 = -2\ln\lambda = -2\ln\left(\frac{\mathcal{L}(s+b)}{\mathcal{L}(b)}\right)$$

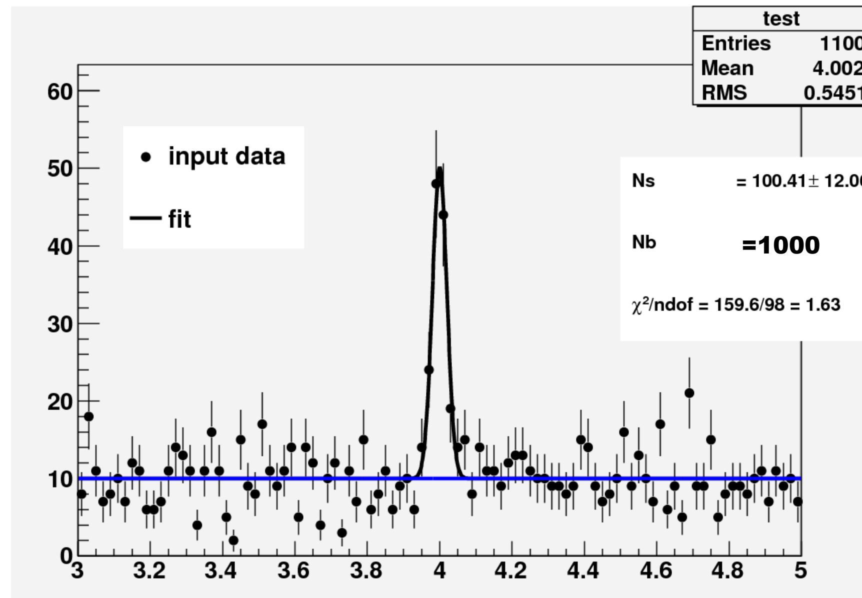
  Für $n=s+b$ groß folgt $\Delta\chi^2$ einer χ^2-Verteilung mit 1 Freiheitsgrad.

  Für gemessene Werte von s und b kann dann der p-Wert für die Signalsignifikanz berechnet werden.

S.S. Wilks, The large-sample distribution of the likelihood ratio for testing composite hypotheses. Ann. Math. Stat. **9**, 60–62 (1938)

Beispiel: Likelihood Quotient bei Zählexperiment

- Häufiger Fall: Suche nach Signal (s) über Untergrund (b)
- Verteilungsdichte für jedes m:
  $n(\mathrm{m}) = s(\mathrm{m}) + b(\mathrm{m})$



- H_0 (nur Untergrund): s=0
  - Anpassung mit freiem Parameter b
  - $\rightarrow \chi^2(b)$
- H_1: Signal-Hypothese $s \neq 0$
  - $b+s$(=Gauß mit festem m und σ)
  - $\rightarrow \chi^2(s+b)$

$$\Delta\chi^2 = -2\ln\lambda = -2\ln\left(\frac{\mathcal{L}(s+b)}{\mathcal{L}(b)}\right)$$ = -73 für diesen Fall

- Wilks: Falls H_0 (nur Untergrund) korrekt, dann folgt $\Delta\chi^2$ einer χ^2-Verteilung mit 1 Freiheitsgrad: p($-\chi^2 = 73$) = $2 \cdot 10^{-16}$, entspricht $z = 8.5\ \sigma$

Beispiel: Gauß-verteilte Stichprobe

- Hypothesentest: Hat eine Gauß-verteilte Stichprobe den Mittelwert μ_0 ?
- Verteilung der Stichprobe:

$$f(x_1, ..., x_N) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^N \exp\left[-\frac{1}{2\sigma^2}\sum_{j=1}^{N}(x_j - \mu)^2\right]$$

- Likelihood-Quotient:

$$\lambda = \exp\left[-\frac{1}{2\sigma^2}\left(\sum_{j=1}^{N}(x_j - \bar{x})^2 - \sum_{j=1}^{N}(x_j - \mu_0)^2\right)\right]$$

max $\mathcal{L}$ im gesamten Raum

μ_0 fest für H_0

- Mit $q = -2\ln\lambda$ folgt:

$$q = \frac{N}{\sigma^2}(\bar{x} - \mu_0)^2$$

Log-Likelihood Differenz

- Variablen Transformation (Nebenrechnung):

$$\Rightarrow \quad g(q) = \frac{1}{\sqrt{2\pi}}\, q^{-\frac{1}{2}} \exp(-\frac{q}{2})$$

χ^2 Verteilung für einen Freiheitsgrad

Wilks' Satz gilt hier für alle *N*

Brandt, Beispiel 8.5

Nebenrechnung

Durchführen einer Variablentransformation:

$$g(q) = \frac{d\bar{x}}{dq} f(\bar{x})$$

$$\text{mit } f(\bar{x}) = N\left(\bar{x}; \mu_0, \frac{\sigma}{\sqrt{N}}\right)$$

Umstellen des Zusammenhangs zwischen $\bar{x}$ und q:

$$\bar{x} = \sqrt{\frac{\sigma^2}{N}}\, q^{\frac{1}{2}} + \mu_0 \quad \Rightarrow \quad \frac{d\bar{x}}{dq} = \frac{1}{2}\sqrt{\frac{\sigma^2}{N}}\, q^{-\frac{1}{2}}$$

alles einsetzen:

$$\Rightarrow \quad g(q) = \frac{1}{\sqrt{2\pi}}\, q^{-\frac{1}{2}} \exp\left(-\frac{q}{2}\right)$$

