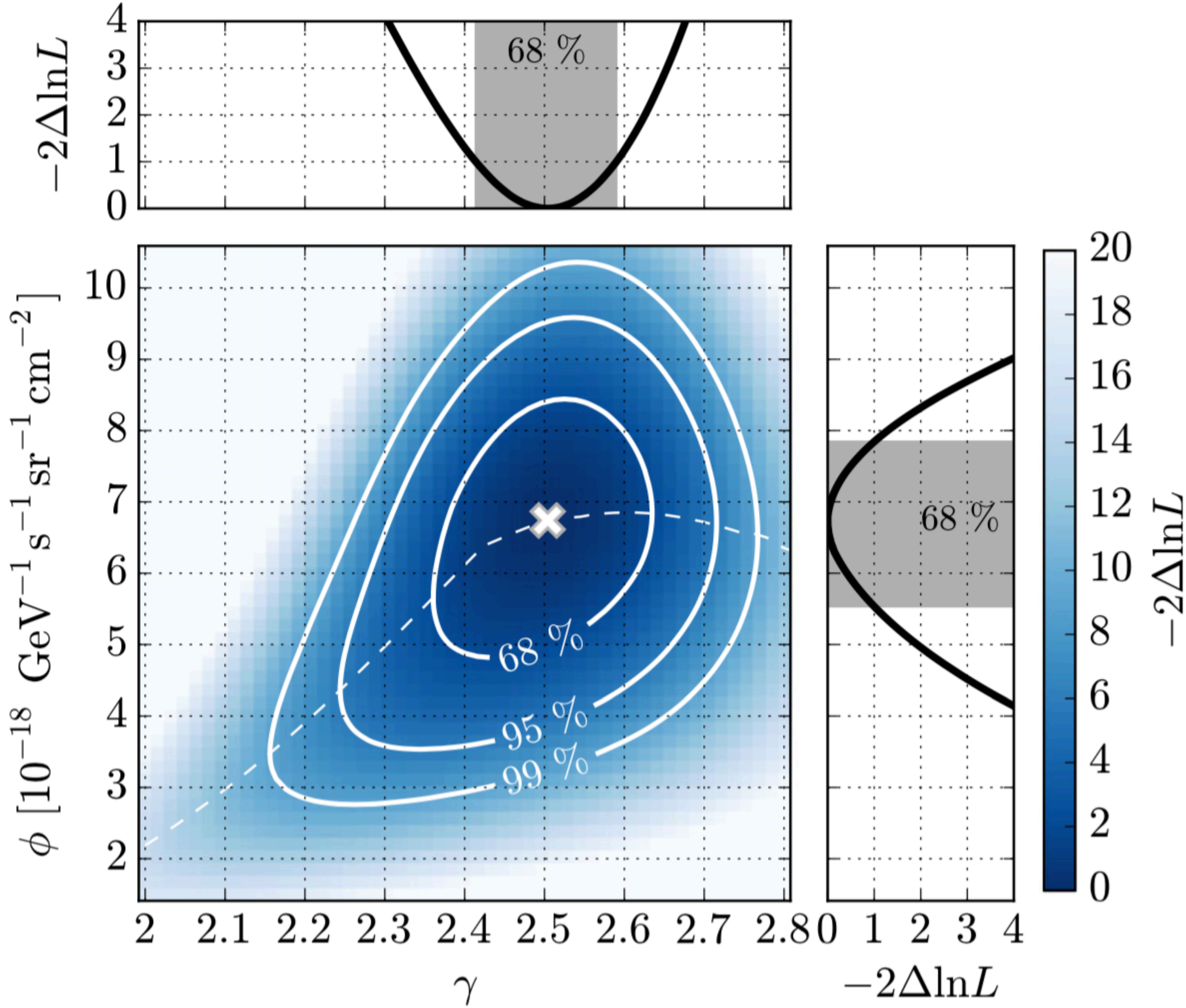
technische universität dortmund

Experimentelle Physik Vb
Astroteilchenphysik

Vorlesung
Statistische Methoden der Datenanalyse
Prof. Dr. Dr. Wolfgang Rhode

Intervallschätzungen und Hypothesentests



$-2\Delta\ln L$
68 %
ϕ [10^{-18} GeV^{-1}s^{-1}sr^{-1}cm^{-2}]
68 %
95 %
99 %
γ
$-2\Delta\ln L$
68 %
$-2\Delta\ln L$

Inhalt

- Intervallschätzung
  - Konfidenzintervalle
  - Neyman Konstruktion
  - Feldman-Cousins Konfidenzbänder
  - Bayesische Konfidenzbänder
- Testen von Hypothesen
  - Typ-I und Typ-II Fehler
  - P-Values
- Statistische Tests
  - Likelihood-Quotienten Test
  - Gauß-, t-, F-Test
  - Kolmogorow-Smirnow Test
  - Chi-Quadrat Test („*goodness of fit*“)

