

Moderne Methoden der Datenanalyse

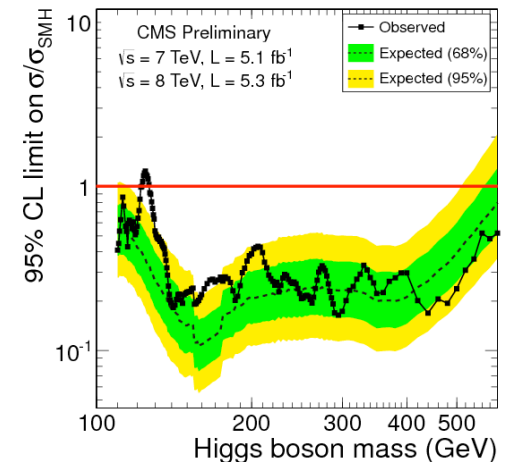
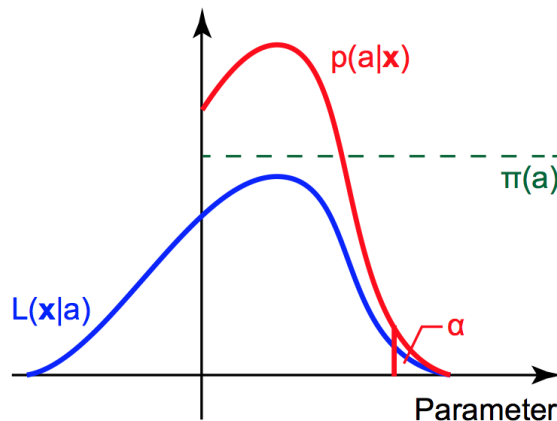
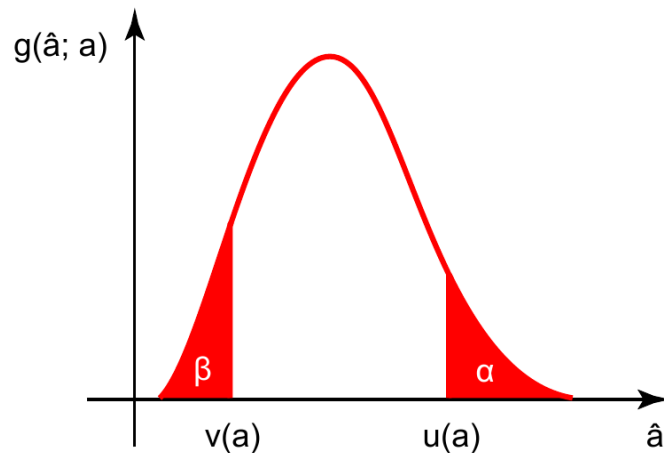
Konfidenz-Intervalle (1)

Andreas B. Meyer



13. Juni 2017

Fakultät für Physik
Institut für Experimentelle Kernphysik - IEKP



- 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

■ 3. Hypothesentests

- Null-Hypothese
- Signifikanz, Effizienz und Reinheit
- Goodness-of-Fit Tests
- Neyman-Pearson Lemma: Likelihood Quotient

■ 4. Parameterschätzung

- Maximum Likelihood
- Kleinste Quadrate (Spezialfall von ML)
- Anpassungen und Güte

■ 5. Optimierungsmethoden

- Suchmethoden
- Markov-Ketten
- Simuliertes Ausglühen

6. **KONFIDENZ-INTERVALLE**

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 FREQUEN-
TISTISCHER" 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, Goodness-of-fit, p-Wert
- Neyman-Pearson Lemma
- Beispiele

6.4
ERINNERUNG
HYPOTHESENTESTS

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

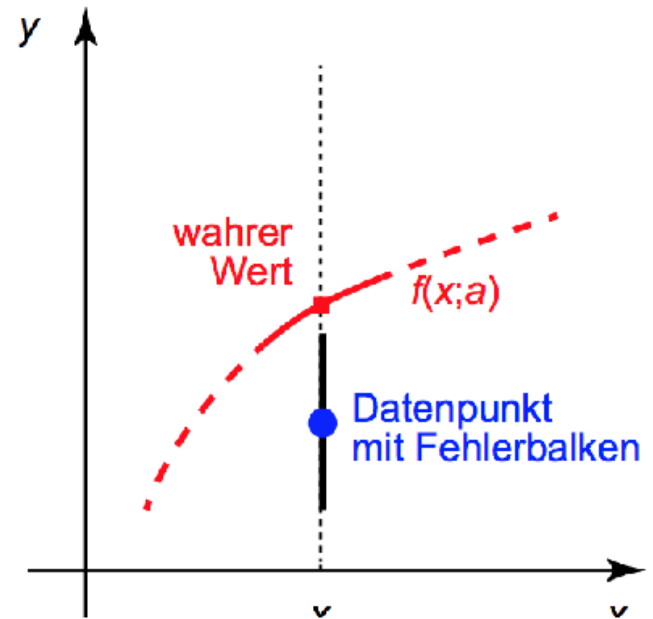
6.5
PROFILE-LIKELIHOOD
QUOTIENT

■ 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>

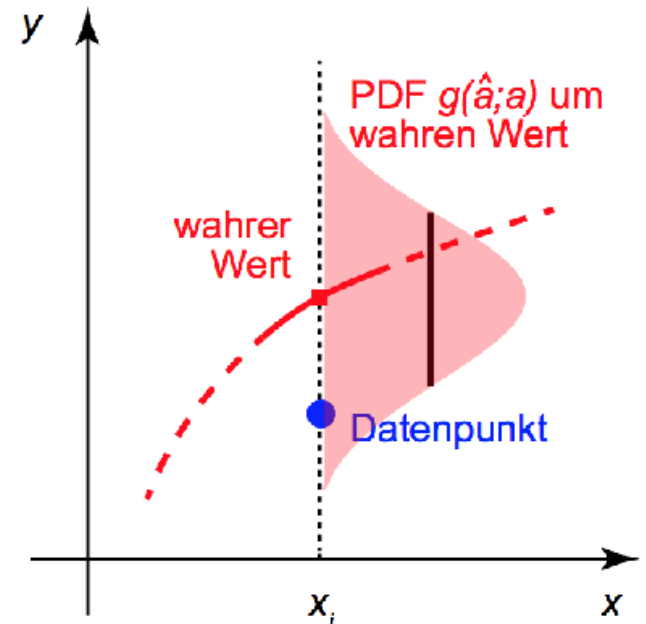
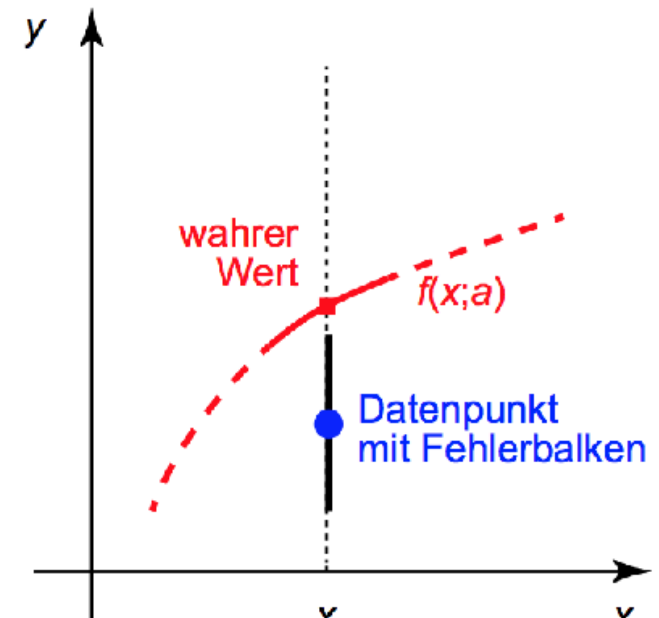
Intervall-Schätzung

- Bisher: Punktschätzung
- Übliche Darstellung von Messungen:
Schätzgröße mit Fehler: $\hat{a} \pm \sigma_a$
- Interpretation: Intervall $[\hat{a} - \sigma_a, \hat{a} + \sigma_a]$ deckt 68.3% Wahrscheinlichkeit ab.

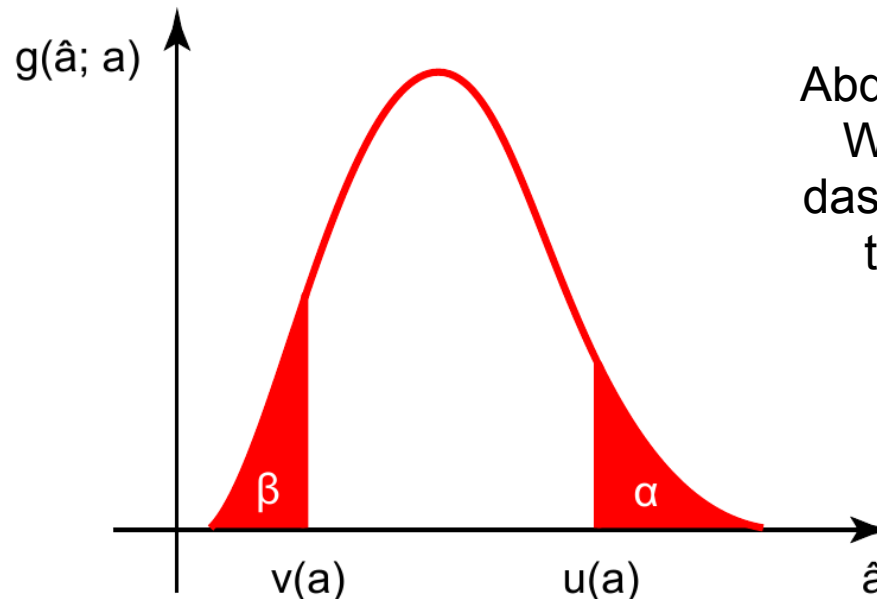


Intervall-Schätzung

- Bisher: Punktschätzung
- Übliche Darstellung von Messungen:
Schätzgröße mit Fehler: $\hat{a} \pm \sigma_a$
- Interpretation: Intervall $[\hat{a} - \sigma_a, \hat{a} + \sigma_a]$ deckt 68.3% Wahrscheinlichkeit ab.
- Eigentliche Bedeutung: Messergebnis $\hat{a}$ ist Zufallszahl verteilt um wahren Wert a . PDF $g(\hat{a}|a)$ ist Verteilungsdichte um den wahren Wert a .
- Äquivalent, wenn $g(\hat{a}|a)$ Gauß-verteilt. Das ist sehr häufig der Fall ($\rightarrow$ zentraler Grenzwertsatz) aber nicht immer!



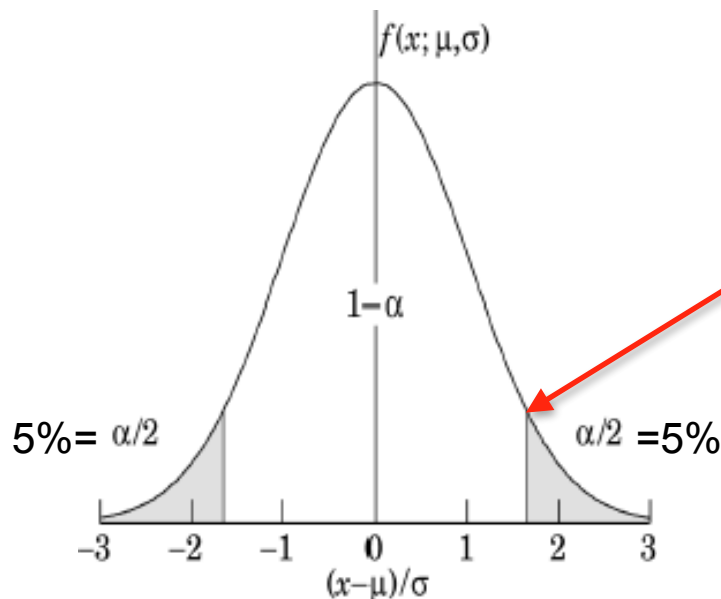
- Verwende Messung von $\hat{a}$, um Konfidenz-Intervall zu bestimmen, in dem der wahre Wert a mit vorgegebener Wahrscheinlichkeit liegt.
- Wahl einer vorgegebenen Wahrscheinlichkeit als “Konfidenz-Niveau”, auch “Vertrauensniveau” (engl. Confidence Level “CL”)
- Ein gewähltes Konfidenz-Niveau gibt angestrebte Abdeckung vor.
- Üblich sind $CL=68.3\%$, 90% oder 95%



Abdeckung (engl. Coverage):
Wahrscheinlichkeit $1-\alpha-\beta$,
dass wahrer Wert im Intervall
tatsächlich enthalten ist

Intervall-Schätzung

- Üblich sind CL=68.3%, 90% oder 95%, oder entsprechende Quantile der Gauß-Verteilung
- Zweiseitige Intervalle: Messungen
- Einseitige Intervalle: Obere oder untere Schranken, z.B. bei Suchen



Root: TMath::ChisquareQuantile

$\alpha_{\text{zweiseitig}}(\delta)$

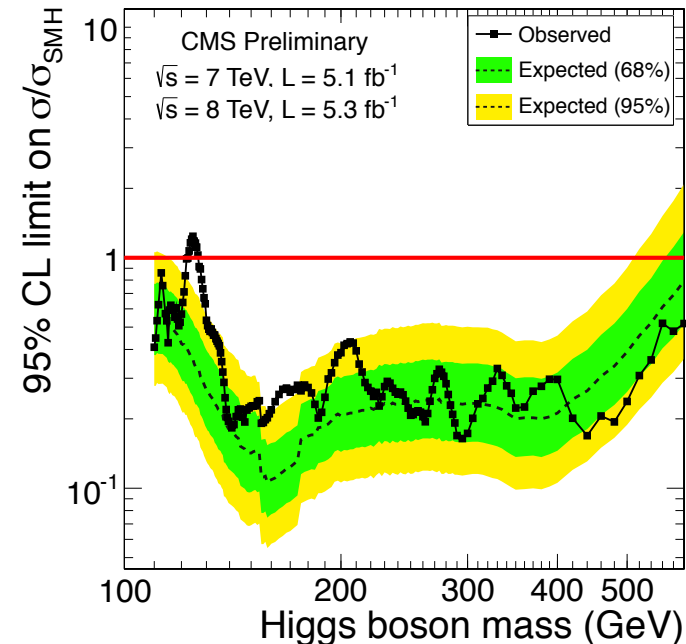
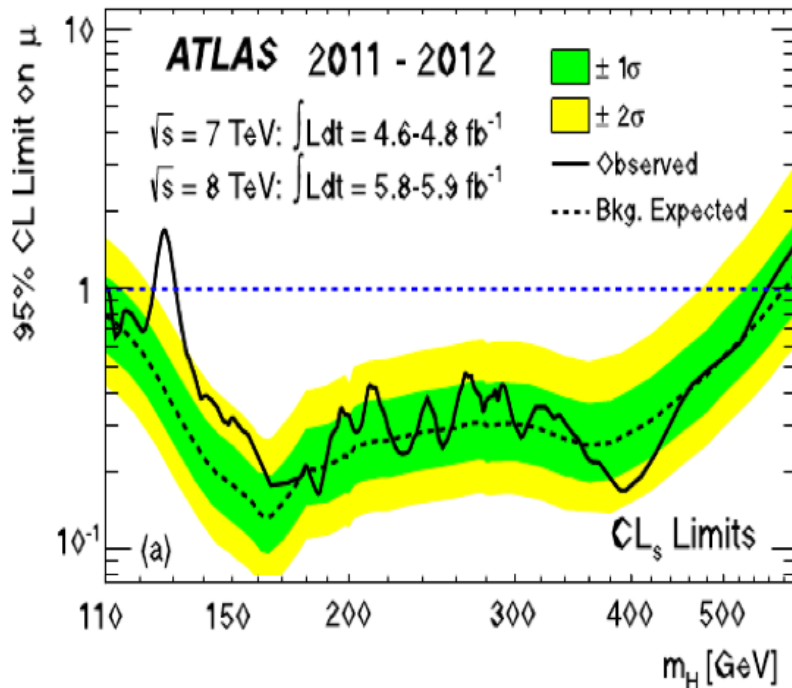
$1 - \alpha$	α	δ
0.683	0.317	1σ
0.90	0.1	1.65σ
0.95	0.05	1.96σ
0.9545	0.0455	2σ
0.9973	0.0027	3σ
	5.7×10^{-7}	5σ

$$\alpha_{\text{einseitig}}(\delta) = \alpha_{\text{zweiseitig}}(\delta)/2$$

Beispiel Obere Schranke: Higgs-Entdeckung

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

Daten Stand: Juli 2012

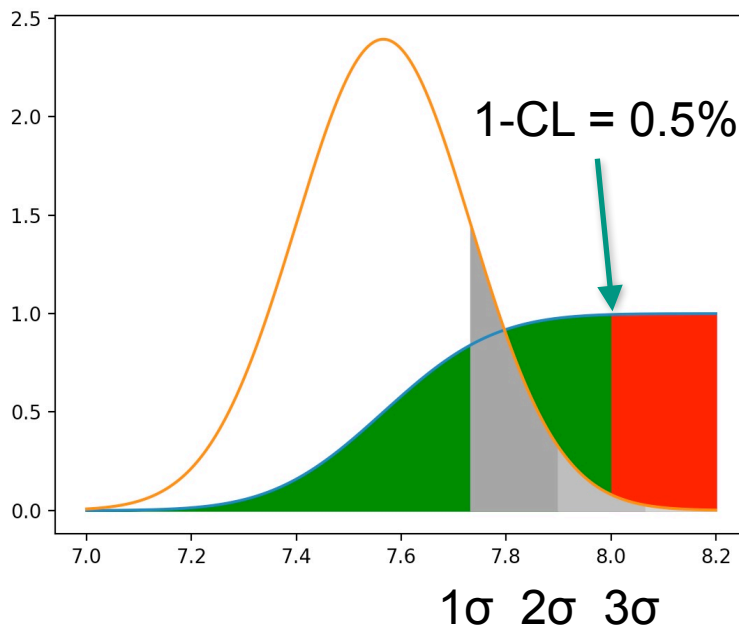


HIG-12-028

Ziel am Ende der nächsten Vorlesung: Verständnis dieser Abbildungen

Beispiel Untere Schranke: Arbeitsweg

- Ein Angestellter muss jeden Tag im 8 Uhr morgens bei der Arbeit sein. Der Arbeitsweg dauert im Mittel 30 Minuten, mit einer hier als Gaußförmig angenommenen Streuung σ von 10 Minuten.
- Wann muss er das Haus verlassen, um nur einmal im Jahr zu spät zu sein? Bei 200 Arbeitstagen also 0.5% der Fälle



$$1 - CL = \int_{t_{\text{Ankunft}}}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$$

$$t_{\text{Ankunft}} = 8\text{h}00$$

$$\sigma = 10 \text{ min}$$

$$1-CL = 0.5\%$$

Einseitiges Limit:

0.5% entspricht etwa 2.6σ

$$\begin{aligned} t_{\text{Abfahrt}} &= 8\text{h}00 - 30 \text{ min} - 2.6 \cdot 10 \text{ min} \\ &= 7\text{h}04 \end{aligned}$$

Beispiel: [arbeitsweg.py](#)

- Fehlerintervalle sollten das Ergebnis der Messung so objektiv wie möglich beschreiben (z.B. unabhängig von Modellannahmen).
- Der gesuchte Parameter sollte mit einer bestimmten, wählbaren Wahrscheinlichkeit enthalten sein (Konfidenz-Niveau und Abdeckung).
- Das Intervall sollte die Streuung der Messwerte des Parameters widerspiegeln.
- Definition sollte konsistent und effizient sein: Im Grenzfall genauer Messungen wird der wahre Wert erreicht, und genauere Messungen führen zu kleineren Fehlern.
- Schlussfolgerungen, die (explizite) (Modell-)annahmen enthalten, sollten möglich sein. Mittelung von Messwerten und Fehlerfortpflanzung.
- Das Verfahren sollte möglichst einfach sein.