Beispiel: Zählexperiment mit Poisson-Statistik

Karlsruher Institut für Technologie

- Beobachtung von n Ereignissen mit kleinem Signal s

$$P_0(n;b) = \frac{1}{n!} b^n e^{-b} \qquad P_1(n;s+b) = \frac{1}{n!}(s+b)^n e^{-(s+b)}$$

$$q = -2\ln\lambda = 2(n\ln(1+\frac{s}{b}) - s)$$

- Annahme b bekannt, dann ist *n* = *b* + *s*:

$$q = 2(b+s)\ln\left(1+\frac{s}{b}\right) - 2s$$

- Für *s* « *b* ist:

$$\sqrt{q} = s/\sqrt{b} + \mathcal{O}((s/b)^2)$$

- In Wilks Näherung gilt für einen einzelnen Freiheitsgrad *N*=1:
  Signifikanz des Signals *s*, ausgedrückt durch das Gaußsche Quantil *z*:

$$z = \sqrt{\Delta\chi^2} = \sqrt{q} = s/\sqrt{b}$$

Cowan, Cranmer, Gross and Vitelis "Asymptotic formulae for likelihood-based tests of new physics"
https://arxiv.org/abs/1007.1727

Nebenrechnung

- Für *s*≪*b*: Taylor-Entwicklung für den Logarithmus:

$$\ln(1+x) = \sum_{\nu=1}^{\infty} \frac{(-1)^{\nu+1}}{\nu} x^{\nu} \qquad \text{für } x \in (-1, 1]$$

- Einsetzen von *x* = *s*/*b* und Abbruch nach erstem Term:

$$\begin{aligned} q &= 2(b+s)\ln\left(1+\frac{s}{b}\right) - 2s \\ &\approx 2(b+s)\frac{s}{b} - 2s \\ &= 2s^2/b \end{aligned}$$

6.5 PROFILE-LIKELIHOOD QUOTIENT

CCGV: https://arxiv.org/abs/1007.1727
CMS+ATLAS: https://cds.cern.ch/record/1379837

Signalstärke μ

- Likelihood für Zählexperiment:

$$\mathcal{L}(\text{data}|\mu) = \text{Poisson}(\text{data}|\mu \cdot s + b)$$

- Produkt der Poisson-Likelihoods in Bin *i* n_i Ereignisse zu messen

$$\text{Poisson}(\text{data}|\mu \cdot s + b) = \prod_i \frac{(\mu \cdot s_i + b_i)^{n_i}}{n_i!} e^{-(\mu \cdot s_i + b_i)}$$

- Signalstärke μ:
  - $\mu=0$: H_0 Nur Untergrund
  - $\mu=1$: H_1 Erwartetes Signal

Störparameter θ

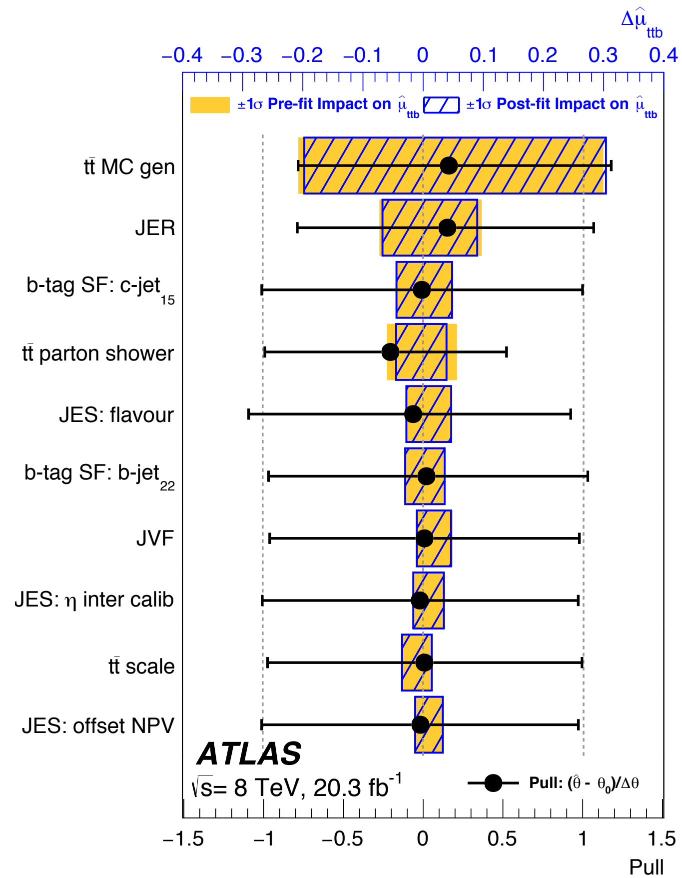
Störparameter θ beeinflussen Messung von *s* und *b*

- Likelihood für Zählexperiment:

$$\mathcal{L}(\text{data}|\mu) = \text{Poisson}(\text{data}|\mu \cdot s(\theta) + b(\theta))$$

- Störparameter θ: Parameter, die nicht primär von Interesse sind, die aber für die Bestimmung von Signal und Untergrund berücksichtigt werden müssen, also i.a. systematische Unsicherheiten.