INTERVALLSCHÄTZUNG

Konfidenzintervalle

- Varianzen, Momente
  - Unabhängig von der zugrunde liegenden Verteilung
  - Wohldefinierte Fehlerfortpflanzung
  - Fehlerkombination unabhängiger Messungen
- Konfidenzintervalle
  - Alternative Methode um Unsicherheiten anzugeben
  - Abhängig von der zugrundeliegenden Verteilung
  - Fehlerpropagation und Kombination schwierig
  - **Frequentistisch** ⇔ Bayesisch

Konfidenzintervalle – Frequentistische Definition

- Definition:
  Für eine Observable x mit einer vom Parameter θ abhängigen p.d.f. ist ein Konfidenzintervall $[\theta_1, \theta_2]$ bezüglich eines Konfidenzlevels α Teil einer Menge mit der Eigenschaft

  $$P(\theta \in [\theta_1, \theta_2]) = \alpha$$

  wobei θ_1 und θ_2 Funktionen von x sind

- **Was bedeutet das genau?**

Konfidenzintervalle – Interpretation

- Die Obervable x ist das, was im Experiment gemessen wird
- Dazu gibt es eine vorher bestimmte p.d.f., die von den Daten x und Parametern θ abhängt

$$\text{Likelihood: } \mathcal{L}(\theta|x) = P(x|\theta)$$

→ Wahrscheinlichkeit der Daten x, unter der Bedingung, dass θ wahr ist

Konfidenzintervalle – Interpretation (Fortsetzung)

- Interpretation der Definition ist subtil:

In einem Anteil α von durchgeführten Experimenten enthält das jeweils konstruierte Konfidenzintervall den wahren Parameter θ_{Wahr}

- Für einen (beliebigen, aber festen) Wert θ_{Wahr} ergeben verschiedene Messungen verschiedene Konfidenzintervalle
  - Die untere/obere Grenze θ_1, θ_2 ist abhängig von den jeweils gemessenen x
- Der Anteil α an erhaltenen Intervallen enthält dann den Wert θ_{Wahr}

- Vorsicht: Das heißt **NICHT**, dass …
  - … der wahre Parameter θ_{Wahr} mit Wahrscheinlichkeit α in $[\theta_1, \theta_2]$ liegt
  - … eine direkte Aussage über den Wert von θ_{Wahr} gemacht wird

Konfidenzintervalle – Deskriptive Statistik (Vorbereitung)

- Sei die p.d.f. bekannt → alle Parameter θ liegen fest → P(x)
- Das Konfidenzintervall [x_-, x_+] zum Konfidenzlevel α ist dann

$$P(x_- \leq x \leq x_+) = \int_{x_-}^{x^+} P(x)\ \mathrm{d}x = \alpha$$

- Freiheit bei der Wahl des Intervalls
  - Symmetrisch um den Erwartungswert μ: $x_+ - \mu = \mu - x_-$
  - Kürzestes Intervall: Der Abstand $x_+ - x_-$ ist minimal
  - Zentrales Intervall: $\int_{-\infty}^{x_-} P(x)\mathrm{d}x = \int_{x_+}^{\infty} P(x)\mathrm{d}x = \frac{1-\alpha}{2}$
- Bei symmetrischen p.d.f. sind alle Intervalle equivalent

Konfidenzintervalle - Deskriptive Statistik (Vorbereitung)

- Zusätzlich gibt es obere/untere Grenzen (*Upper/Lower Limits*)
- Upper Limit

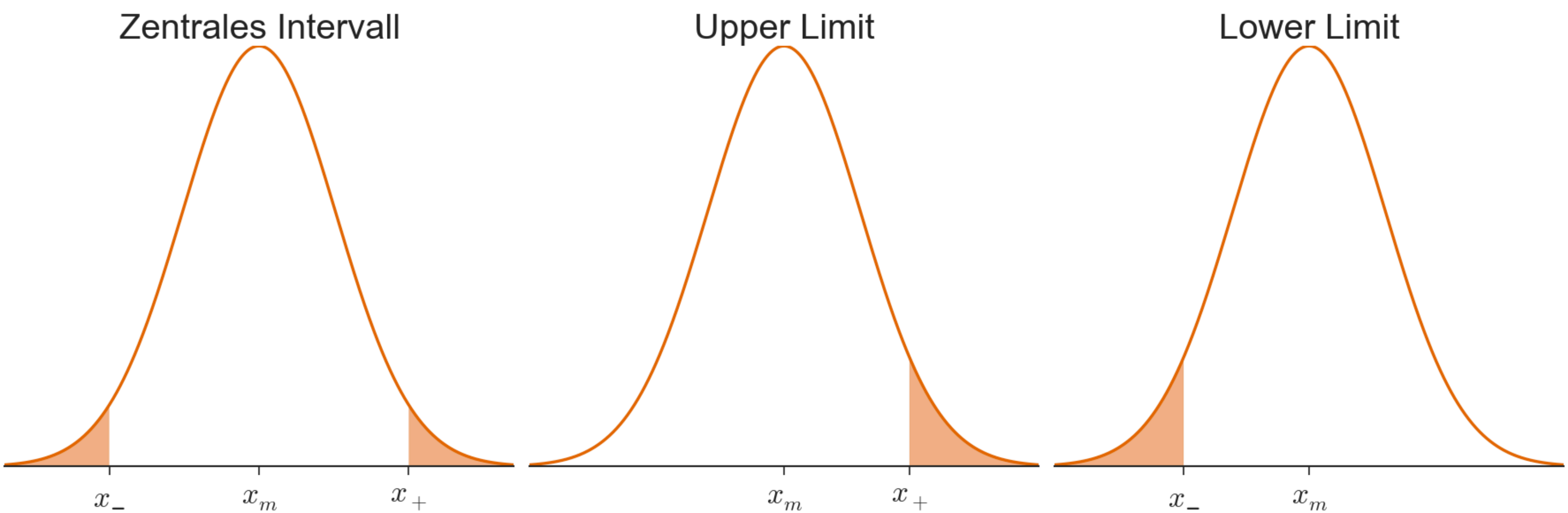
$$P(x < x_+) = \int_{-\infty}^{x_+} P(x)\mathrm{d}x = \alpha$$