■ Frequentistische Sichtweise: Es existiert ein fester wahrer Wert a

- Wahre Werte sind wahr, haben also keine Unsicherheit
- Wahrscheinlichkeitsbegriff bezogen auf prinzipiell messbare Häufigkeit (z.B. QM)

Das Konfidenz-Intervall zu einem Konfidenz-Niveau $CL = p\%$ deckt den wahren Wert a in $p\%$ aller Fälle ab.

- Wahrscheinlichkeitsaussagen nur über das Intervall, nicht den wahren Wert
- “Inversion” durch Neyman-Konstruktion, Abdeckung “by design”

■ Bayes'sche Sichtweise

- Der wahre Wert a hängt von Voraussetzungen ab.
- “Prior” beschreibt degree of belief, dass a bestimmte Werte annehmen kann.
- Wahrer Wert besitzt eine Unsicherheit, die von der Messung abhängig ist

Konstruiere Wahrscheinlichkeitsdichte für unbekannten Wert a

$$f(a|\hat{a}) \propto \mathcal{L}(\hat{a}|a) \cdot \pi(a)$$

- Posterior $f(a|\hat{a})$ benötigt Likelihood-Funktion $\mathcal{L}(\hat{a}|a)$ und Prior $\pi(a)$
- Abdeckung muss explizit überprüft werden (z.B. mit toy-MC)

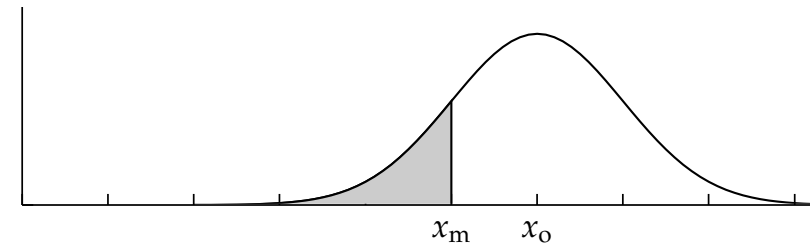
6.1

FREQUENTISTISCHER ANSATZ

Klassisches (frequentistisches) Konfidenz-Intervall

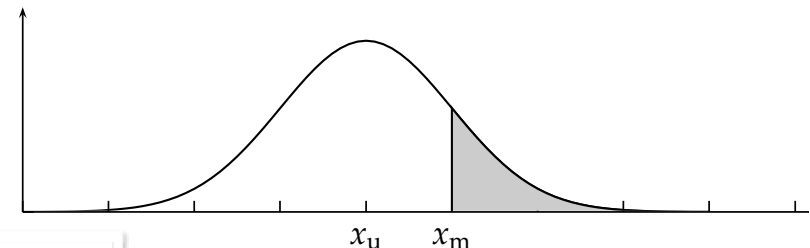
- Messung einer Größe x mit einem Apparat bekannter, Gauß-förmiger Auflösung. Messung liefert Wert x_m .
- Konstruktion eines zweiseitigen, symmetrischen Intervalls $[x_{unten}, x_{oben}]$ aus verschobenen Verteilungen
- 68% Konfidenz-Intervall soll den wahren Wert in 68% der Fälle enthalten.

- Obere Grenze x_{oben} :
Für angenommenen wahren Wert x_{oben} ist die Wahrscheinlichkeit einen Wert x_m oder kleiner zu messen: $(1 - 0.68)/2 = 16\%$



$$\alpha_0 = \int_{-\infty}^{x_m} \frac{1}{\sqrt{2\pi}\sigma} \exp(-(x - x_o)^2 / 2\sigma^2) dx = 16\%$$

- Untere Grenze x_{unten} :
Für angenommenen wahren Wert x_{unten} ist die Wahrscheinlichkeit einen Wert x_m oder größer zu messen: $(1 - 0.68)/2 = 16\%$



Aus Blobel/Lohrmann, Abschnitt 6.7

Poisson-Verteilung

■ Beispiel: 90% Konfidenz-Intervall (symmetrisch, zweiseitig) bei $n = 9$ beobachteten Ereignissen

■ Poisson-Wahrscheinlichkeit: $p(n|\mu) = e^{-\mu} \mu^n / n!$

■ Untere Grenze μ_{unten} :

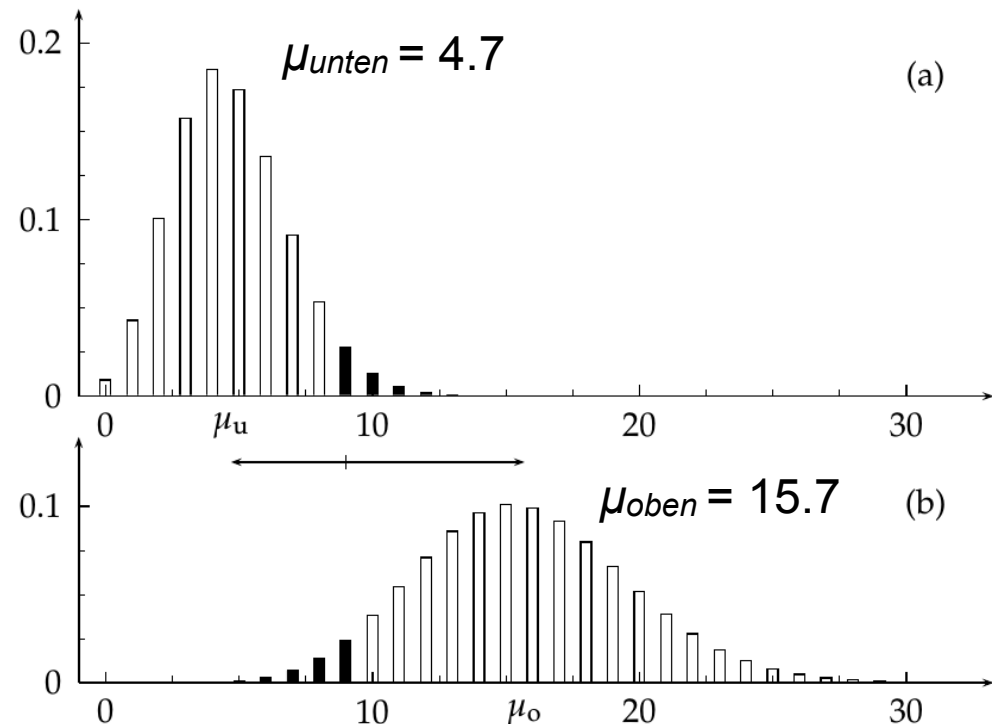
$$\begin{aligned}\alpha &= \sum_{k=n}^{\infty} P(k|\mu_u) = 5\% \\ &= 1 - \sum_{k=0}^{n-1} P(k|\mu_u)\end{aligned}$$

■ Obere Grenze μ_{oben} :

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

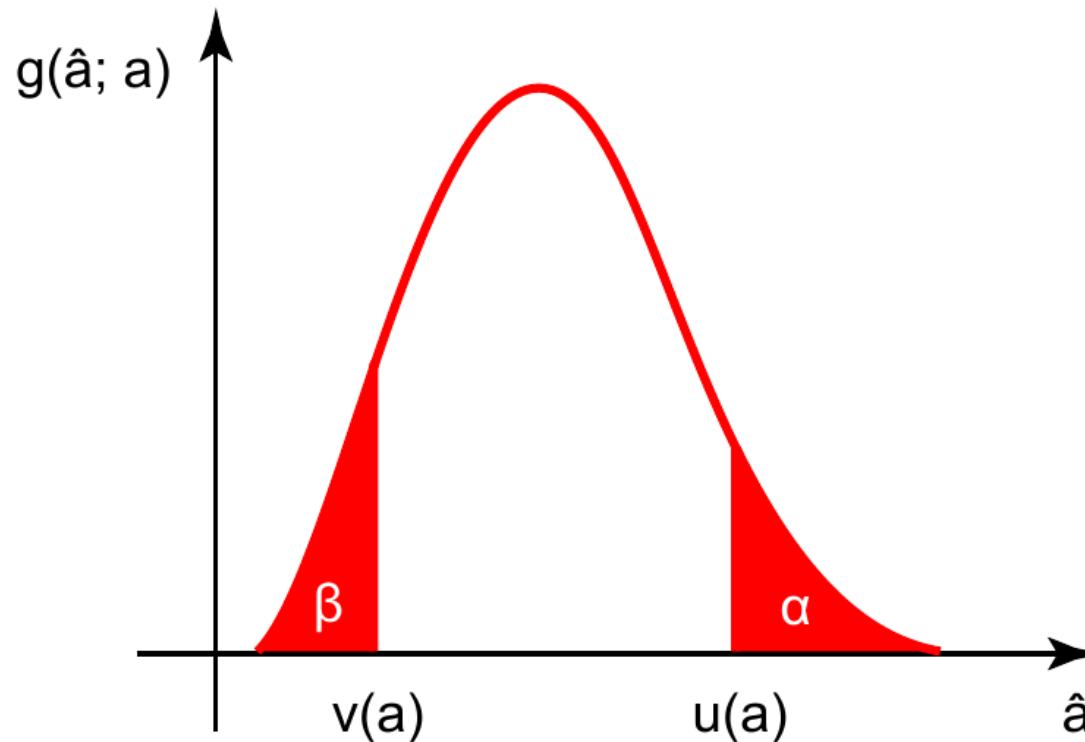
■ Lösung: $\mu \in [4.7, 15.7]$ @ 90 CL

■ Vergleich Gauß-Verteilung: $9 \pm 1.645 \cdot \sqrt{9} = 9 \pm 4.9$ (1.645 σ entspricht $\alpha=5\%$)



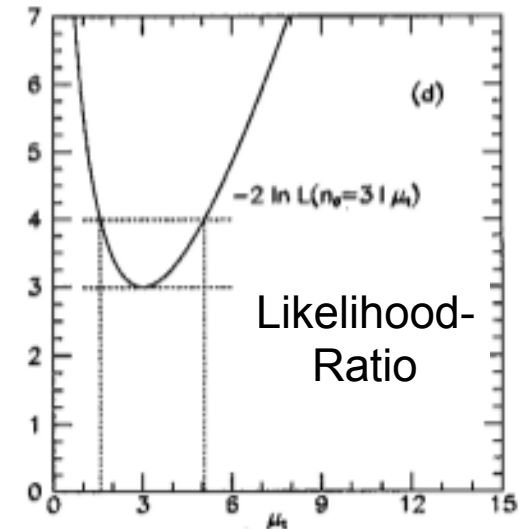
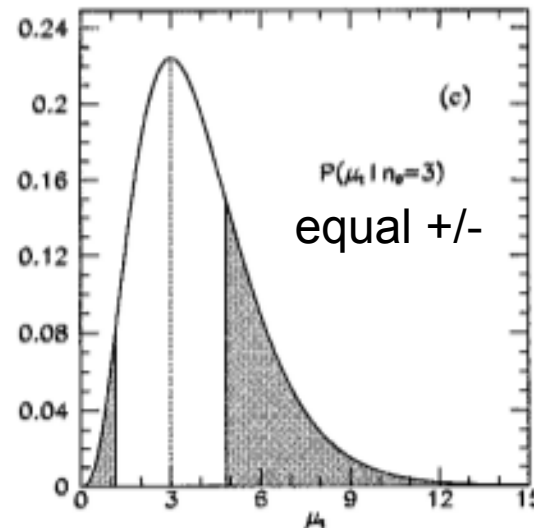
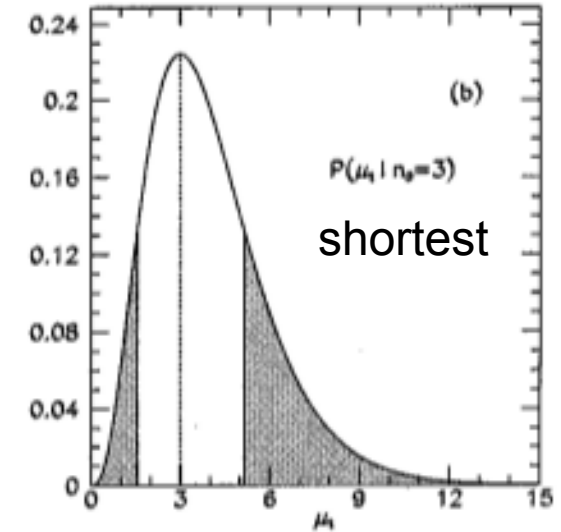
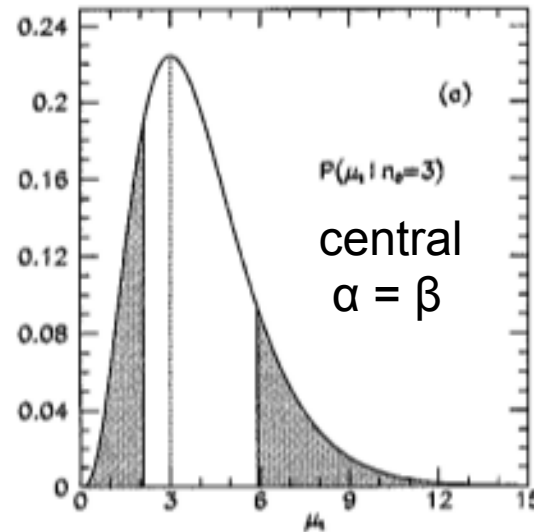
Typen von Konfidenz-Intervallen

- Allgemein sind verschiedene Intervall-Typen gebräuchlich



- Bisher: “symmetrisch” oder “zentral” im Sinne von $\alpha = \beta$
- Betrachtet: 2-seitig 68% und 90%, 1-seitig 0.5% (Arbeitsweg)

Typen von Konfidenz-Intervallen (Ordnungsprinzip)



Die Entscheidung, welche Definition für die Intervallgrenzen verwendet wird, nennt man "Ordnungsprinzip"

Aus: R. Cousins, „Why isn't every Physicist a Bayesian?“

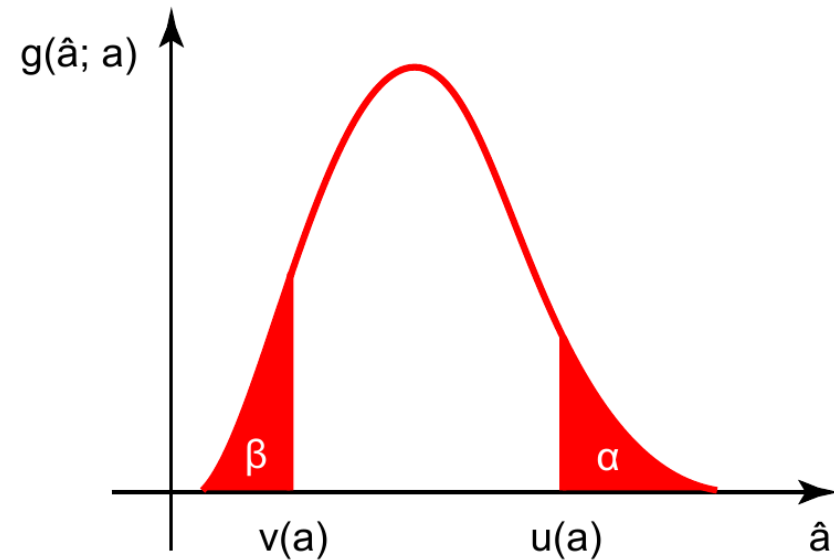
- Systematische Vorschrift aus frequentistischer (“klassischer”) Statistik: Konstruktion von Konfidenz-Intervallen mit vorgegebenem Konfidenz-Niveau (J.Neyman, 1937)
- Konstruiere Intervall $I=[a_{\min}, a_{\max}]$ im Parameterraum aus Daten, so dass wahrer Wert von Parameter a in einem festen Anteil (z.B. 68.3%) wiederholter Experimente in $I=[a_{\min}, a_{\max}]$ liegen würde.
- Aussage über die Abdeckung des wahren Werts durch das Intervall. Neyman-Konstruktion garantiert (per Vorschrift) exakte Abdeckung für kontinuierliche Verteilungen (Überabdeckung für diskrete Verteilungen)