- I.a. viele $\mathcal{O}(10-100)$ Störparameter



ATLAS: arXiv:1508.06868

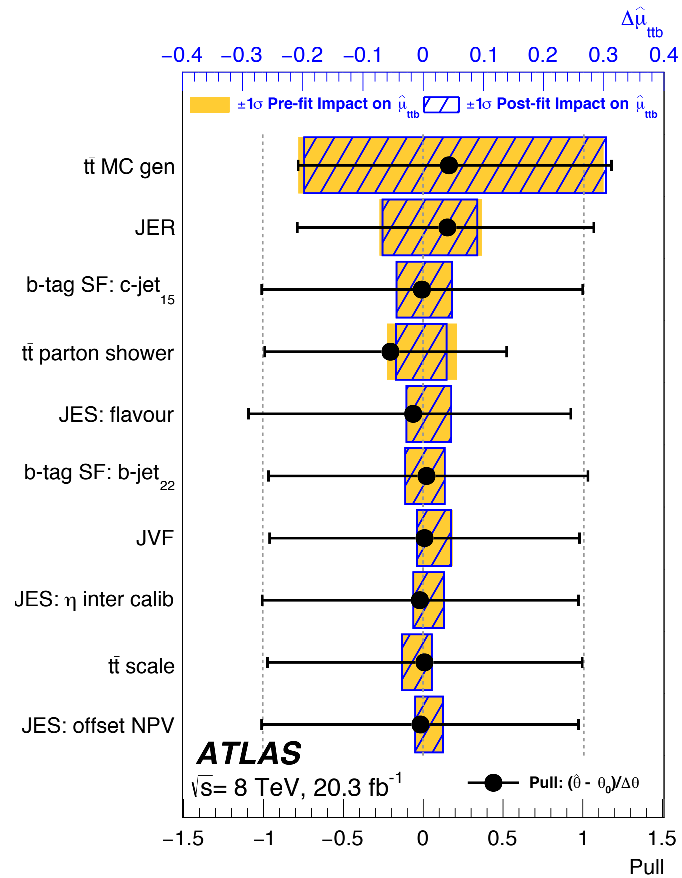
Randbedingungen (“Priors”)

- Likelihood für Zählexperiment:

Aus anderen Messungen bestimmte PDF (“Priors”)

$$\mathcal{L}(\mathrm{data}|\mu) = \mathrm{Poisson}(\mathrm{data}|\mu \cdot s(\theta) + b(\theta)) \cdot PDF(\theta)$$

- PDF(θ): PDF von Hilfsmessungen → verwendet als Randbedingungen (“Constraints”) für die frequentistische Likelihood der Hauptmessung.



ATLAS: arXiv:1508.06868

Profile-Likelihood Quotient

- Verwendet bei Suchen (d.h. Setzen oberer Grenzen oder Entdeckung) am LHC (z.B. Higgs-Boson) mit Signal-Hypothese μ.
- Profile-Likelihood: Bestimmung des <u>Intervalls</u> für (wahre) Signalstärke μ, für optimale Störparameter $\hat{\theta}_\mu$, normiert auf das <u>globale</u> Maximum der Likelihood
- Profile = Scan, d.h bestimme q_μ für alle μ

Prüfgröße

Signalstärke

Störparameter $\hat{\theta}_\mu$: maximiert $\mathcal{L}(\text{data}|\mu,\hat{\theta}_\mu)$ für μ

$$q_\mu = -2\ln\frac{\mathcal{L}(\text{data}|\mu,\hat{\Theta}_\mu)}{\mathcal{L}(\text{data}|\hat{\mu},\hat{\Theta})}\ ,\ \ 0 \le \hat{\mu} \le \mu$$

Best-fit Werte aller Parameter

erzwingt einseitige Grenze

- Für $\hat{\mu} < 0$: $\hat{\mu} = 0$ (d.h. $\hat{\mu} = \mu_{best}$ kann nicht negativ werden → μ positiv)
- Für $\hat{\mu} > \mu$ $q_\mu = 0$ (d.h. einseitige Grenze)

CCGV, Absch. 2.5 und CMS+ATLAS 2.1

Profile-Likelihood Quotient

Large Sample Approx.

$$q_\mu = -2 \ln \lambda(\mu) = -2\Delta \ln \mathcal{L}$$

$$\approx \frac{(\mu - \hat{\mu})^2}{\sigma^2}$$

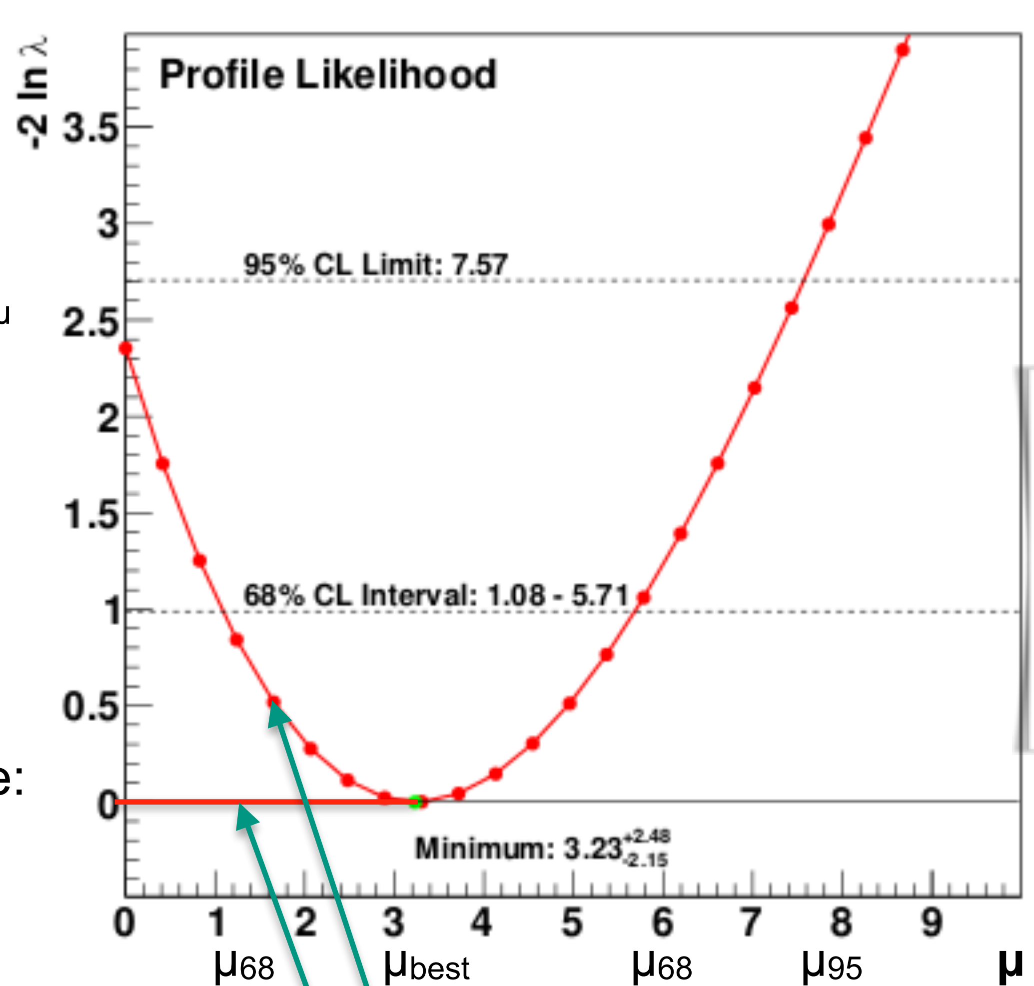
- Im Limit großer Statistik folgt q_μ einer χ^2-Verteilung (Parabel)
- Profile-Likelihood liefert alle relevanten Größen:
- Bester Schätzwert von μ:

  Minimum
- Ausschluss der Null-Hypothese:

  $q_{(\mu=0)} = z^2 = (\text{Signifikanz})^2$

  Hier: $z \sim \sqrt{2.4} \approx 1.5$
- Obere Grenze μ_{95}:

$$-2\Delta \ln \mathcal{L}(\mu_{95}) = 1.645^2 = 2.71$$



Allgemein: 2-seitiges Intervall

q_μ=0 für $\mu<\hat{\mu}$: 1-seitige Grenze

CCGV, Abschnitt 3.7

Profile-Likelihood Quotient

- Falls asymptotische Formeln nicht anwendbar sind, und zum Testen der asymptotischen Formeln: Bestimmung der erwarteten Verteilungen von q_μ für Untergrund (μ=0) und Signal (μ=1) durch Pseudo-Experimente.

Prüfgröße

Signalstärke

Störparameter $\hat{\theta}_\mu$: maximiert $\mathcal{L}(\text{data}|\mu,\hat{\theta}_\mu)$ für μ

$$q_\mu = -2\ln\frac{\mathcal{L}(\text{data}|\mu,\hat{\hat{\Theta}}_\mu)}{\mathcal{L}(\text{data}|\hat{\mu},\hat{\Theta})}\ ,\ \ 0 \le \hat{\mu} \le \mu$$

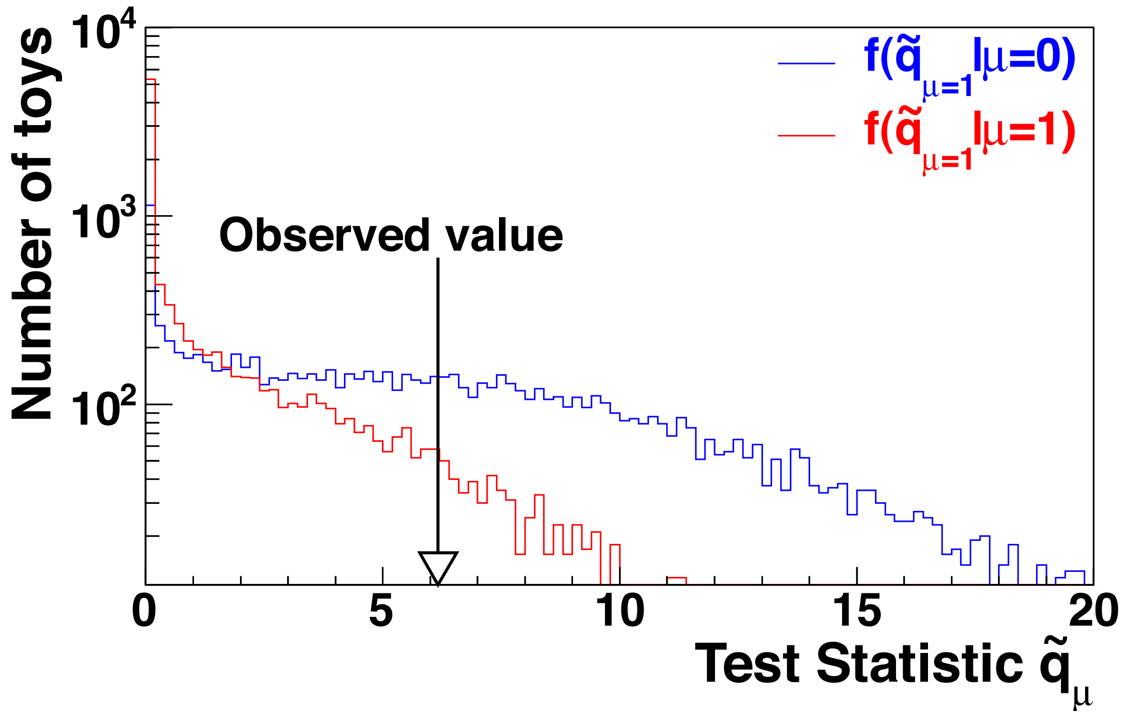
Best-fit Werte aller Parameter

erzwingt ein-seitige Grenze

CCGV, Absch. 2.5 und CMS+ATLAS 2.1

Profile-Likelihood Quotient

- Generiere Verteilungen aus Pseudo-Daten (oder Toy-MC) von q_μ für die Hypothesen Signal+Untergrund (μ=1) und Nur-Untergrund (μ=0).
- Verwende dafür gemessene Störparameter $\hat{\theta}_{\mu,obs}$