- Lower Limit

$$P(x > x_-) = \int_{x_-}^{\infty} P(x)\mathrm{d}x = \alpha$$

- Achtung: Das Upper/Lower Limit ist **NICHT** gleich dem oberen/unteren Ende des Konfidenzintervalls zum selben α, z.B.:
  - 95% zentrales Intervall: 2,5% liegen oberhalb von x_+
  - 95% Upper Limit: 5% liegen oberhalb von x_+

Beispiel – 95% Konfidenzintervall, Upper/Lower Limit



(Abb. 1)

Konfidenzintervalle – Parameterschätzung

- Nun: Parameterabhängige p.d.f.

$$\text{Likelihood: } \mathcal{L}(\theta|x) = P(x|\theta)$$

→ Konstruiere Konfidenzintervall für den unbekannten Parameter θ_{Wahr}

- Für jeden (fixen) Wert θ_0 des Parameters θ gibt es eine von den Daten x abhängige p.d.f. $P(x \mid \theta = \theta_0)$
- Z.B. Gaußverteilung mit bekannten σ und zu schätzendem μ:

$$P(x|\mu) \propto e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

  - Für jedes $\mu = \mu_0$ ergibt sich eine andere p.d.f. und es können Konfidenzintervalle oder Upper/Lower Limits wie in *Abb. 1* konstruiert werden

Konfidenzintervalle – Neyman Konstruktion

- Ziel: Konfidenzintervall für zu schätzende Parameter θ

- **Intervallkonstruktion nach Neyman**
  1. **Vor** der Messung wird für jeden möglichen Wert für $\theta = \theta_0$ das zugehörige Interval für die p.d.f. $P(x \mid \theta = \theta_0)$ bestimmt
     - Für in θ kontinuierliche Variablen geschieht dies auf einem feinen Gitter
  2. Dadurch ergeben sich Konfidenzbänder, die für jeden Messwert x_0 im Vorfeld das jeweilige Konfidenzintervall für θ festlegen
  3. Lies **nach** erfolgter Messung $x=x_0$ das Konfidenzintervall $\theta \in [\theta_1, \theta_2]$ vertikal ab
     - Der obere/untere Schnittpunkt der Vertikalen durch x_0 mit dem konstruierten Konfidenzbändern ist die obere/untere Grenze des Konfidenintervalls für θ