Neyman-Konstruktion

- PDF $g(\hat{a}|a)$ der Schätzwerte bekannt

- Für vorgegebenes Konfidenz-Niveau, (d.h. Wahrscheinlichkeiten α und β) und festen wahren Wert a , finde $v(a)$ und $u(a)$, so dass

$$\beta = \int_{-\infty}^{v(a)} g(\hat{a}; a) d\hat{a} \quad \text{und} \quad \alpha = \int_{u(a)}^{\infty} g(\hat{a}; a) d\hat{a}$$



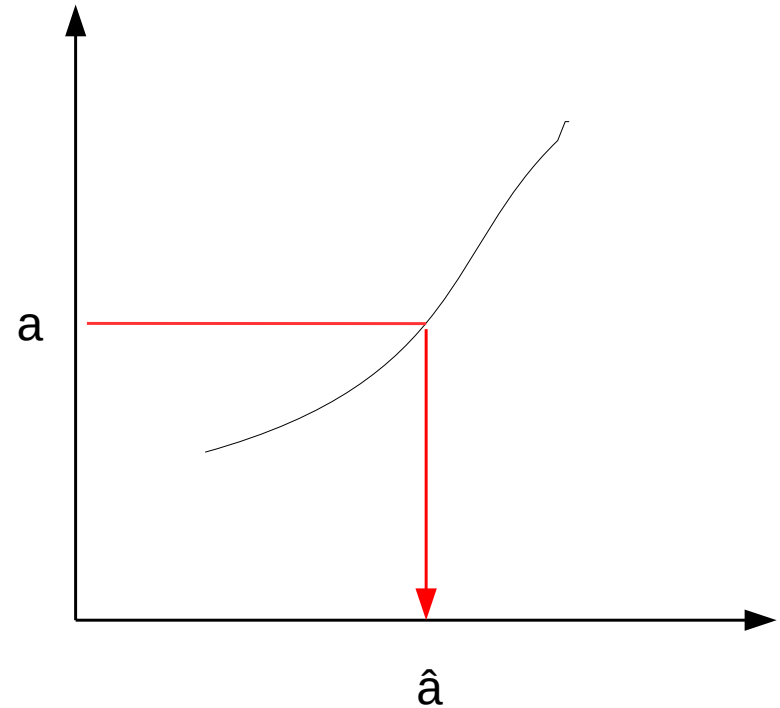
- Messung von $\hat{a}$: Um Konfidenz-Intervalls $[a_{\min}(\hat{a}), a_{\max}(\hat{a})]$ zu bestimmen, benötige Umkehrfunktionen von $v(a)$ und $u(a)$
- Nutze aus: Wahrscheinlichkeitserhaltung bei Variablen-Transformation

$$\alpha = P(\hat{a} > u(a)) = P(a > a_{\max}(\hat{a}))$$

$$\beta = P(\hat{a} < v(a)) = P(a < a_{\min}(\hat{a}))$$

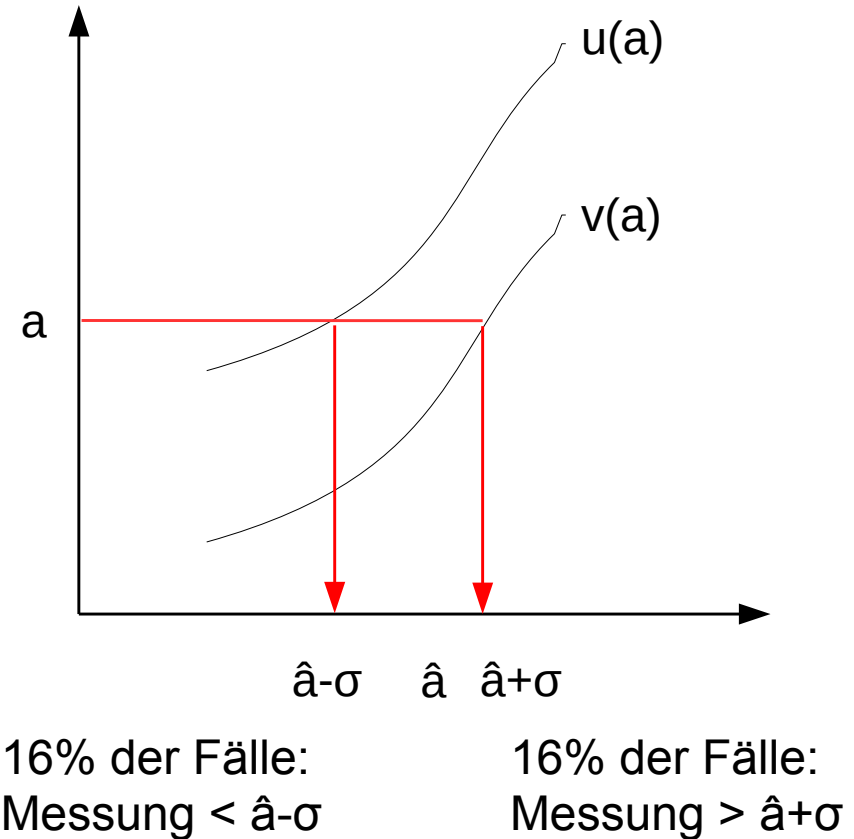
Neyman-Konstruktion

- Für einen gegebenen wahren Wert a gibt es eine Messung $\hat{a}=a_{\text{obs}}$



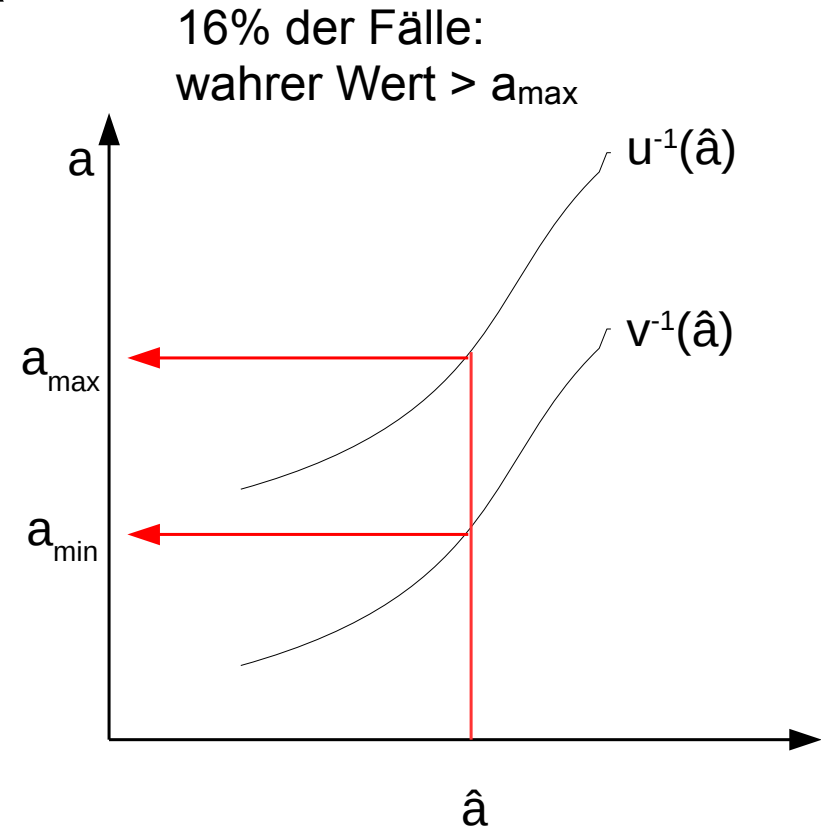
Neyman-Konstruktion

- Für einen gegebenen wahren Wert a gibt es eine Messung $\hat{a}=a_{\text{obs}}$
- mit Unsicherheit σ
 $\hat{a}-\sigma$ und $\hat{a}+\sigma$ sind Funktionen von a ,
nämlich $u(a)$ und $v(a)$



Neyman-Konstruktion

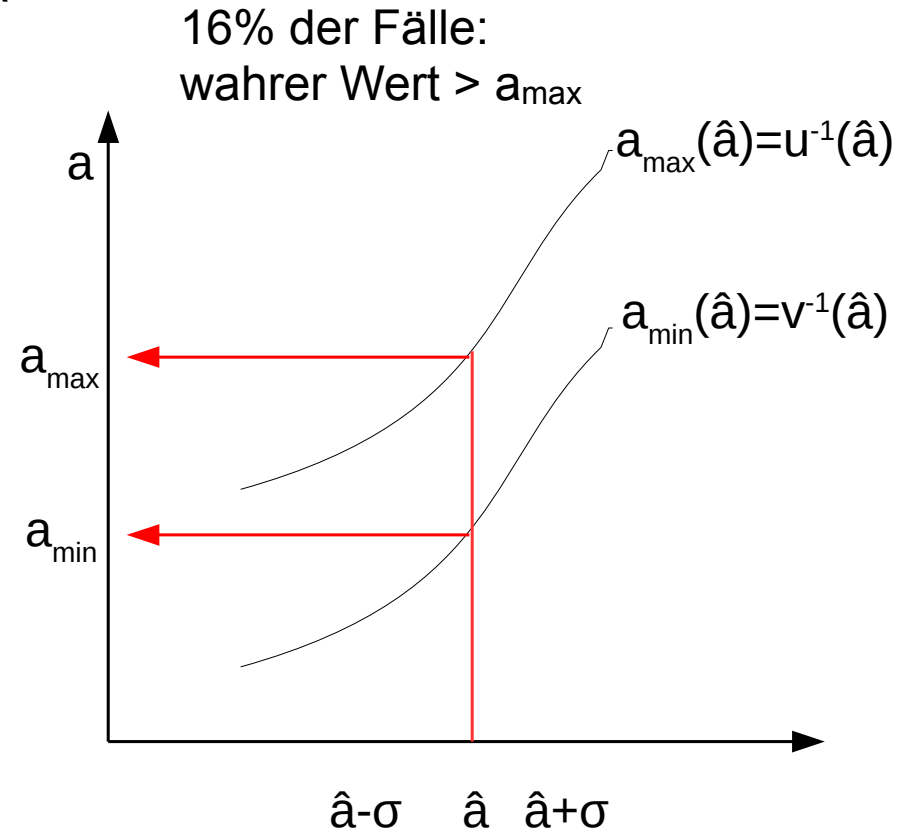
- Für einen gegebenen wahren Wert a gibt es eine Messung $\hat{a}=a_{\text{obs}}$
- mit Unsicherheit σ
 $\hat{a}-\sigma$ und $\hat{a}+\sigma$ sind Funktionen von a , nämlich $u(a)$ und $v(a)$
- Für eine konkrete Messung $\hat{a}$ kann ein Konfidenzintervall konstruiert werden.
- Die Funktionen $v^{-1}(\hat{a})$ und $u^{-1}(\hat{a})$ sind die Umkehrfunktionen von $u(a)$ und $v(a)$



16% der Fälle:
wahrer Wert $< a_{\text{min}}$

Neyman-Konstruktion

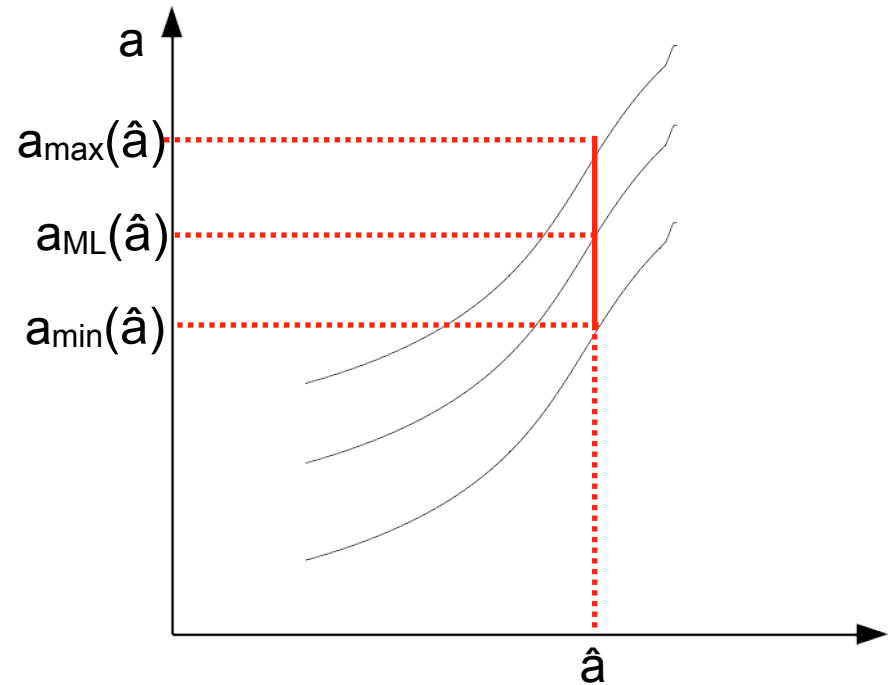
- Für einen gegebenen wahren Wert a gibt es eine Messung $\hat{a}=a_{\text{obs}}$
- mit Unsicherheit σ
 $\hat{a}-\sigma$ und $\hat{a}+\sigma$ sind Funktionen von a , nämlich $u(a)$ und $v(a)$
- Für eine konkrete Messung $\hat{a}$ kann ein Konfidenzintervall konstruiert werden.
- Die Funktionen $a_{\min}(\hat{a})$ und $a_{\max}(\hat{a})$ sind der Konfidenz-Gürtel
- D.h. der Gürtel wird horizontal für angenommene wahre Werte a konstruiert. Für eine konkrete Messung $\hat{a}$ wird das Konfidenz-Intervall vertikal abgelesen



16% der Fälle:
wahrer Wert $< a_{\min}$

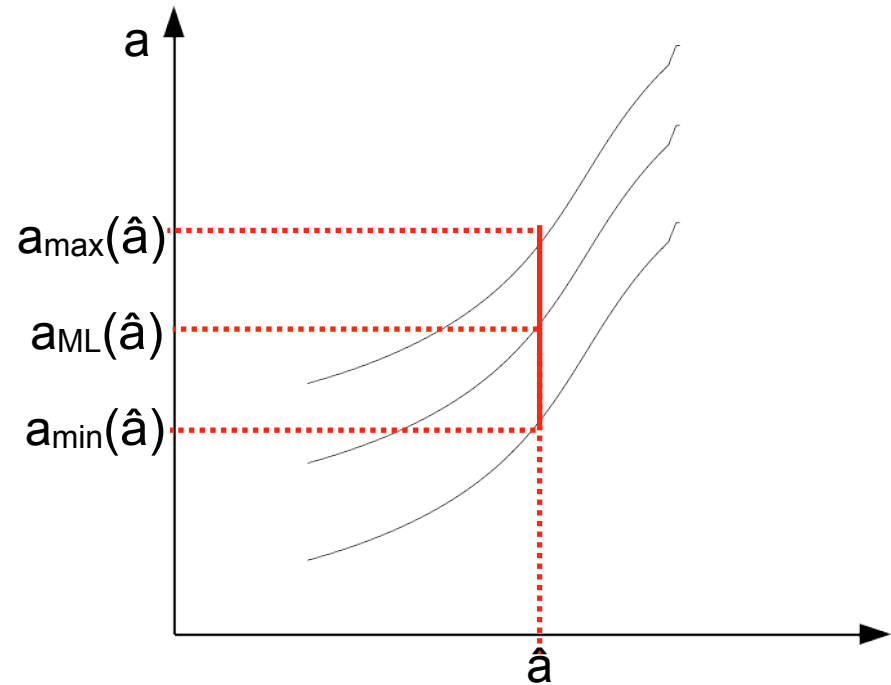
Neyman-Konstruktion

- Die Funktionen $u(a)$ und $v(a)$ beschreiben die Intervall-Grenzen für ein Konfidenz-Niveau als Funktion eines hypothetischen wahren Werts
- Umgekehrt liefern die Umkehrfunktionen von $u(a)$ und $v(a)$ die Intervall-Grenzen $a_{\max}(\hat{a}) = u^{-1}(\hat{a})$ und $a_{\min}(\hat{a}) = v^{-1}(\hat{a})$ ausgehend vom Messwert $\hat{a}$
- Bedingung: $u(a)$ und $v(a)$ monoton und Umkehrfunktionen existieren
- Hier ist jetzt auch der wahrscheinlichste Wert für a eingezeichnet. Streng frequentistisch hat a aber keine Unsicherheit