- Aus diesen PDF $f(q_\mu|\mu,\hat{\theta}_{\mu,obs})$ ermittele p-Werte für CL_{S+B} und CL_B

CMS+ATLAS, Figure 1

Profile-Likelihood Quotient mit CL$_S$

- Erinnerung: CL$_S$ - Rezept ist robust gegen Abwärtsfluktuationen des Untergrunds

$$CL_S = \frac{p_\mu}{1 - p_b}$$

- Konstruiere p-Werte für aus PDF für q_μ

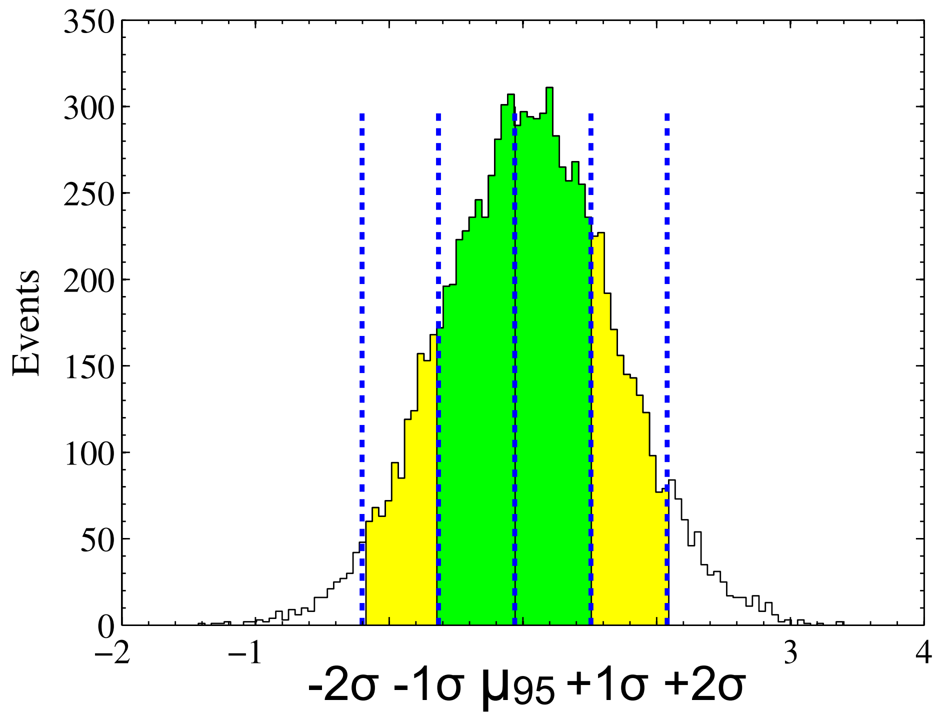
$$p_\mu \;=\; P(\,\tilde{q}_\mu \geq \tilde{q}_\mu^{obs}\,|\,\text{signal+background}) \;=\; \int_{\tilde{q}_\mu^{\text{obs}}}^{\infty} f(\tilde{q}_\mu|\mu, \hat{\theta}_\mu^{\text{obs}})\,d\tilde{q}_\mu$$

$$1 - p_b \;=\; P(\,\tilde{q}_\mu \geq \tilde{q}_\mu^{obs}\,|\,\text{background-only}) \;=\; \int_{q_0^{\text{obs}}}^{\infty} f(\tilde{q}_\mu|0, \hat{\theta}_0^{\text{obs}})\,d\tilde{q}_\mu$$

- CL$_S$ quantifiziert Übereinstimmung mit Signalhypothese:
  Falls für μ=1, CL$_S$ < α, dann ist das Signal mit 1-α CL$_S$ Vertrauens-Niveau ausgeschlossen.

Beispiel: Higgs-Entdeckung

- Typischerweise Angabe des Wertes μ_{95}, der mit 95% CL_S Vertrauens-Niveau ausgeschlossen ist.
- Pseudo-Experimente zur Bestimmung der Verteilung an der für die Untergrund-Hypothese erwarteten 95% CL Grenze,
d.h. des Medians und der Bereiche für ±1σ und ± 2σ um μ_{95}.