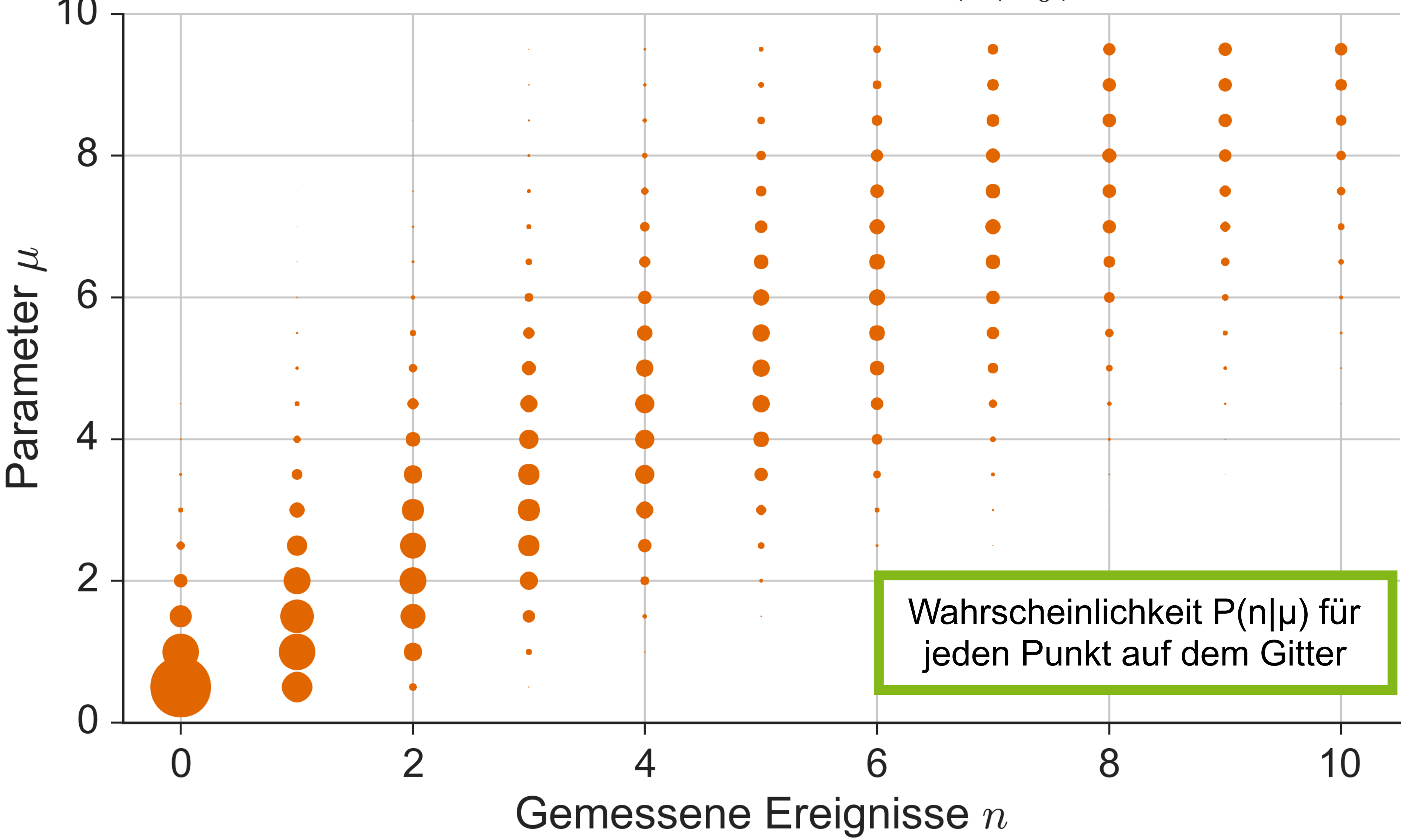
Beispiel – Neyman Konstruktion

- Beispiel mit poissonverteilter p.d.f.

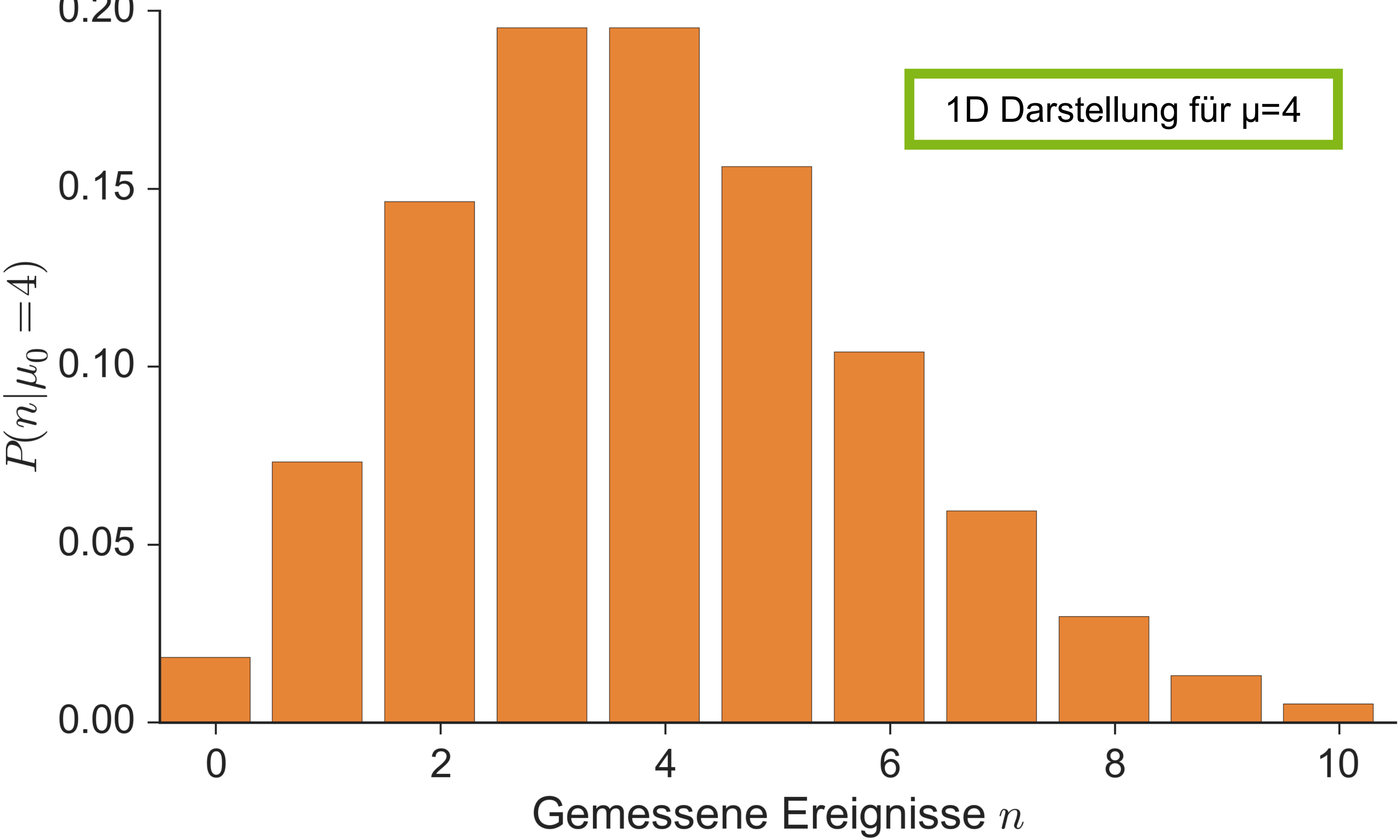
$$P_n = P(n|\mu) = \frac{\mu^n e^{-\mu}}{n!}$$