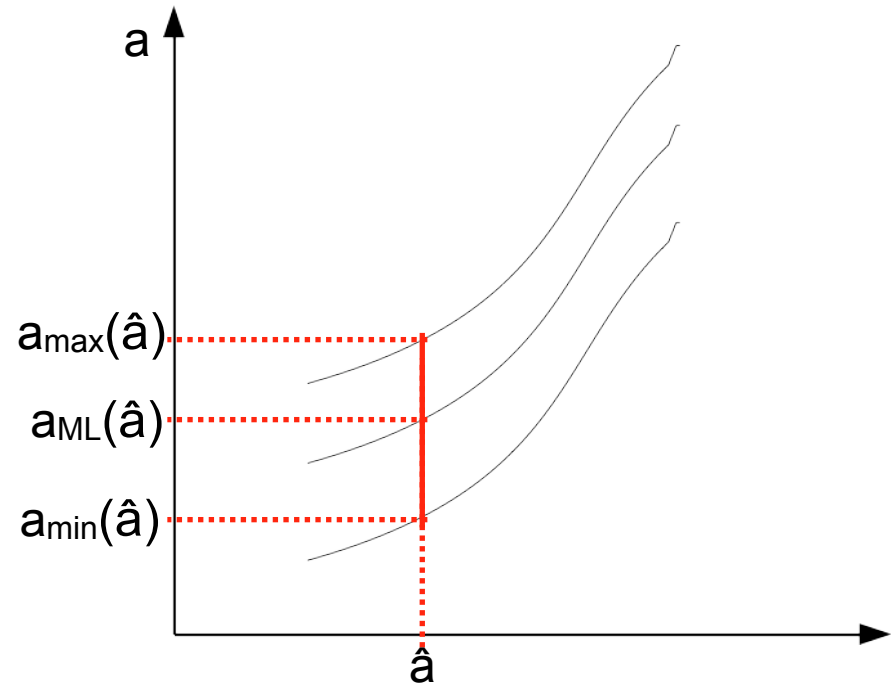
Neyman-Konstruktion

- Die Funktionen $u(a)$ und $v(a)$ beschreiben die Intervall-Grenzen für ein Konfidenz-Niveau als Funktion eines hypothetischen wahren Werts
- Umgekehrt liefern die Umkehrfunktionen von $u(a)$ und $v(a)$ die Intervall-Grenzen $a_{\max}(\hat{a}) = u^{-1}(\hat{a})$ und $a_{\min}(\hat{a}) = v^{-1}(\hat{a})$ ausgehend vom Messwert $\hat{a}$
- Bedingung: $u(a)$ und $v(a)$ monoton und Umkehrfunktionen existieren
- Hier ist jetzt auch der wahrscheinlichste Wert für a eingezeichnet. Streng frequentistisch hat a aber keine Unsicherheit



Neyman-Konstruktion

- Die Funktionen $u(a)$ und $v(a)$ beschreiben die Intervall-Grenzen für ein Konfidenz-Niveau als Funktion eines hypothetischen wahren Werts
- Umgekehrt liefern die Umkehrfunktionen von $u(a)$ und $v(a)$ die Intervall-Grenzen $a_{\max}(\hat{a}) = u^{-1}(\hat{a})$ und $a_{\min}(\hat{a}) = v^{-1}(\hat{a})$ ausgehend vom Messwert $\hat{a}$
- Bedingung: $u(a)$ und $v(a)$ monoton und Umkehrfunktionen existieren
- Hier ist jetzt auch der wahrscheinlichste Wert für a eingezeichnet. Streng frequentistisch hat a aber keine Unsicherheit

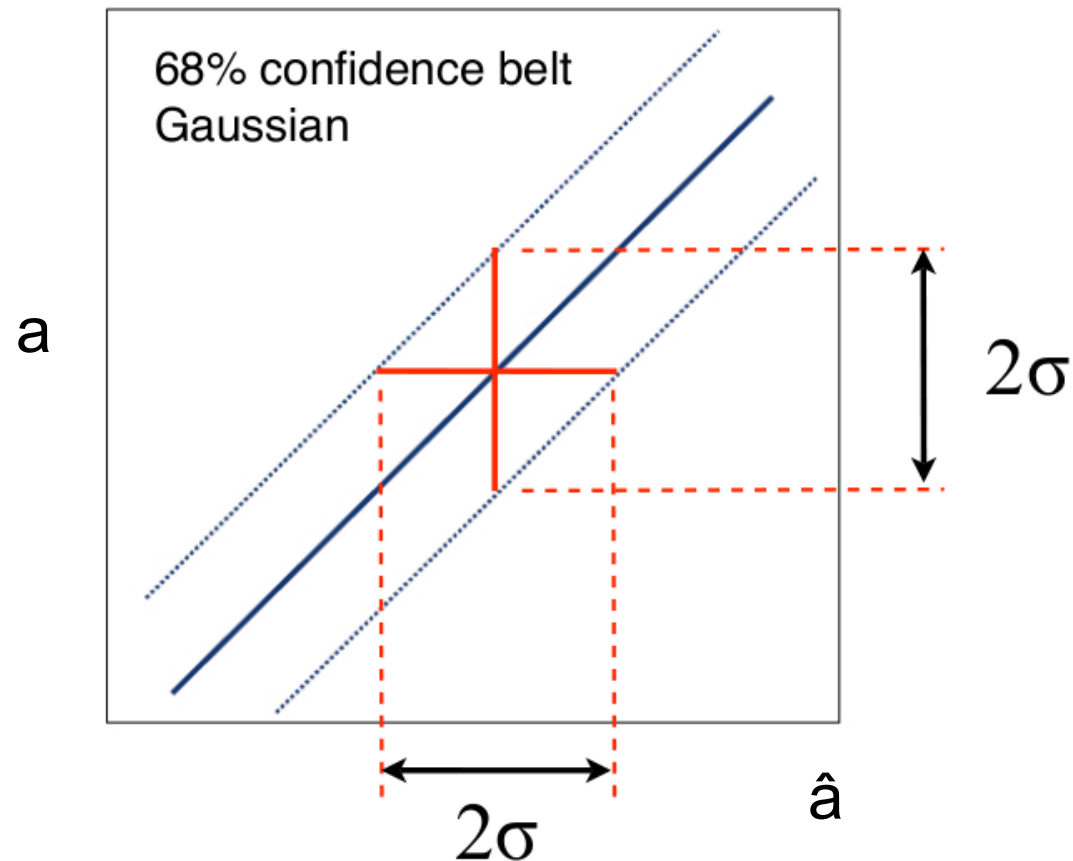


Neyman-Konstruktion für die Gauß-Verteilung

- Für die Gauß-Verteilung besteht das Konfidenz-Band aus einer Geraden mit Steigung 1.

- Anmerkung:
Deshalb ist es üblich, den Fehlerbalken an den gemessenen Wert zu zeichnen - und ihn dann als Konfidenzbereich für den wahren Wert a zu interpretieren!

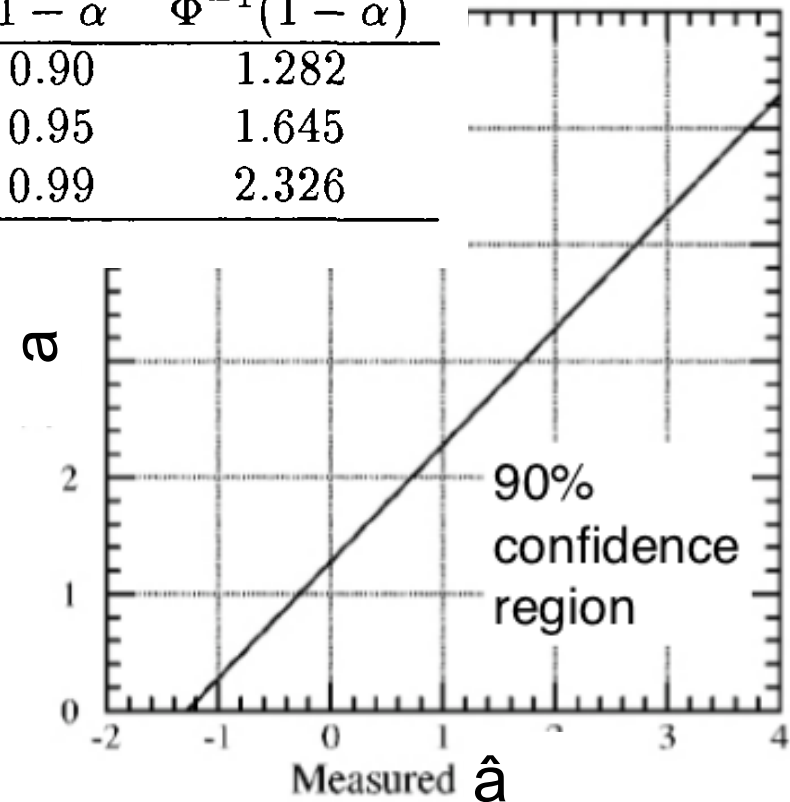
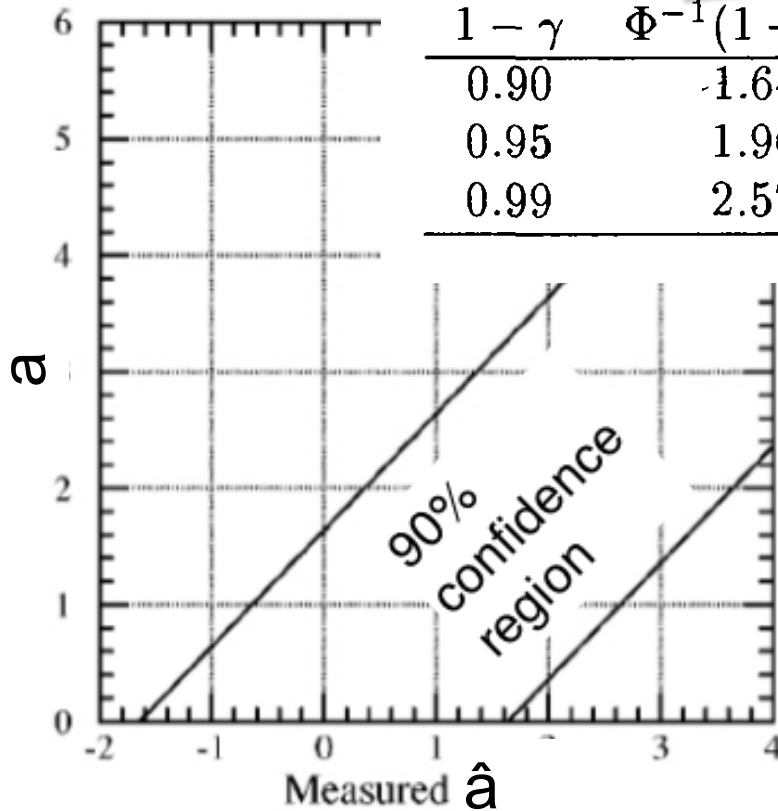
Dies gilt aber nur, wenn der Fehler nicht vom Messwert selbst abhängt (also z.B. nicht bei Poisson-Verteilung)



Neyman-Konstruktionen für die Gauß-Verteilung

Cowan: Tabelle 9.2

$1 - \gamma$	$\Phi^{-1}(1 - \gamma/2)$	$1 - \alpha$	$\Phi^{-1}(1 - \alpha)$
0.90	1.645	0.90	1.282
0.95	1.960	0.95	1.645
0.99	2.576	0.99	2.326

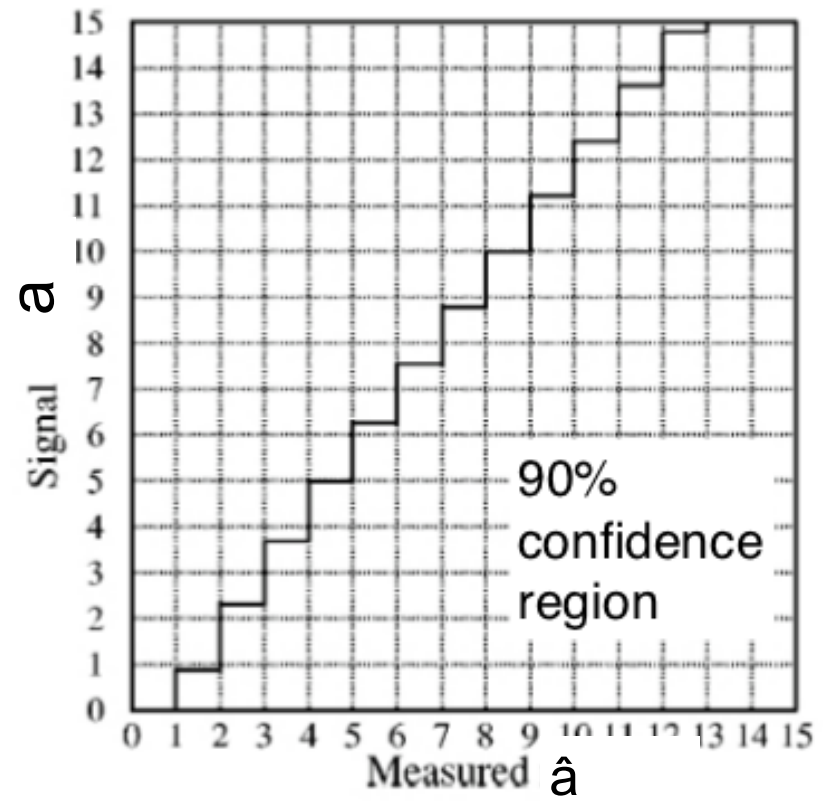
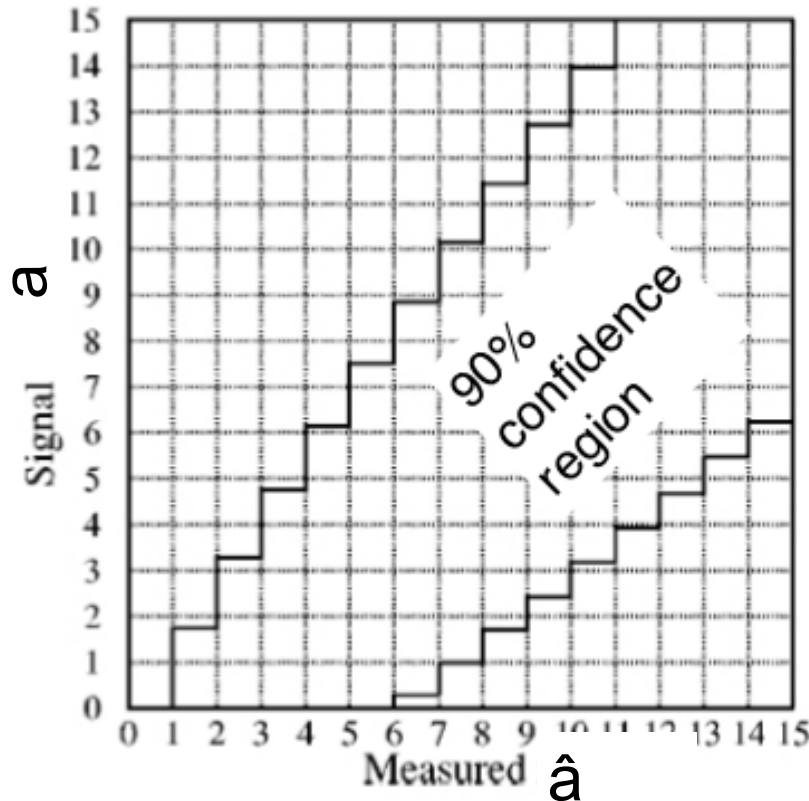


Zweiseitiges und einseitiges
Konfidenzband einer Gauß-Verteilung

5%+5%: 1.645σ 10%: 1.282σ

Neyman-Konstruktionen für die Poisson-Verteilung

- 90% Konfidenz-Interval für ein unbekanntes Poisson-verteiltes Signal mit (hier genau bekanntem) Untergrund von 3



Zweiseitiges und einseitiges
Konfidenzband einer Poisson-Verteilung

- In diesem Fall ist für $n=0$ ist das Band leer, ein wohlbekanntes Problem

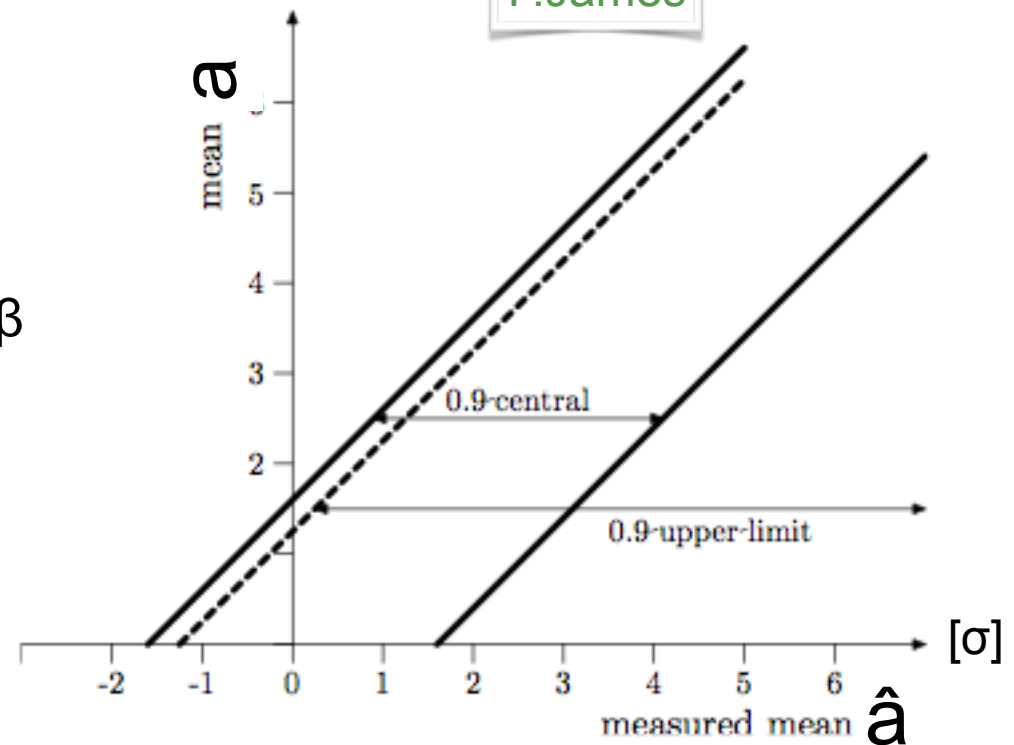
- Festlegung vor der Messung:
 - Konfidenz-Niveau (z.B. 68%, 90%, 95%)
 - Einseitiges (Limit) oder Zweiseitiges Intervall (Messung)
- Zweiseitige Konfidenz-Intervalle $1-\alpha-\beta \rightarrow$ Messungen
 - Symmetrisches (auch “zentrales”) Konfidenz-Intervall (gleiche Wahrscheinlichkeiten rechts und links)
 - Kürzest mögliches Intervall $[a_{\max}-a_{\min}]$ minimal
- Einseitige Konfidenz-Intervalle $\rightarrow$ Schranken (engl. Limits)
 - $a_{\min}$ = untere Schranke für a mit Wahrscheinlichkeit $1-\alpha$
 - $a_{\max}$ = obere Schranke für a mit Wahrscheinlichkeit $1-\beta$
- Bisher Gauß und Poisson betrachtet. In der Praxis sind PDF häufig komplizierter $\rightarrow$ MC-Methoden

Ein- und zweiseitige Intervalle

F. James

- Festlegung vor der Messung:
 - Konfidenz-Niveau
(z.B. 68%, 90%, 95%)
 - Einseitiges Limit $1-\alpha$ oder
Zweiseitiges Intervall $1-\alpha-\beta$
(Messung)

- Je nach Wahl verschiebt sich
der Konfidenz-Gürtel



- Häufige Praxis: Wahl der Intervallgrenzen in Abhängigkeit vom
Messergebnis
 - $\hat{a} > 3\sigma \rightarrow$ Messung (2-seitiges Konfidenz-Band, $1-\alpha-\beta$, 90% C.L.)
 - $\hat{a} < 3\sigma \rightarrow$ Obere Schranke (1-seitiges Konfidenz-Band, $1-\alpha$, 90% C.L.)