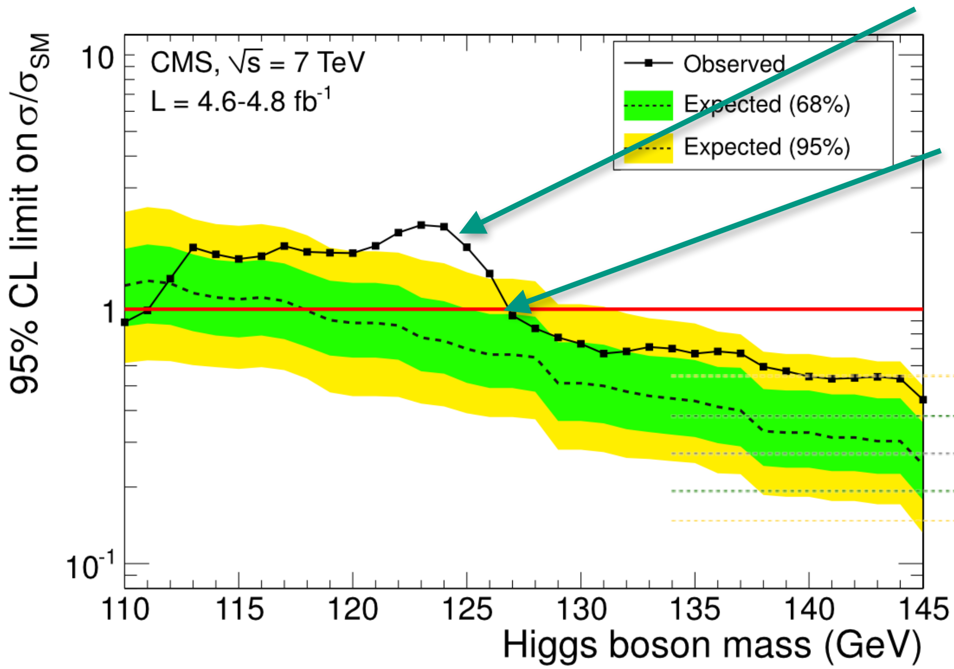


Beispiel: Higgs-Entdeckung

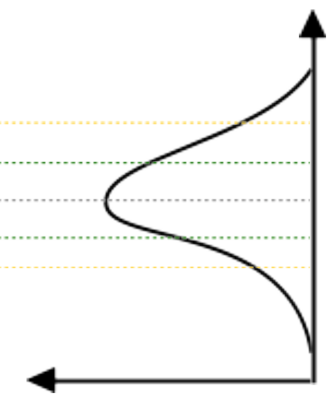
Stand Dezember 2011

- Zahl der erwarteten Ereignisse hängt von Masse des Higgs Bosons ab.

Überschuss bei einer Higgs-Masse von ~124 GeV



Higgs-Masse > 127 GeV wird mit 95% Vertrauens-Niveau ausgeschlossen



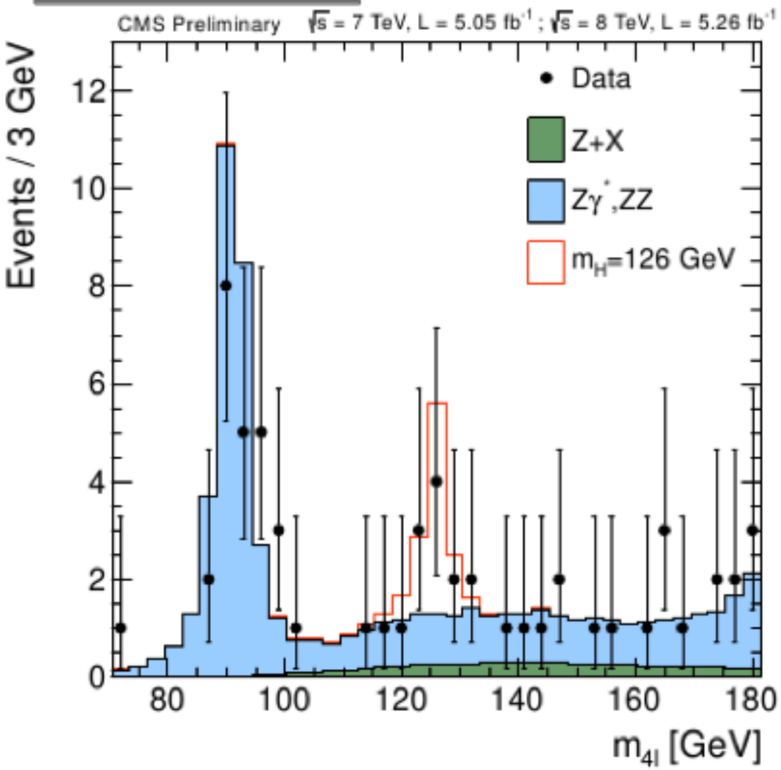
Gemessenes Limit für 133 GeV

- Verdeckte Analyse: Analyse und alle Kriterien wurden vor dem Anschauen der Daten festgelegt.