  - n : Zählrate (diskret)
  - μ : Erwartungswert (kontinuierlich, hier Gitter mit Schrittweite 0,5)
    - *Wird im Realfall feiner gewählt, hier zur besseren Übersicht sehr grob*
  - Konfidenzlevel gewählt als α = 90%

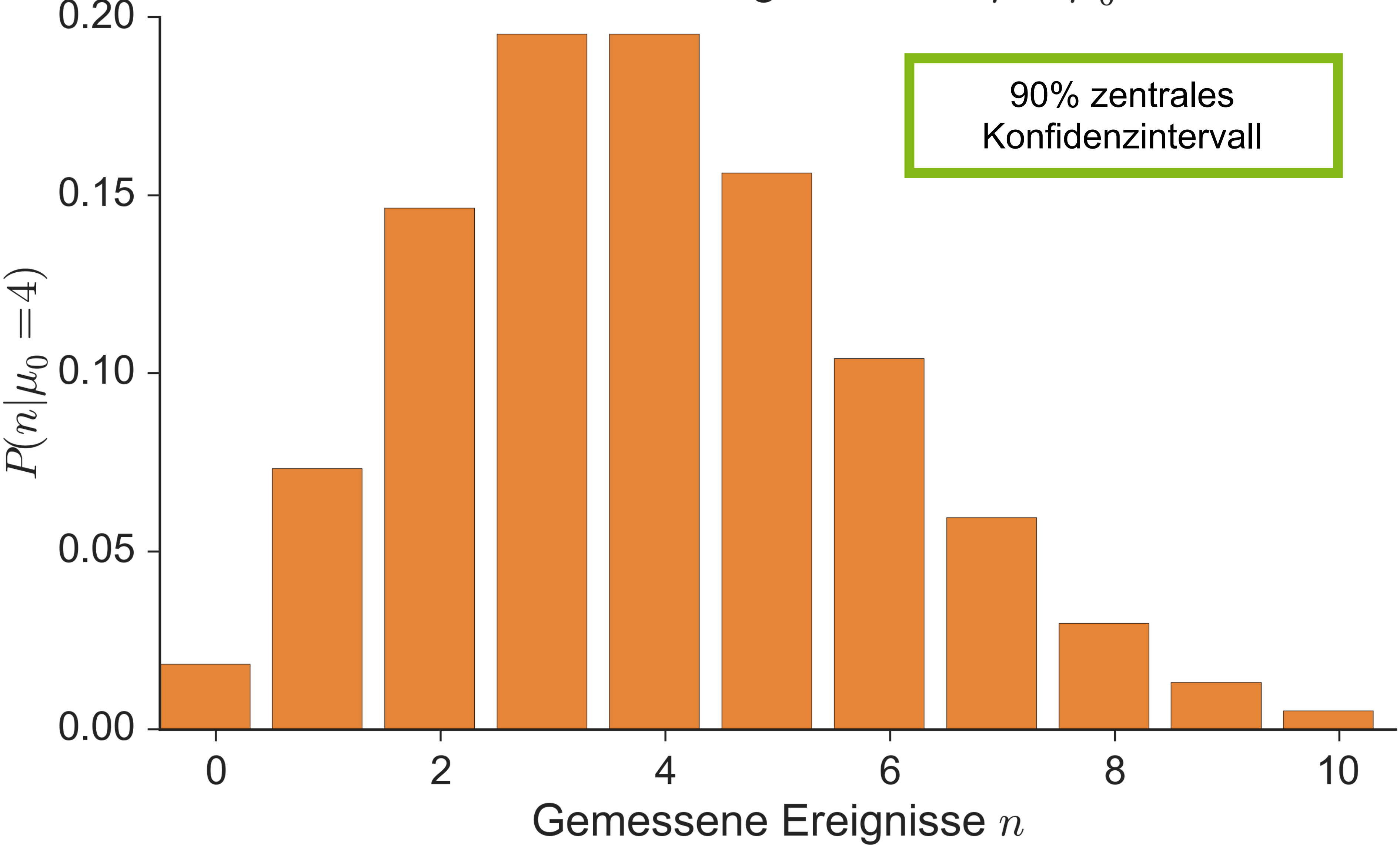
Wahrscheinlichkeiten $P(n|\mu_0)$



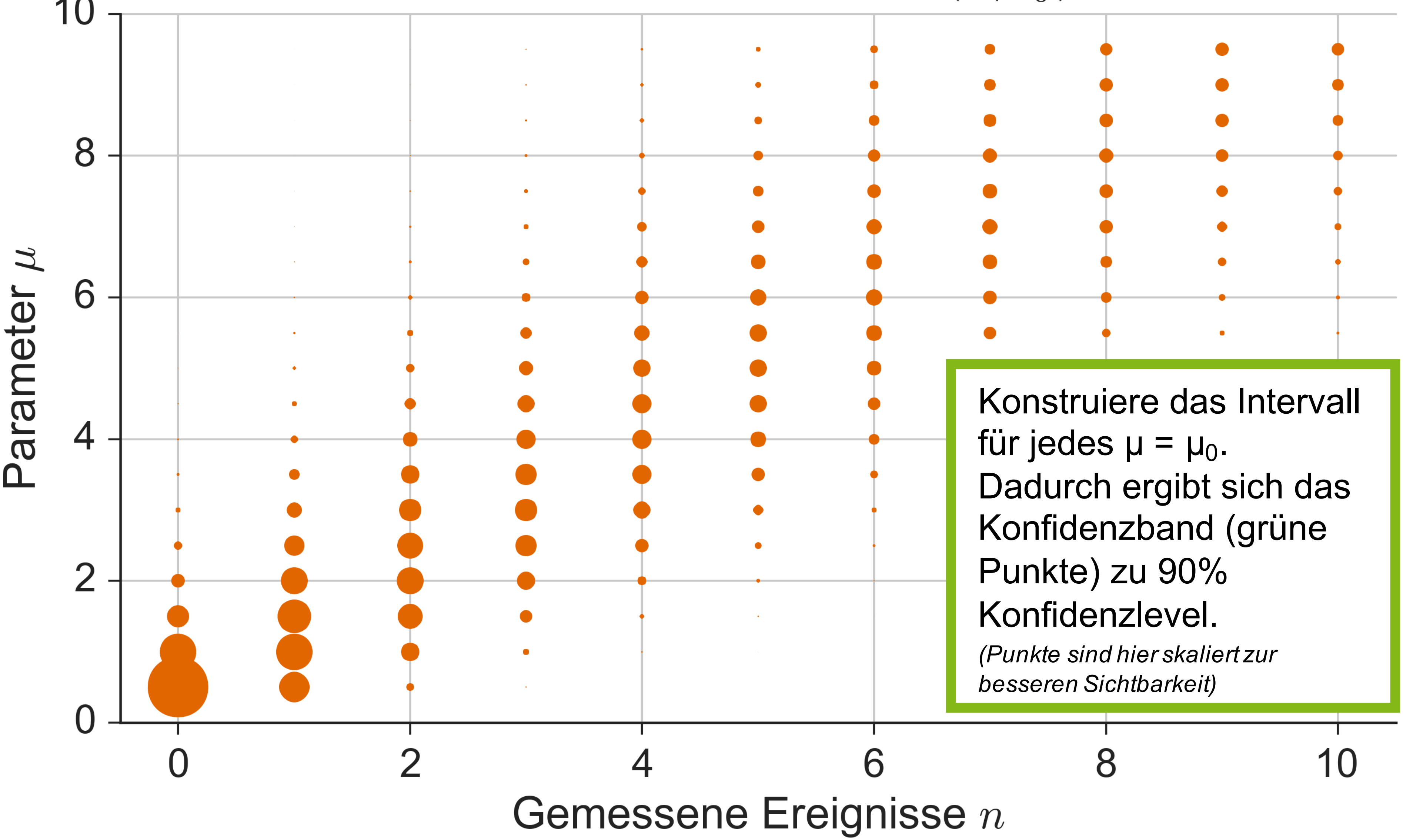
Schnitt durchs Diagramm bei $\mu = \mu_0 = 4$