Ein- und zweiseitige Intervalle → Flip-Flopping

F.James, Abb. 9.7

■ Abbildung: für $a = 2.5$:

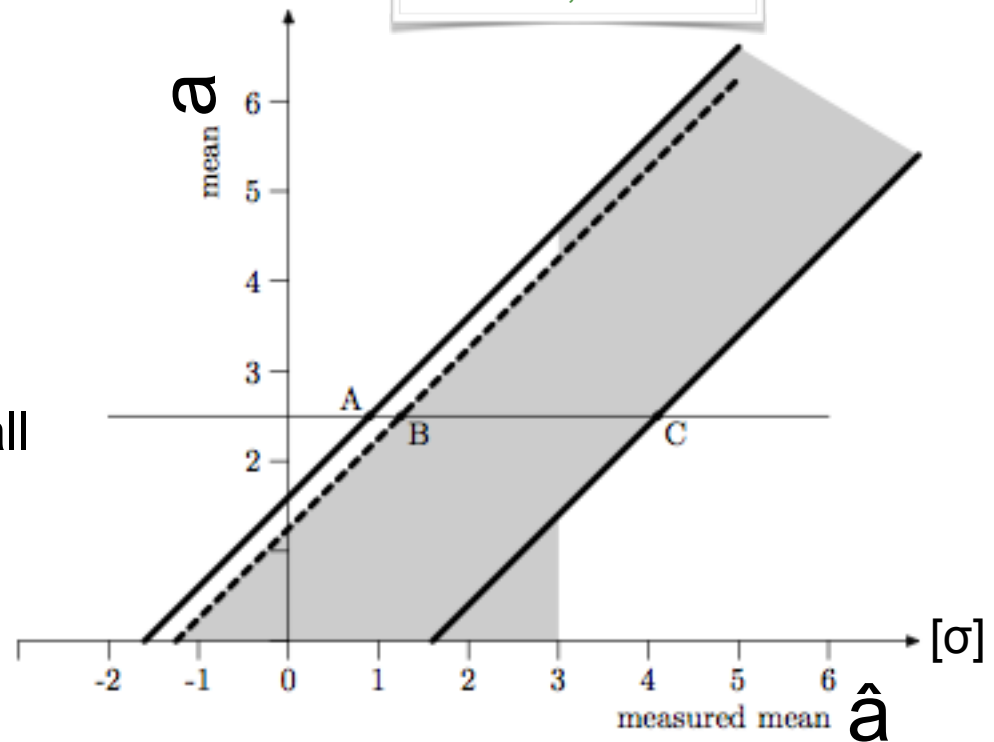
■ AC: 90% CL zweiseitig

■ B^∞ : 90% CL einseitig

■ BC: 85% CL

■ Nur 85% Abdeckung im Intervall
 $1.2 < a < 4.3$

■ Zudem: Leeres Intervall, d.h.
nicht definiert z.B. für $\hat{a} = -2$



■ Häufige Praxis: Wahl der Intervallgrenzen in Abhängigkeit vom Messergebnis

■ $\hat{a} > 3\sigma \rightarrow$ Messung (2-seitiges Konfidenz-Band, $1-\alpha-\beta$, 90% C.L.)

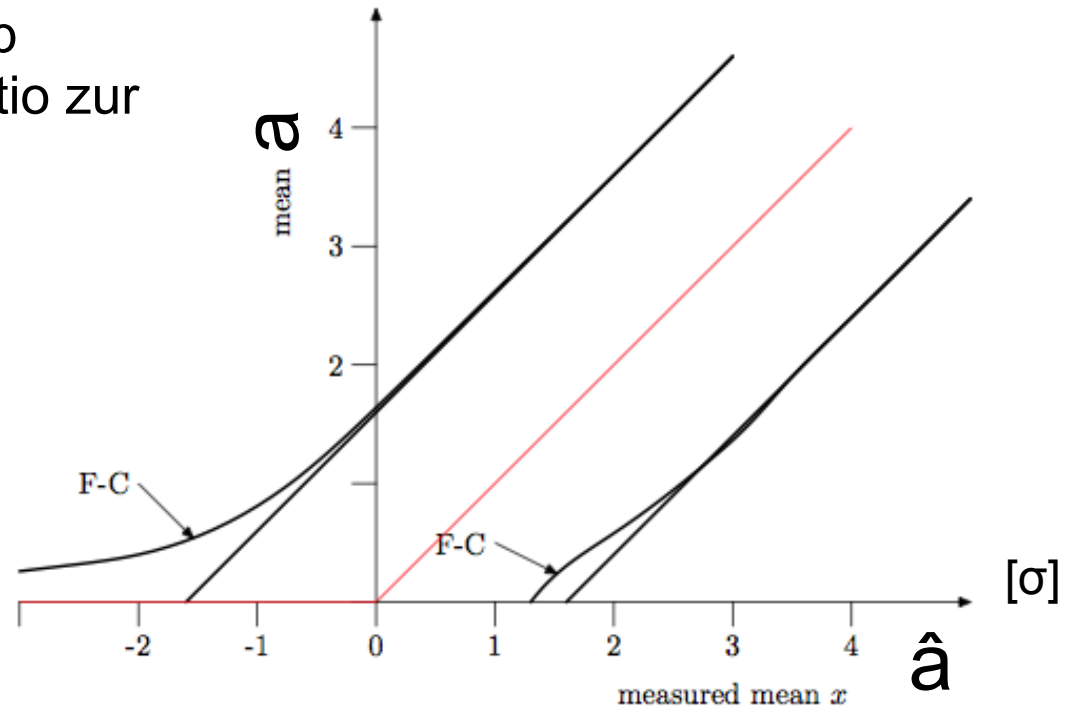
■ $\hat{a} < 3\sigma \rightarrow$ Obere Schranke (1-seitiges Konfidenz-Band, $1-\alpha$, 90% C.L.)

F.James, Abb. 9.8

- Feldman-Cousins Methode:
Verwendet Ordnungsprinzip
basierend of Likelihood-Ratio zur
Konstruktion des Intervalls

$$R(\hat{a}|a) = \frac{g(\hat{a}|a)}{g(\hat{a}|a_{\text{best}})}$$

- $a_{\text{best}} = a$ mit größtem $g(\hat{a}|a)$

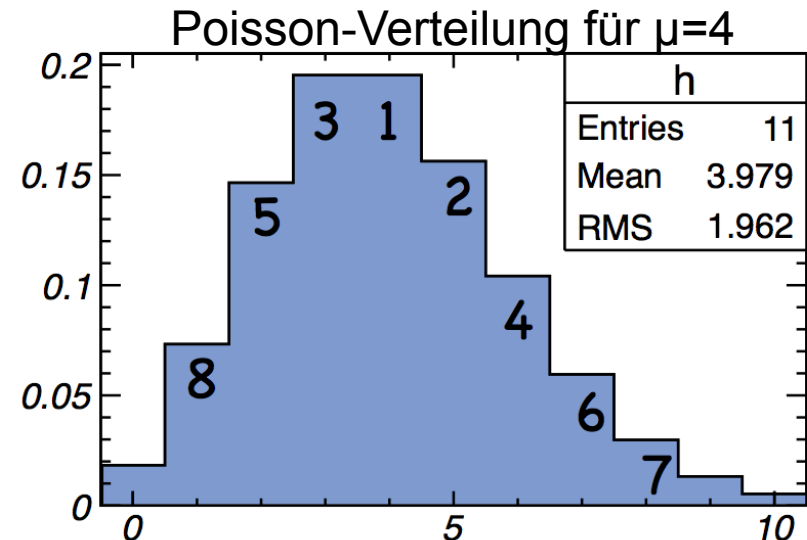


- Ordnungsprinzip:
 - Addieren von $\hat{a}$ Werten nach absteigenden Werten von R bis $g(\hat{a}|a)$ den angestrebten C.L. erreicht.
 - Für $\hat{a} < 0$: Hinzufügen von Beiträgen auf der linken Seite
- Automatische Entscheidung, ob Messung oder Schranke → “Unified”

Feldman-Cousins Unified Approach: Beispiel

- Ordnungsprinzip: Entscheidung, welches Kriterium für die Definition der Intervallgrenzen verwendet wird.
- Likelihood-Quotient: $R(\hat{a}|a) = \frac{g(\hat{a}|a)}{g(\hat{a}|a_{\text{best}})}$; a_{best} ist a , für das $g(\hat{a}|a)$ maximal
- Beispiel: Poisson-Wahrscheinlichkeit mit Untergrund ($a=4 \rightarrow \mu=4$ und $\hat{a} \rightarrow n$)
- Berechne $R(n|\mu=4)$ für jeden möglichen Messwert n
- R definiert Rangfolge der Intervalle
- Addiere Intervalle bis gewünschte Abdeckung erreicht ist
- Erinnerung Poisson:

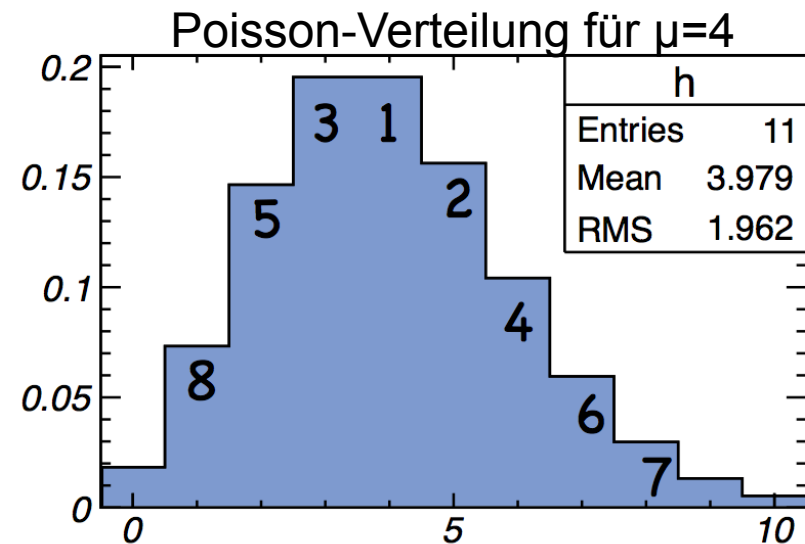
$$f(n; \mu) = \frac{e^{-\mu} \mu^n}{n!}$$



Feldman-Cousins Unified Approach: Beispiel

- Ordnungsprinzip: Entscheidung, welches Kriterium für die Definition der Intervallgrenzen verwendet wird.
- Likelihood-Quotient: $R(\hat{a}|a) = \frac{g(\hat{a}|a)}{g(\hat{a}|a_{\text{best}})}$; a_{best} ist a , für das $g(\hat{a}|a)$ maximal
- Beispiel: Poisson-Wahrscheinlichkeit mit Untergrund ($a=4 \rightarrow \mu=4$ und $\hat{a} \rightarrow n$)

n	$R(n \mu)$	$f(n \mu)$	$\sum f$
4	1.000	0.195	0.195
5	0.891	0.156	0.352
3	0.872	0.195	0.547
6	0.649	0.104	0.651
2	0.541	0.147	0.798
7	0.400	0.060	0.857
8	0.213	0.030	0.887
1	0.199	0.073	0.960



- Ergebnis: Konfidenz-Intervall $[1,8]$ mit 96% Abdeckung
- Komplexere Verteilungen: Rechenaufwand $\rightarrow$ Asymptotische Verfahren

Zusammenfassung: Neyman-Konstruktion

- Systematische Vorschrift aus frequentistischer Sicht zur
“Inversion der Wahrscheinlichkeitsaussagen”:
 - Gegeben ein wahrer Wert a , die Wahrscheinlichkeit $\hat{a}$ zu messen ist ...
 - Gegeben eine Messung $\hat{a}$, die Wahrscheinlichkeit, dass wahrer Wert a im Intervall liegt, ist ...
- Aussage über Abdeckung des wahren Werts durch das Intervall, nicht über Wahrscheinlichkeit des wahren Werts selbst
- Neyman-Konstruktion garantiert exakte Abdeckung (gemäß Vorgabe) und Überabdeckung für diskrete Verteilungen
 - Zentrale Konfidenz-Intervalle → Messungen
 - Einseitige Konfidenz-Intervalle → obere oder untere Schranken
 - Unified Frequentist Approach: verwendet Likelihood-Ratio als Ordnungsprinzip (Vermeidung von “Flip-flopping” und “leeren Intervallen”)
- Bemerkung: Betrachte Vorschriften immer im Verhältnis zur relativen Größe der statistischen Effekte und zum Aufwand. Andere Fehler als statistisch-methodische, z.B. systematische Unsicherheiten, spielen bei Messungen meist eine große Rolle

6.2

BAYES'SCHE STATISTIK

■ Degree of belief, dass a bestimmten Wert annimmt, d.h. wahrer Wert is Zufallsvariable mit PDF

■ Bayes' Posterior Wahrscheinlichkeit:
Wahrer Wert a liegt mit CL Wahrscheinlichkeit in Konfidenz-Intervall

$$f(a|\hat{a}) \propto \mathcal{L}(\hat{a}|a) \cdot \pi(a)$$

benötigt Prior-Annahme

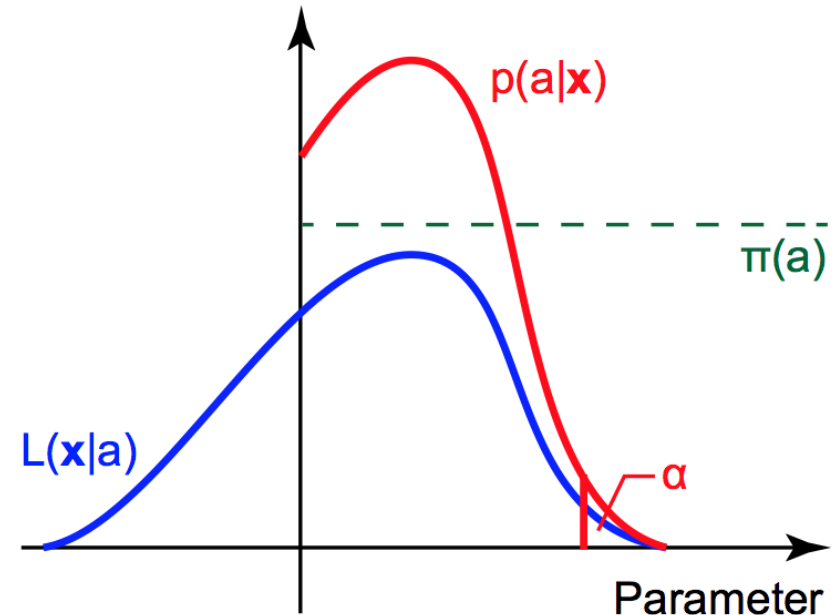
■ Vergleich: Frequentistische Likelihood $\mathcal{L}(\hat{a}|a)$ liefert p-Wert:
p% aller aus Messungen $\hat{a}$ konstruierten Konfidenz-Intervalle enthalten den wahren Wert a .

- Likelihood-Funktion und Prior-Annahmen geben Posterior-Wahrscheinlichkeit (hier x =Daten)

$$p(a|x) = \frac{L(x|a) \cdot \pi(a)}{\int L(x|a') \cdot \pi(a') da'}$$

- Posterior enthält alles Wissen über den Parameter
- Wahl des Priors ist entscheidend:
 - $\pi(a) = \text{const}$ („flacher Prior“) $\rightarrow p(a|x) \sim L(x|a) \rightarrow$ Bayes'scher und klassischer Schätzwert identisch
 - Problem 1: $\pi(a)$ nicht invariant bei Re-parametrisierung $\rightarrow$ Unterschiedliche Resultate z.B. für $p(a|x)$, $p(\ln(a)|x)$
 - Problem 2: sind alle Parameterwerte wirklich a priori gleich wahrscheinlich ? $\rightarrow$ z.B. Zahl der Higgs-Bosonen in den LHC Daten $\pi(100) = \pi(10^{10})$?