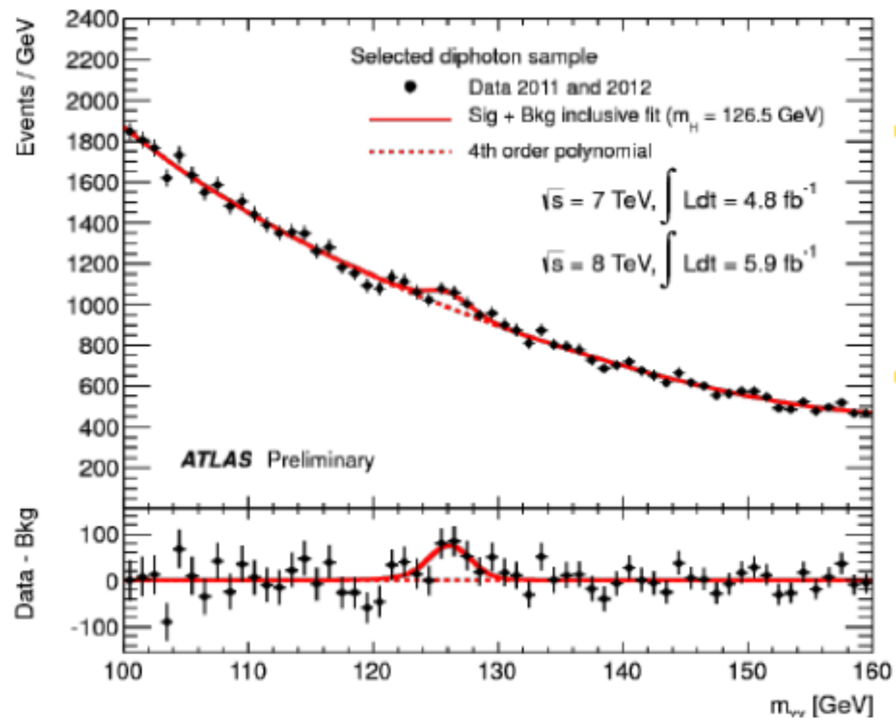
Beispiel: Higgs-Entdeckung

Stand Juli 2012

$H \to ZZ$



$H \to \gamma\gamma$



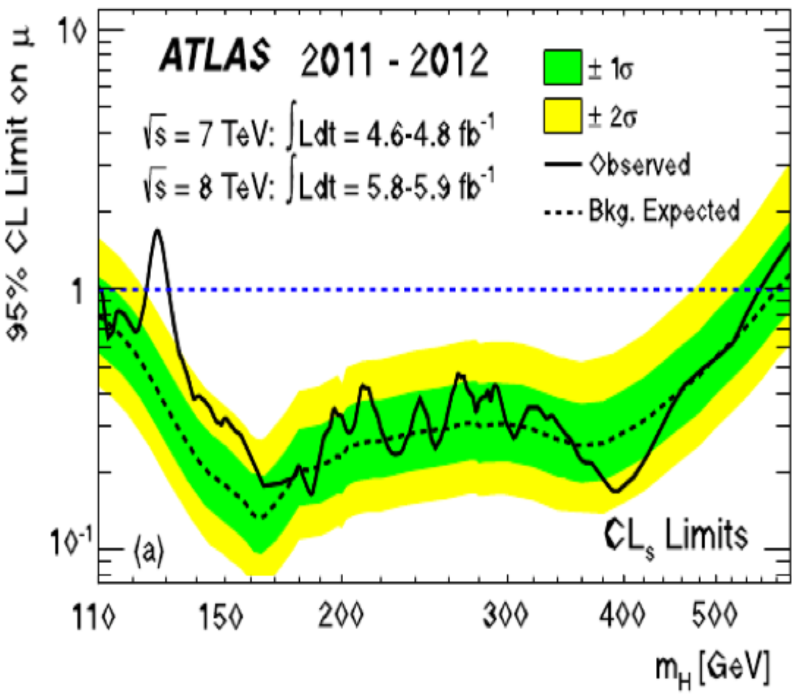
Ereignisüberschuss in ZZ und $\gamma\gamma$ bei ATLAS und CMS