Schnitt durchs Diagramm bei $\mu = \mu_0 = 4$



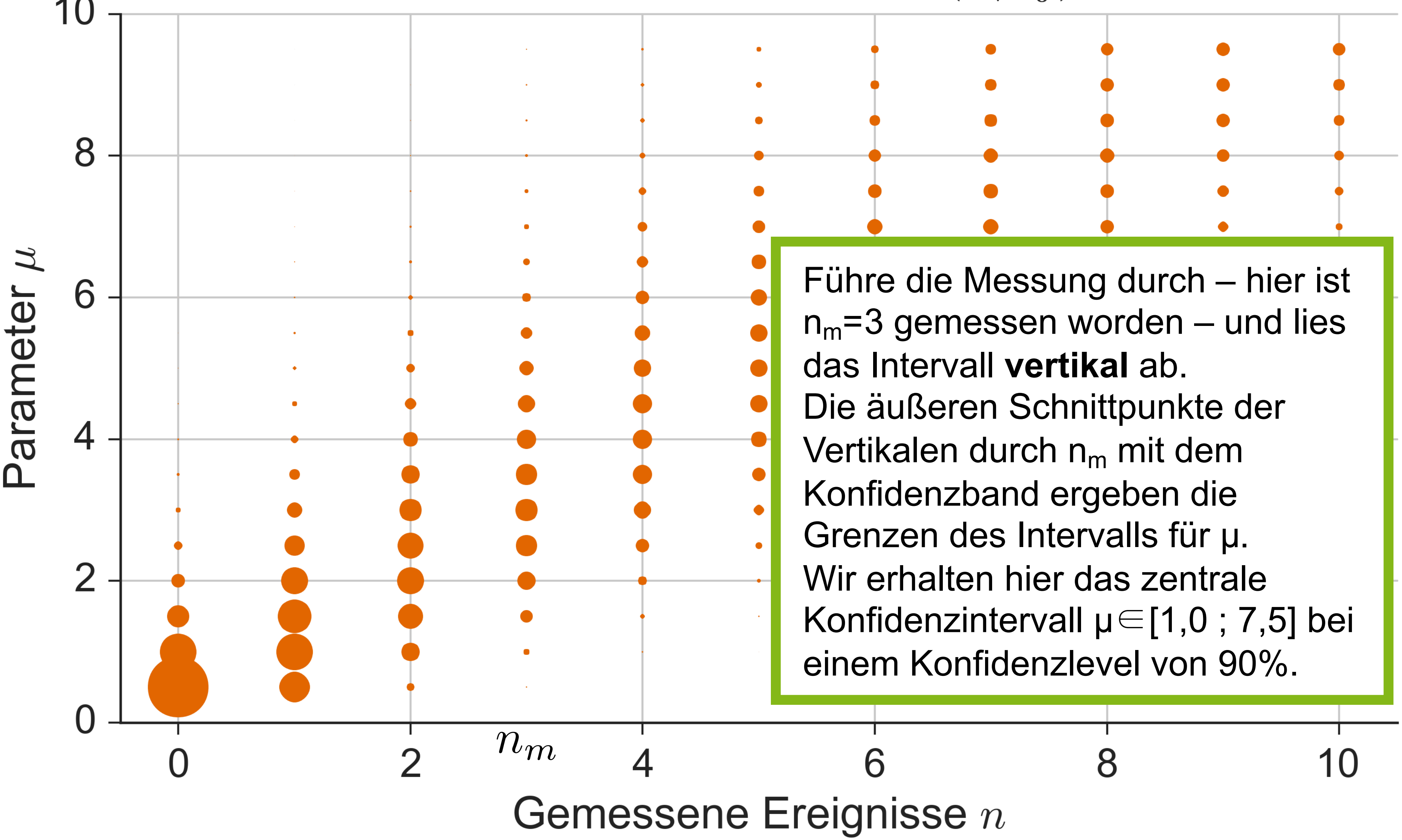
Wahrscheinlichkeiten $P(n|\mu_0)$



Konstruiere das Intervall für jedes $\mu = \mu_0$.
Dadurch ergibt sich das Konfidenzband (grüne Punkte) zu 90% Konfidenzlevel.

(Punkte sind hier skaliert zur besseren Sichtbarkeit)

Wahrscheinlichkeiten $P(n|\mu_0)$



Führe die Messung durch – hier ist n_m=3 gemessen worden – und lies das Intervall **vertikal** ab.
Die äußeren Schnittpunkte der Vertikalen durch n_m mit dem Konfidenzband ergeben die Grenzen des Intervalls für μ.
Wir erhalten hier das zentrale Konfidenzintervall $\mu \in [1{,}0\ ;\ 7{,}5]$ bei einem Konfidenzlevel von 90%.