- Konfidenzintervall aus Posterior-Verteilung $p(a|x)$
- Wahl der Prior-Verteilung $\pi(a)$ beinhaltet physikalisches Urteil
- Physikalische Grenzen sind in Prior-Verteilung berücksichtigt z.B. falls a immer positiv.
- Frequentistische Abdeckung nicht mehr automatisch garantiert

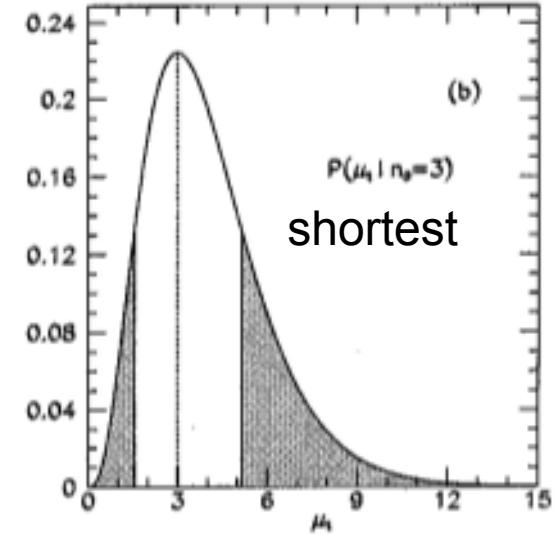
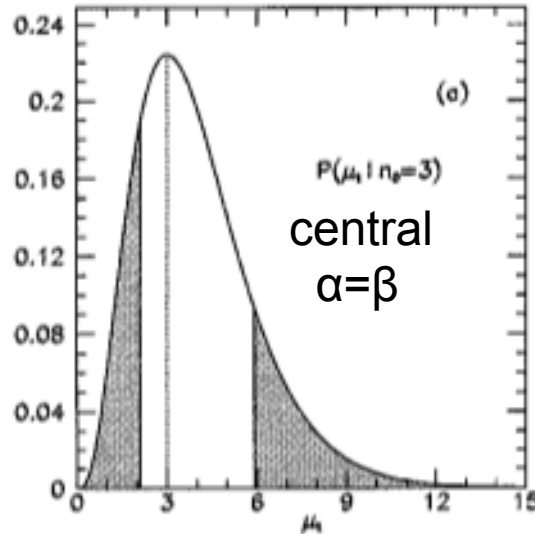


1. Umnormierung der Wahrscheinlichkeit im physikalischen Bereich
2. Konfidenzintervall durch Integration

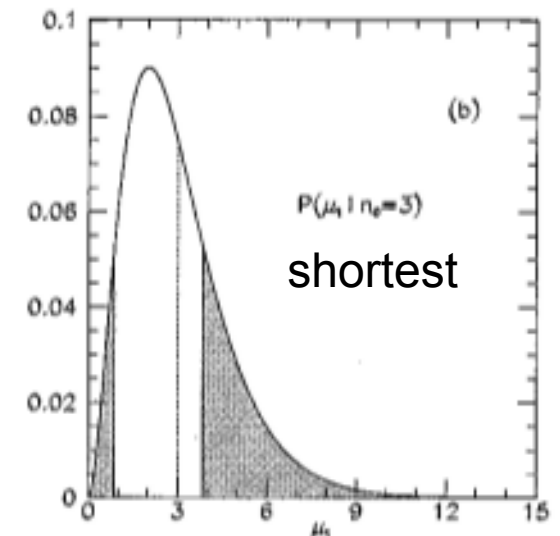
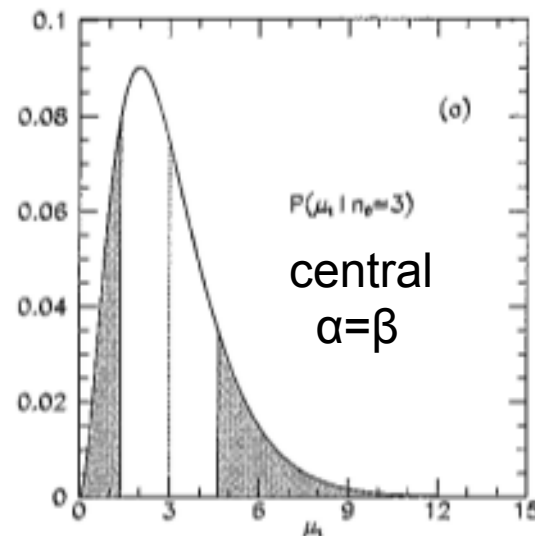
Einfluss des Priors

Bayes Intervalle für $n=3$ beobachtete Ereignisse

■ mit flachem Prior



■ mit $1/\mu$ Prior



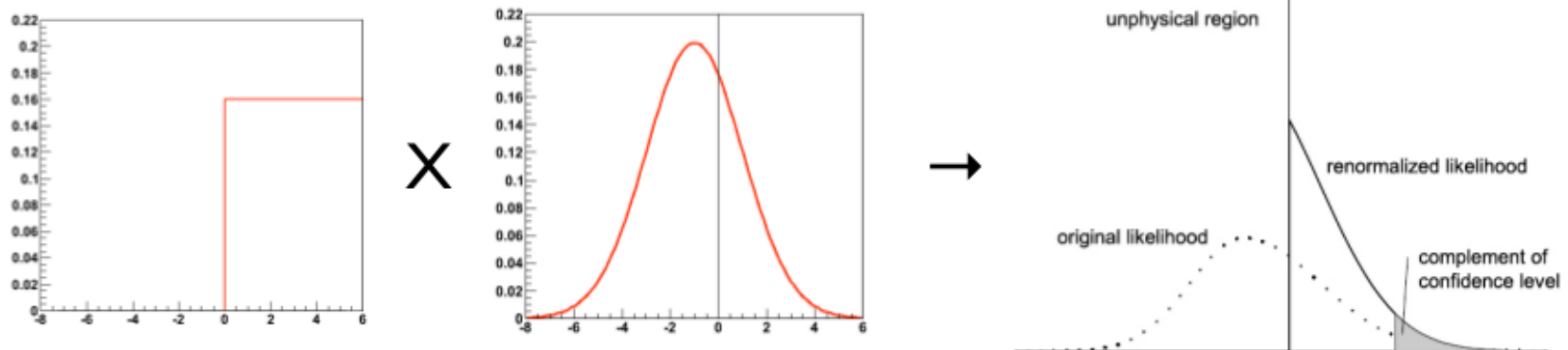
Limits an Physikalischen Grenzen

■ Typische Fälle

- Zahl der Signalereignisse > 0
- Teilchenmasse > 0
- Messung des Quadrats eines Parameters, $a^2 > 0$

■ Beispiel: $x=x-y$; x , y und damit auch z Gauß-förmig verteilt

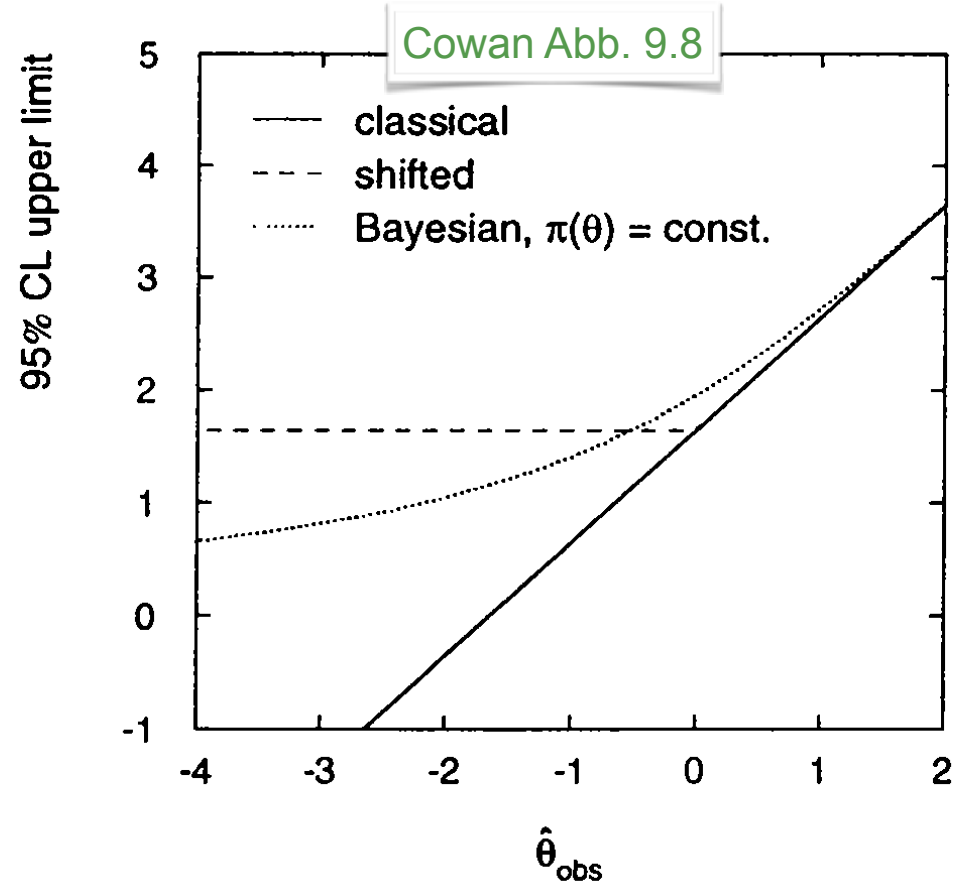
- Klassische 95%-Obergrenze: $z_{95} = z_m + 1.645 \cdot \sigma_z$
- Mit $z_m = -2$ und $\sigma_z = 1$ = -0.355 unphysikalische Obergrenze
- Bayes'sche Statistik bietet einfache Lösung: $f(z)=0$ für $z<0$ und $f(z)=1$ für $z>0$ (Stufenfunktion)



Limits an Physikalischen Grenzen

Vergleich der Methoden:

- frequentistisch
- “verschoben”,
d.h. falls $z_m < 0$, setze
Messwert $z_m = 0$ mit
Obergrenze für 95% CL:
 1.645σ
- bayesisch
mit flacher Stufenfunktion
als Prior



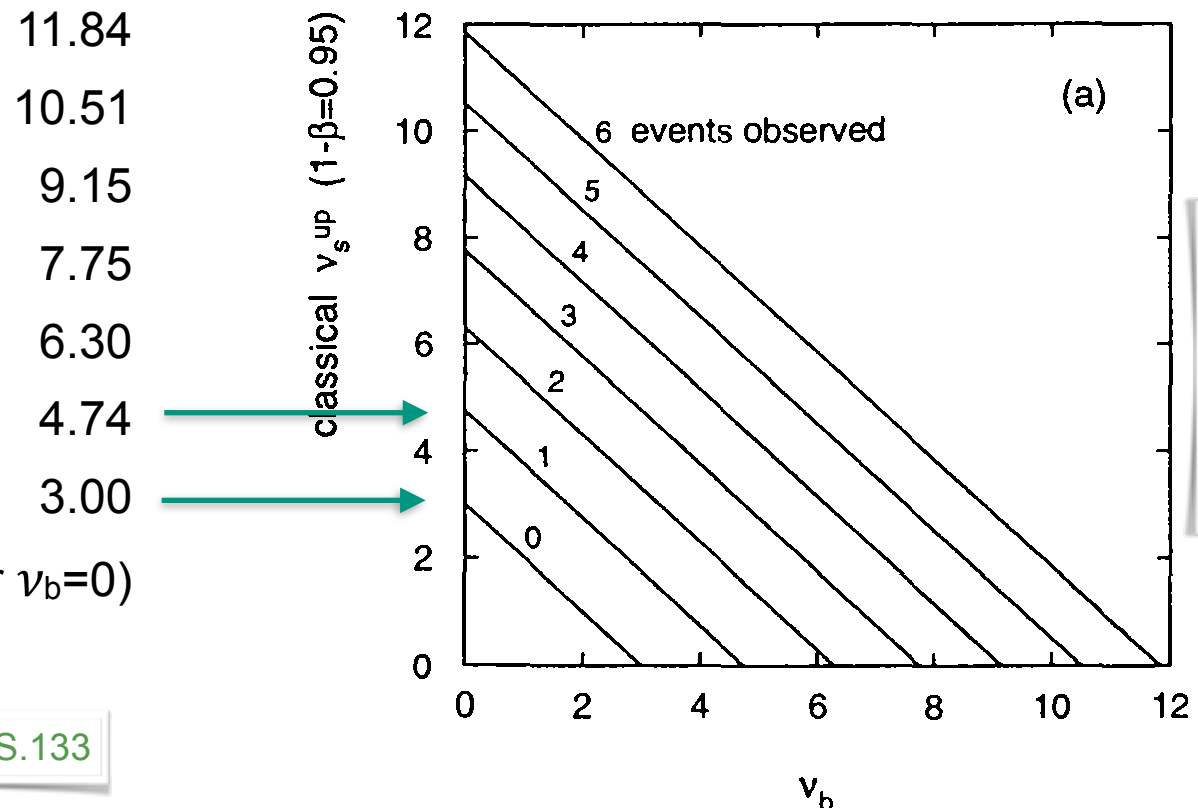
- Im allgemeinen sollte man auch die Messungen von negativen Signalen publizieren. Mittelung über mehrere Messungen gibt dann den korrekten Wert.

- Betrachte typische Suche, d.h. Anzahl der Signalereignisse ν_s klein
- Signal + Untergrund ist Poisson-verteilt: $p(n | \nu_s, \nu_b) = p(n | \nu = \nu_s + \nu_b)$
- Bestimmung ν und Subtraktion von ν_b liefert Schätzwert $\hat{\nu}_s = n - \nu_b$
- Oberes Limit (95% CL) für ν_s als Funktion des erwarteten U.-grunds ν_b , für verschiedene n_{obs} :

$$\sum_{r=n+1}^{\infty} P(r | \nu) = 0.95$$

$$\sum_{r=0}^n P(r | \nu) = 0.05$$

(95% CL Limits für $\nu_b=0$)



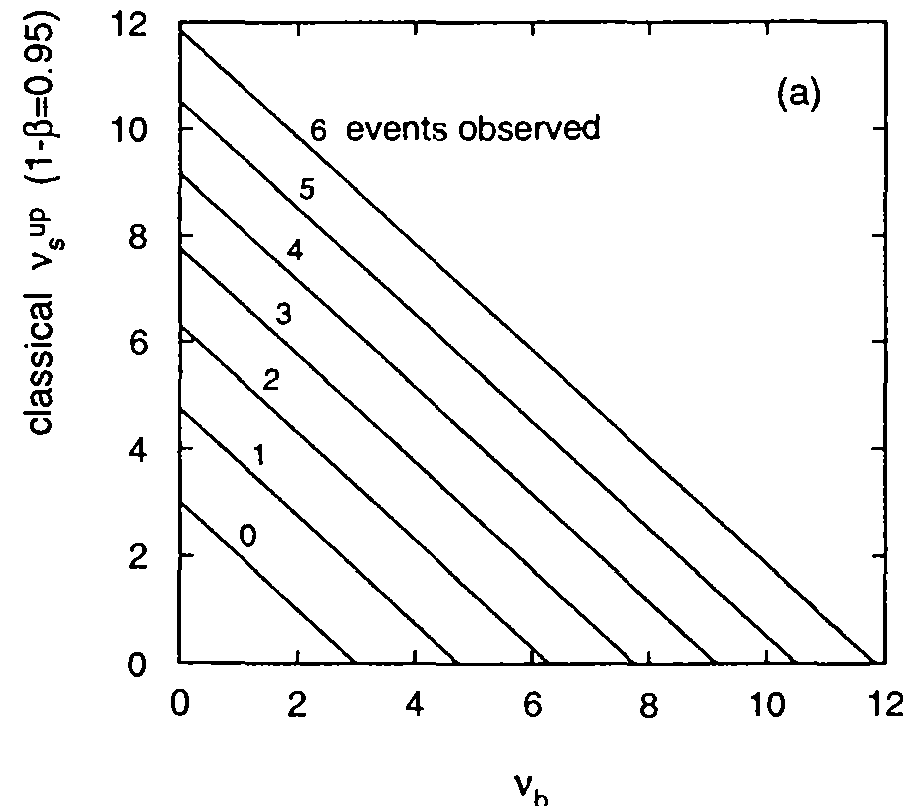
Cowan Abb. 9.9

Tabelle: z.B. Barlow, S.133

- Betrachte typische Suche, d.h. Anzahl der Signalereignisse ν_s klein
- Signal + Untergrund ist Poisson-verteilt: $p(n | \nu_s, \nu_b) = p(n | \nu = \nu_s + \nu_b)$
- Bestimmung ν und Subtraktion von ν_b liefert Schätzwert $\hat{\nu}_s = n - \nu_b$
- Oberes Limit (95% CL) für ν_s als Funktion des erwarteten U.-grunds ν_b , für verschiedene n_{obs} :

■ Probleme mit diesem Ansatz:

- Für n_{obs} klein und ν_b groß existiert kein positives Limit
- Wenn $n_b < \nu_b$ (d.h. stat. Fluktuation des Untergrunds nach unten): Experiment mit mehr Untergrund misst mit gleich vielen Ereignissen bessere obere Grenze!