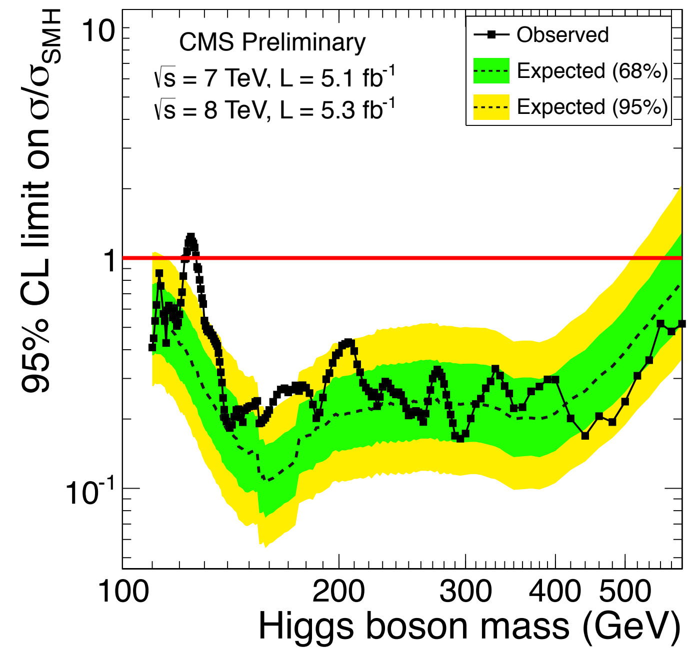
Beispiel: Higgs-Entdeckung

Stand Juli 2012

- 4.Juli 2012: Bekanntgabe der Entdeckung eines "neuen Bosons" am CERN

CMS-PAS-HIG-12-028



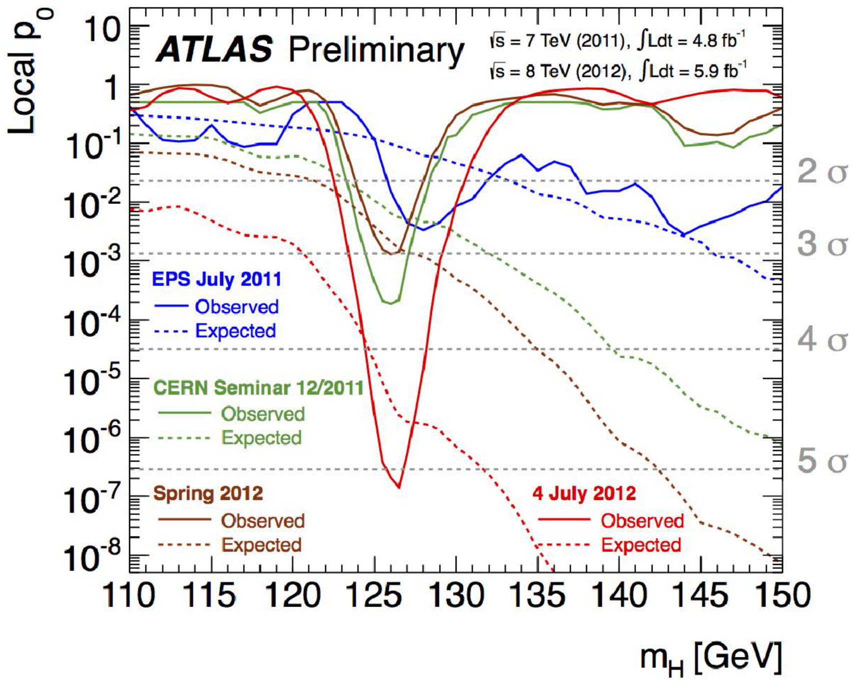


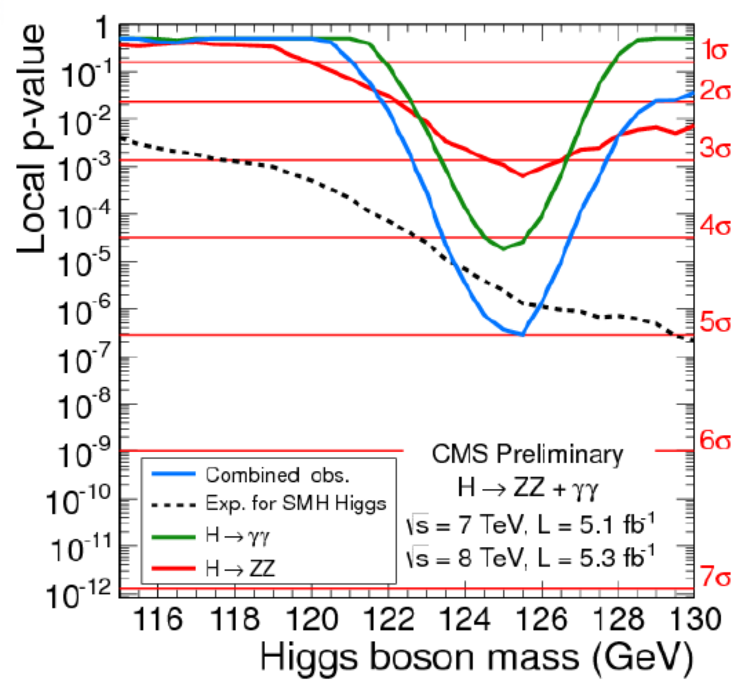
- Ausschluss von Signalen zwischen 131(128) GeV und 523(600) GeV

Beispiel: Higgs-Entdeckung

Stand Juli 2012

- 4.Juli 2012: Bekanntgabe der Entdeckung eines “neuen Bosons” am CERN





- Bestimmung der Signalsignifikanz und Bestimmung des <u>lokalen</u> p-Werts durch Vergleich mit Untergrund-Hypothese

$S_{ATLAS}=5.9\ \sigma$

$S_{CMS}=5.0\ \sigma$

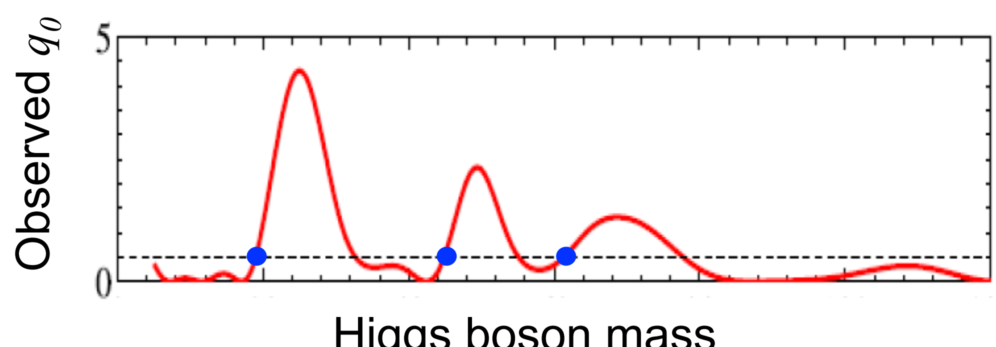
Look-Elsewhere Effekt

- Lokaler p-Wert: Wahrscheinlichkeit, dass die Beobachtung durch eine Aufwärts-Fluktuation des Untergrunds an einem vorgegebenen Wert einer Observablen (z.B. Masse m_H des Higgs-Bosons) verursacht ist.

- In globalen Suchen erhöht sich die Wahrscheinlichkeit mit der Größe des Suchbereichs → “Look-Elsewhere Effect”

$$\text{global p} = \text{Trialfactor} \times \text{local p}$$

- Der Trialfaktor ist i.a. proportional zur Breite der Verteilung und umgekehrt proportional zur Auflösung.
- Bestimmung:
  - Pseudoexperimente (CPU-aufwändig, da seltene Fluktuationen)
  - Oder Häufigkeit der Aufwärts-Fluktuationen in den Daten selbst



CMS+ATLAS, Figure 4

Zusammenfassung

- Intervallschätzung
  - Im nicht-Gaußischen Fall eine komplexe Angelegenheit.
  - Frequentistischer Ansatz: Betrachte alle Messungen â für alle möglichen wahren Werte a => Konfidenzintervall $[a_{min},a_{max}]_{\hat{a}}$ liefert Angabe der Häufigkeit bei wiederholten Experimenten den wahren Wert a in den abgeschätzten Intervallen zu finden.
  - Bayes: Betrachte aktuelles Experiment und A-Priori Wissen darüber.

- Profile-Likelihood Ratio:
  - Frequentistische Methode zur Bestimmung von oberen Grenzen bei Such-Analysen.
  - Verwendet Störparameter und Randbedingungen (Priors) aus Hilfsmessungen.
  - Anwendung der CL_S Methode: Robustheit gegen Untergrund-schwankungen.