Poisson-Signal und Untergrund: mit Bayes' Prior

- Betrachte typische Suche, d.h. Anzahl der Signalereignisse ν_s klein
- Signal + Untergrund ist Poisson-verteilt: $p(n | \nu_s, \nu_b) = p(n | \nu = \nu_s + \nu_b)$
- Bestimmung ν und Subtraktion von ν_b liefert Schätzwert $\hat{\nu}_s = n - \nu_b$
- Oberes Limit (95% CL) für ν_s als Funktion des erwarteten U.-grunds ν_b , für verschiedene n_{obs} :

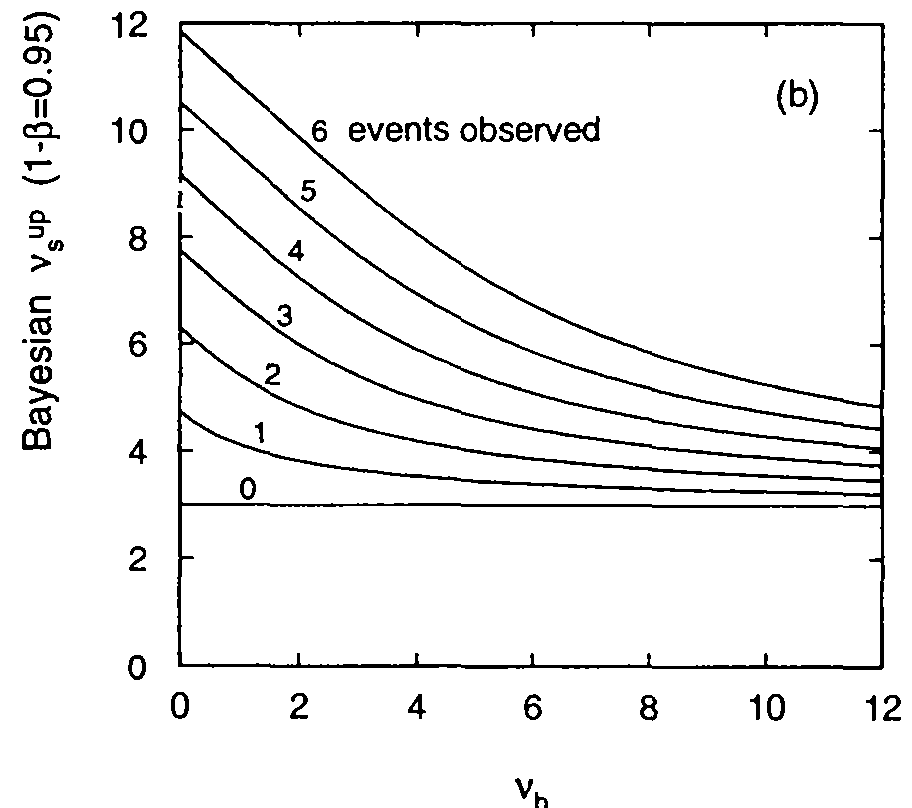
■ Bayes' Prior:

$$\pi(\nu_s < 0) = 0 \text{ and } \pi(\nu_s \geq 0) = \text{const}$$

liefert gute Eigenschaften:

- gleiches Ergebnis für $\nu_b = 0$
- höhere Grenze für $\nu_b > 0$

d.h. Abdeckung ist hinreichend



6.3

MODIFIZIERTER FREQUENTISTISCHER ANSATZ

Modifizierter (Frequentistischer) Ansatz: CL_s

- Gesucht: Robuste obere Schranke (oder Messung) eines hypothetischen Signals (gemessen mit Observabler d), aber ohne Bayes' Prior
- Bestimme obere Grenzen für zwei Hypothesen:
 - H_1 : Gemessene Ereignisse bestehen aus Untergrund und Signal:
 $d = s + b$ → liefert p-Wert = " CL_{b+s} "
 - H_0 : Gemessene Ereignisse bestehen nur aus Untergrund:
 $d = b$, d.h. $s=0$ → liefert p-Wert = " CL_b "

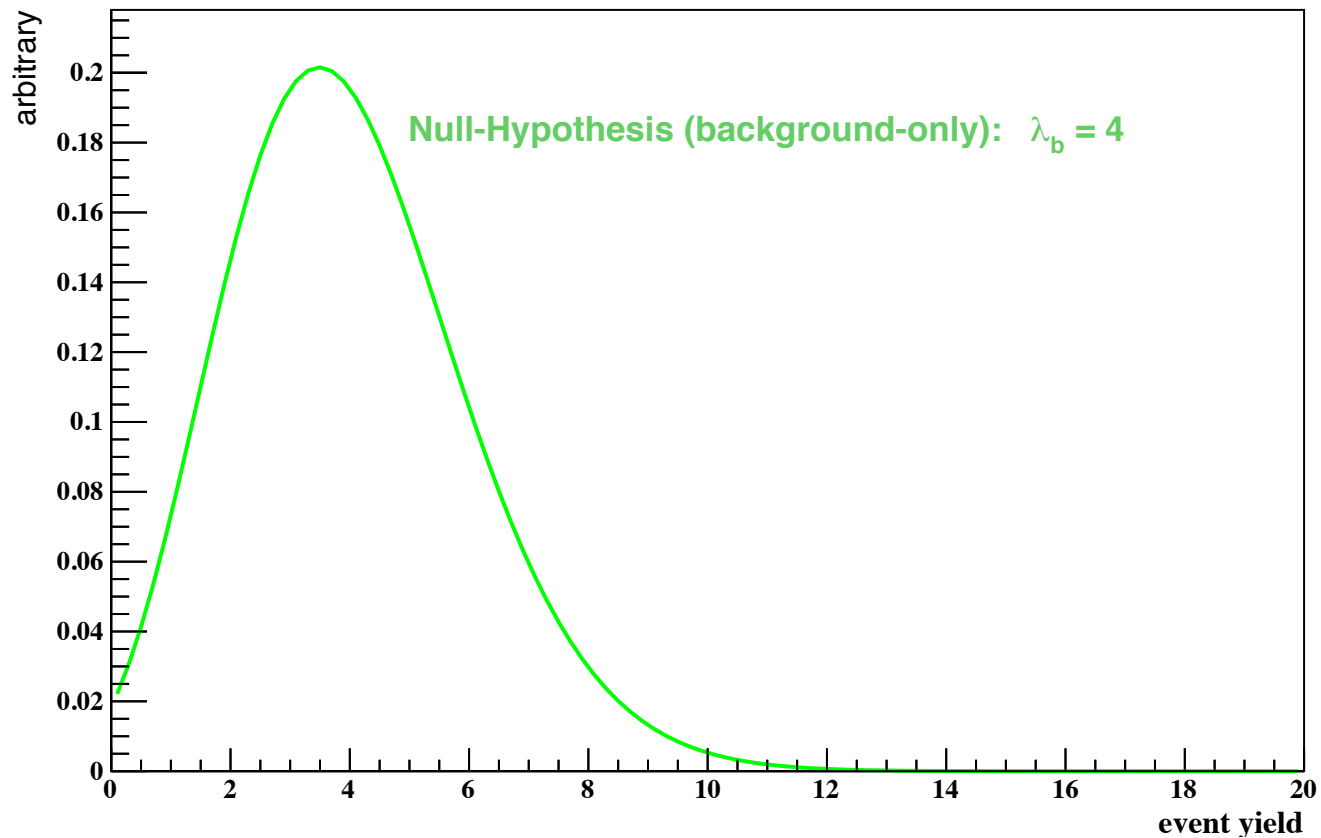
$$CL_s = \frac{CL_{s+b}}{CL_b} = \frac{\sum_{k=0}^d P(k; s + b)}{\sum_{k=0}^d P(k; b)} = \frac{p_\mu}{1 - p_b}$$

- CL_s renormiert gemessene Grenze auf die Abschätzung des Untergrunds:
 - Quantitativ ähnlicher Effekt wie mit konstantem Bayes' Prior
 - Vergleichbarkeit von Experimenten mit unterschiedlichem Untergrund
 - CL_s immer größer als p_μ → Abdeckung ok

G.Cowan, PDG,
Abschnitt 39.4.2.4

Beispiel

■ Erwarteter Untergrund: $b=4$



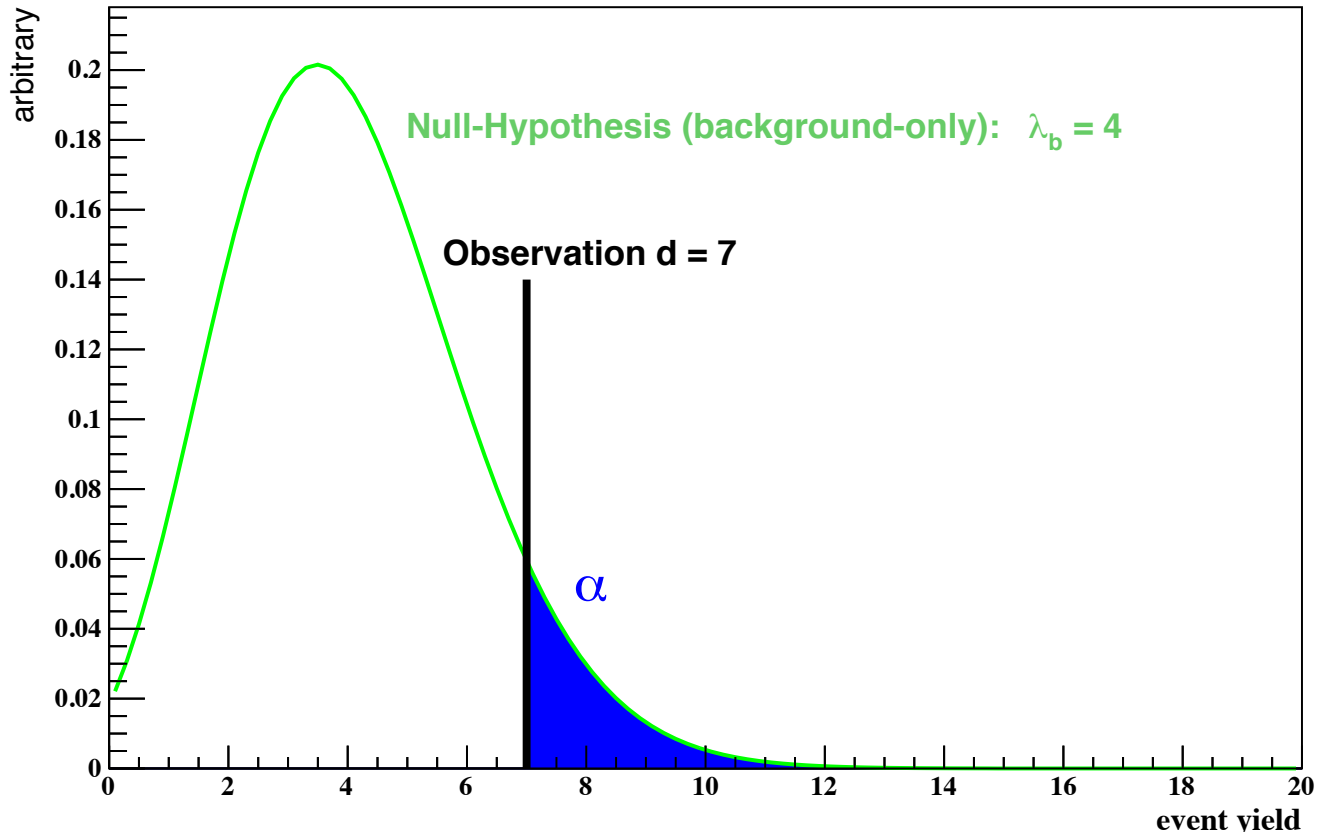
Dank an:
Christian Autermann,
RWTH Aachen

Beispiel

■ Erwarteter Untergrund: $b=4$

■ Messung: $d=7$

$$\alpha = 1 - \int_0^d P(x|b)dx$$

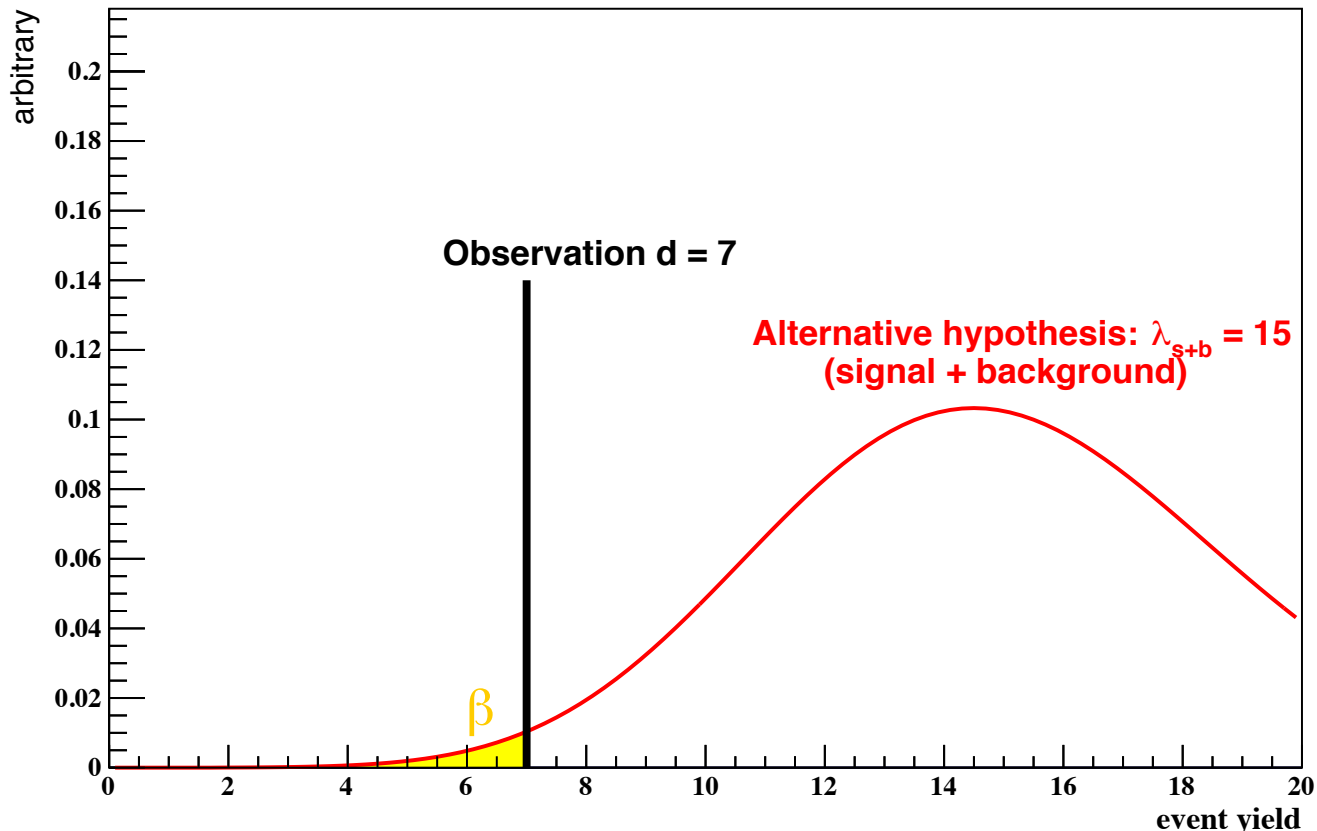


Dank an:
Christian Autermann,
RWTH Aachen

Beispiel

- Erwarteter Untergrund: $b=4$
- Messung: $d=7$
- Erwartetes Signal (z.B. neue Physik): $s=11$

$$\beta = \int_0^d P(x|b + s)dx$$



Dank an:
Christian Autermann,
RWTH Aachen

Beispiel

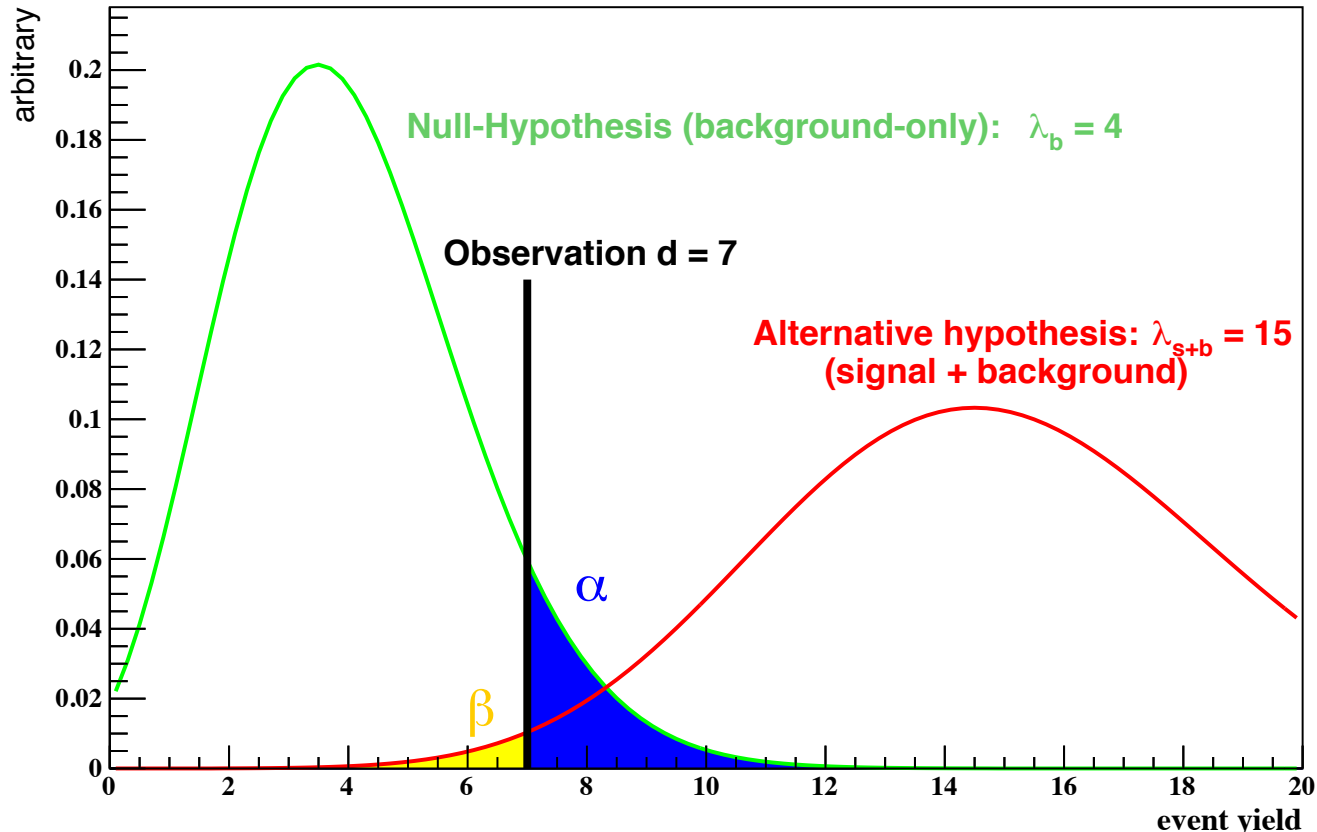
■ Erwarteter Untergrund: $b=4$

■ Messung: $d=7$

■ Erwartetes Signal: $s=11$

$$\text{CL}_{b+s} \quad \beta = \int_0^d P(x|b+s)dx$$

$$1-\text{CL}_b \quad \alpha = 1 - \int_0^d P(x|b)dx$$



Dank an:
Christian Autermann,
RWTH Aachen

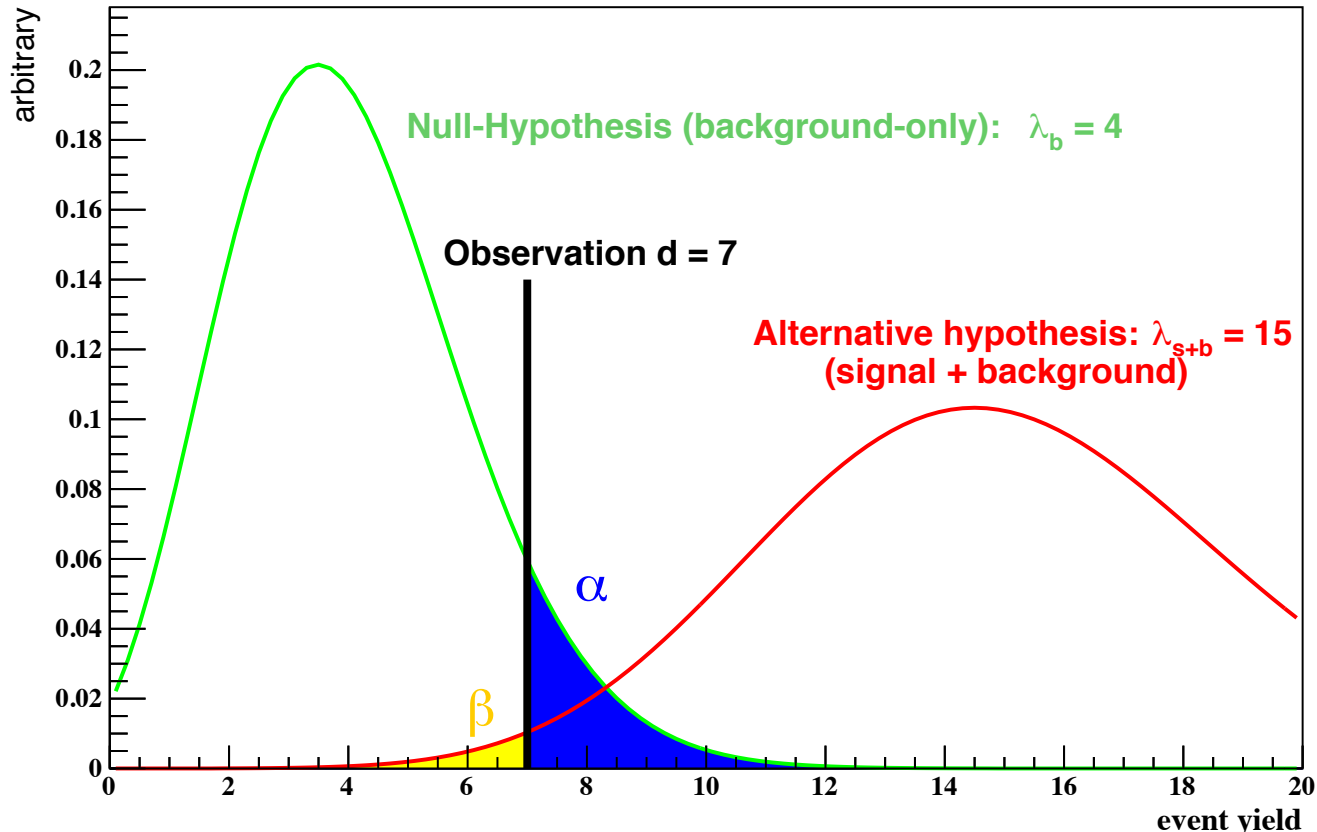
Beispiel

■ Frage: Was ist die gemessene obere Grenze s_{oben} (95% CL_{s+b}) auf das Signal ?

$$\beta = \int_0^d P(x|b + s) dx$$

■ Antwort: $s_{\text{oben}} = 8.5$, d.h. $s+b = 12.5$

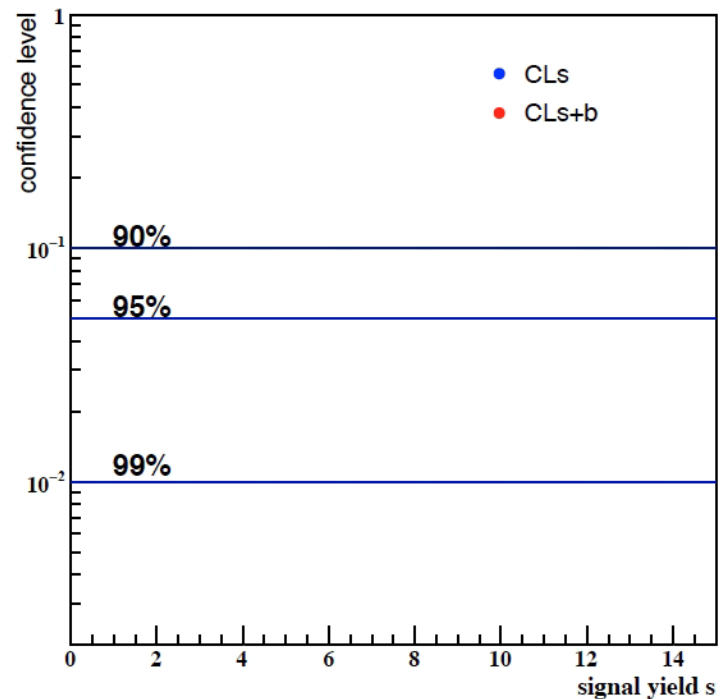
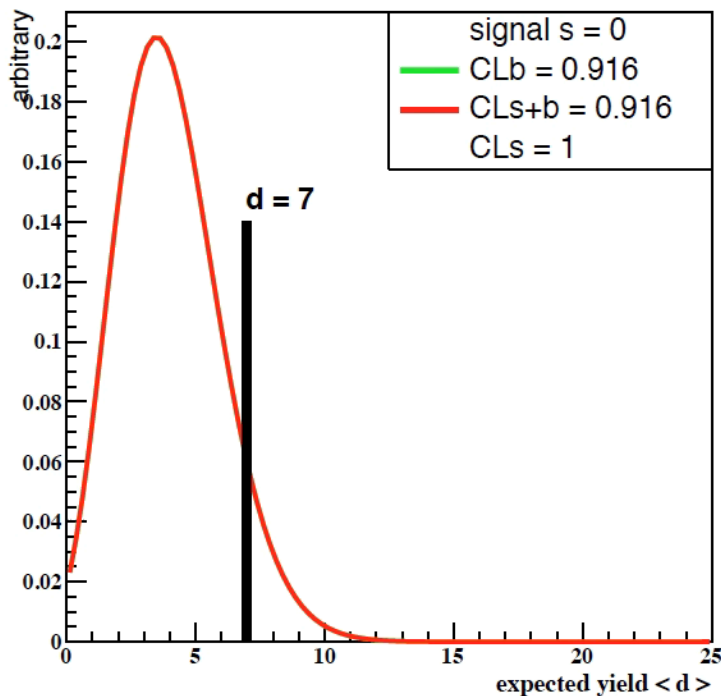
■ D.h. die Messung von $d=7$ schliesst die ursprüngliche Erwartung $s=11$ also $s+b=15$ mit 95% CL_{s+b} aus.



Dank an:
Christian Autermann,
RWTH Aachen

Beispiel

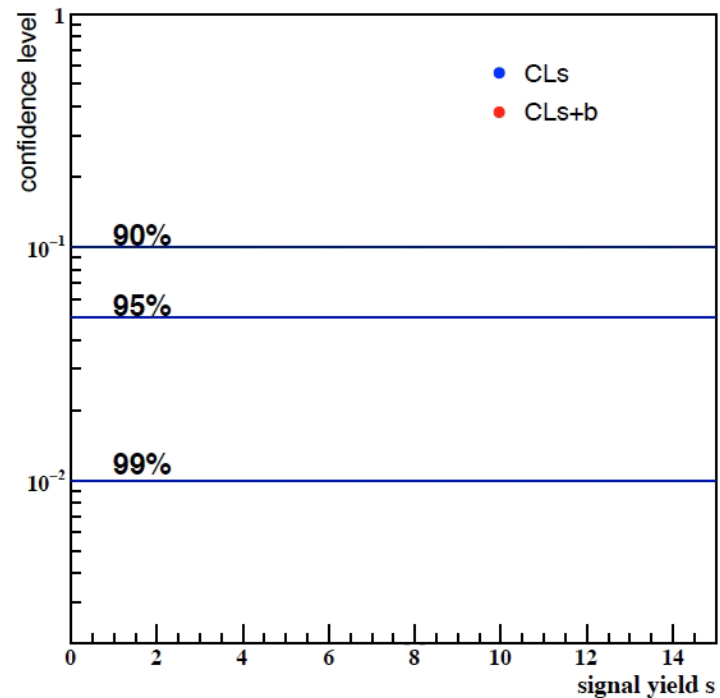
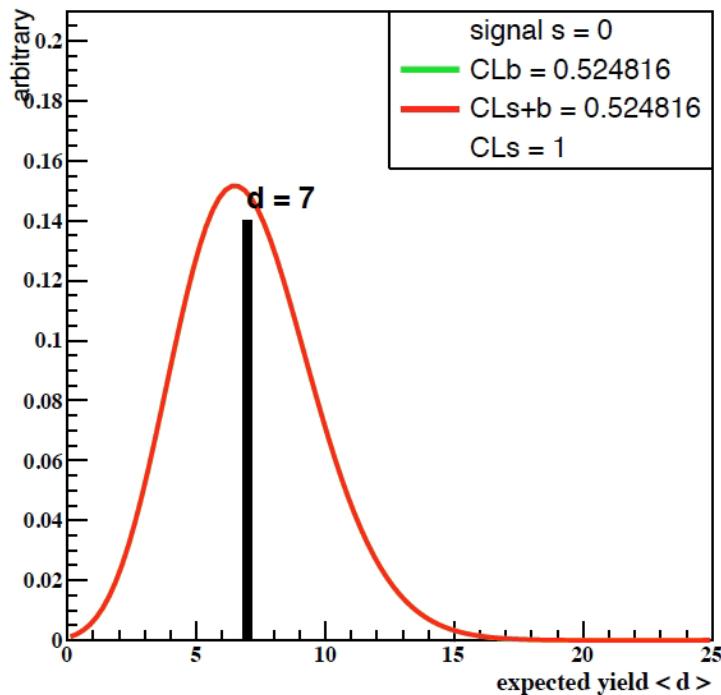
- Bestimmung der Obergrenze durch Scan der Signalhypothesen und Vergleich mit Messung
- Für Poisson-verteilte $b = 4$ und $d = 7$:
 - Obere Grenze für $CL_{S+B} = 95\%$: $s = 8.5 \Rightarrow s+b = 12.5$
 - Obere Grenze für $CL_S = 95\%$: $s = 8.7 \Rightarrow s+b = 12.7$



Dank an:
Christian Autermann,
RWTH Aachen

Beispiel

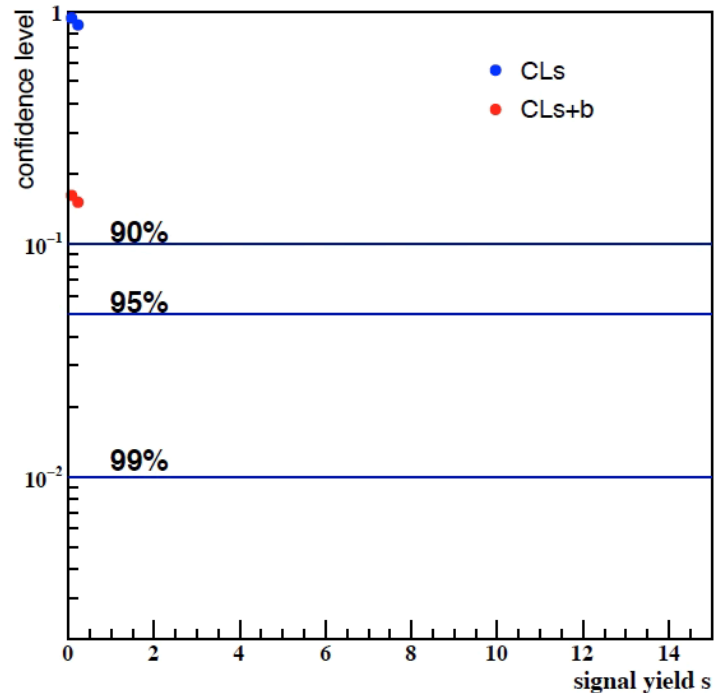
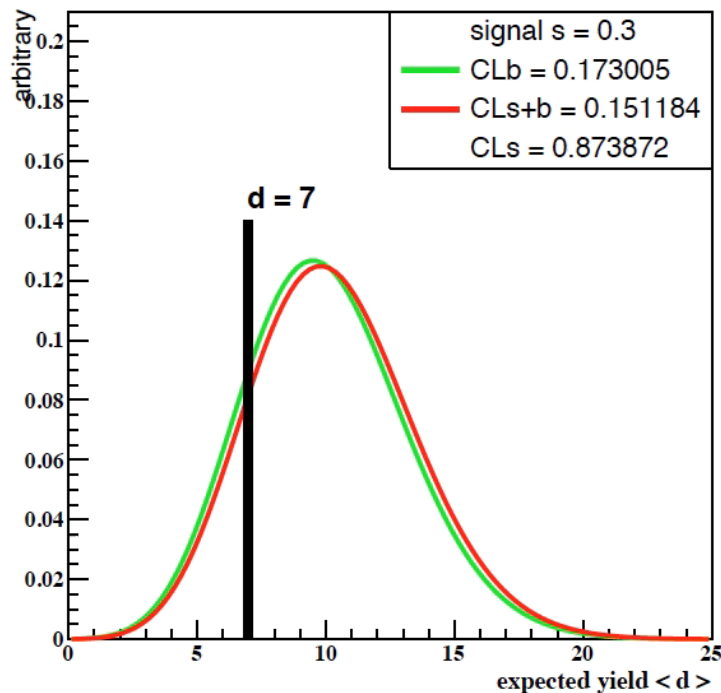
- Bestimmung der Obergrenze durch Scan der Signalhypothesen und Vergleich mit Messung - jetzt mit mehr Untergrund
- Für Poisson-verteilte $b = 7$ und $d = 7$
 - Gleiche obere Grenze für $CL_{S+B} = 95\%$: $s = 5.5 \Rightarrow s+b = 12.5$
 - Schlechtere obere Grenze für $CL_S = 95\%$: $s = 6.6 \Rightarrow s+b = 13.6$



Dank an:
Christian Autermann,
RWTH Aachen

Beispiel

- Bestimmung der Obergrenze durch Scan der Signalhypothesen und Vergleich mit Messung - jetzt mit Abwärtsfluktuation des Untergrunds
- Für Poisson-verteilte $b = 10$ und $d=7$:
 - Gleiche obere Grenze für $CL_{S+B} = 95\%$: $s=2.5 \Rightarrow s+b=12.5$
 - Noch schlechtere obere Grenze für $CL_S=95\%$: $s=5.5 \Rightarrow s+b=15.5$



Dank an:
Christian Autermann,
RWTH Aachen

Zusammenfassung Konfidenz-Intervalle (1)

- Bestimme Intervall, in dem wahrer Wert mit vorgegebener Wahrscheinlichkeit liegt
- Relevant für nicht-Gauß-verteilte PDF

- Frequentistischer Ansatz:
 - Neyman-Konstruktion - Inversion der Wahrscheinlichkeitsverteilungen
 - Korrekte Abdeckung per Definition
 - Unified Frequentist Approach: verwendet Likelihood-Ratio als Ordnungsprinzip (Vermeidung von Flip-flopping und leeren Intervallen)

- Bayes'sche Priors:
 - Zur Vermeidung unphysikalischer Ergebnisse
 - Abdeckung muss separat getestet werden

- Modifizierter frequentistischer Ansatz CL_s :
 - Etablierte, robuste Methode, um statistische Fluktuationen des Untergrunds zu berücksichtigen
 - Überabdeckung per Definition