Konfidenzintervalle – Zusammenfassung Neyman

- Konfidenzbänder werden vor der Messung konstruiert
  - Wähle vorher das Konfidenzlevel α
- Das Konfidenzintervall für θ wird nach erfolgter Messung abgelesen
- Der Anteil α an für verschiedene Mesungen konstruierten Konfidenzbändern enthält dann den wahren Wert θ_{Wahr}
  - Anders ausgedrückt (hier beispielhaft mit zentralem Intervall):
  - $\theta_{Wahr} > \theta_2 \rightarrow$ Wahrscheinlichkeit $(1-\alpha)/2$, den gemessen Wert oder einen Kleineren zu erhalten
  - $\theta_{Wahr} < \theta_1 \rightarrow$ Wahrscheinlichkeit $(1-\alpha)/2$, den gemessen Wert oder einen Größeren zu erhalten
- Bei diskreten Variablen und durch die Notwendigkeit der Diskretisierung überschätzen die Intervalle die Wahrscheinlichkeit: $P > \alpha$
  - Hier nicht zu vermeiden, aber besser als zu unterschätzen

Konfidenzintervalle – Probleme bei beschränkten Parametern

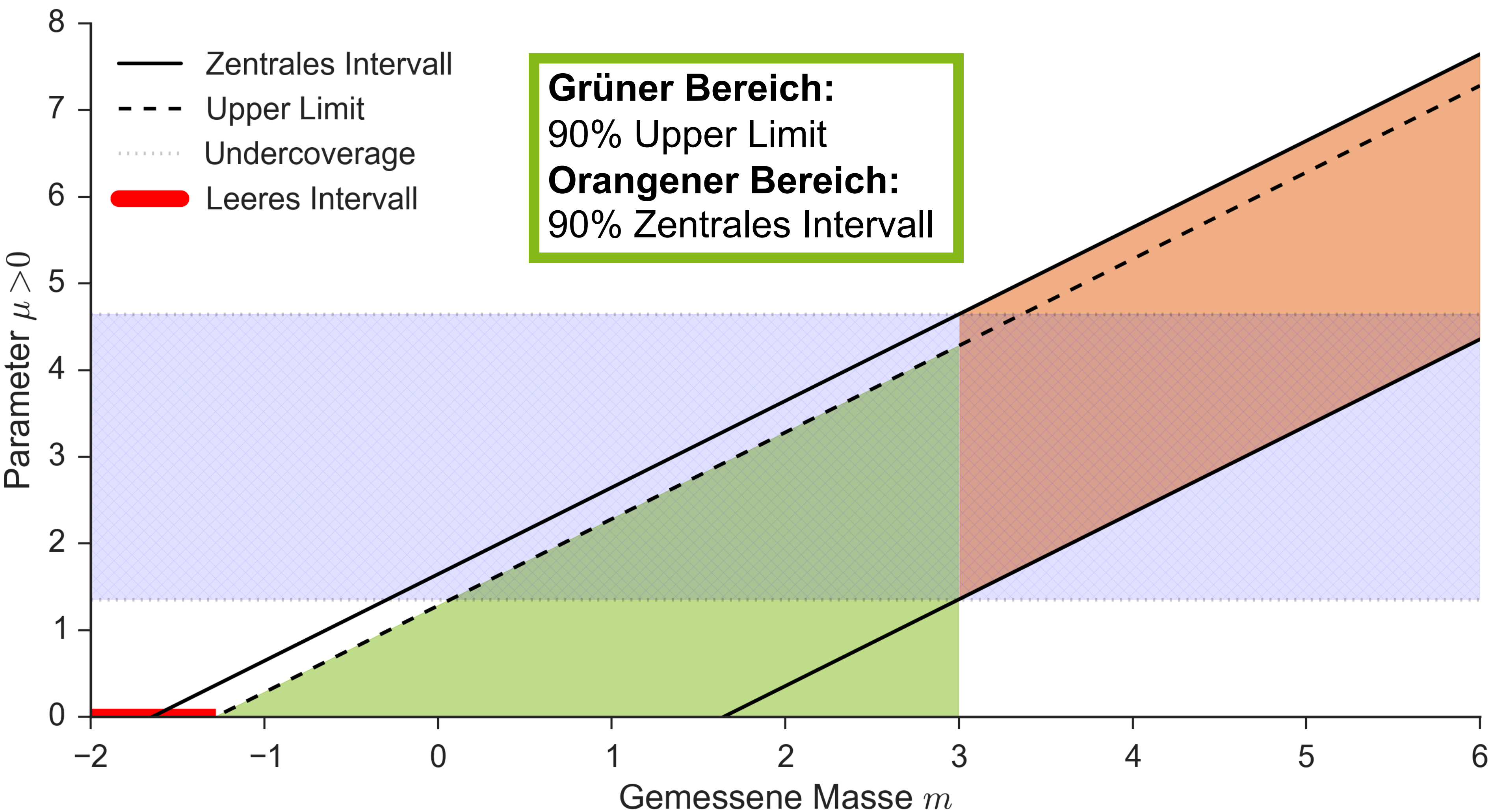
- Problem: Der zu schätzende Parameter ist beschränkt
  - Z.B. Masse >= 0 unter Annahme einer Gaußschen Verteilung
- Das kann zu Problemen führen:

1. **Flip-Flopping**:
   Entscheide anhand der Daten, ob ein Intervall oder Upper Limit verwendet werden soll
   - Führt zu Intervall-Unterschätzung in bestimmten Bereichen
2. **Leere Intervalle**:
   Das konstruierte Konfidenzintervall kann außerhalb des physikalisch erlaubten Bereichs von θ liegen

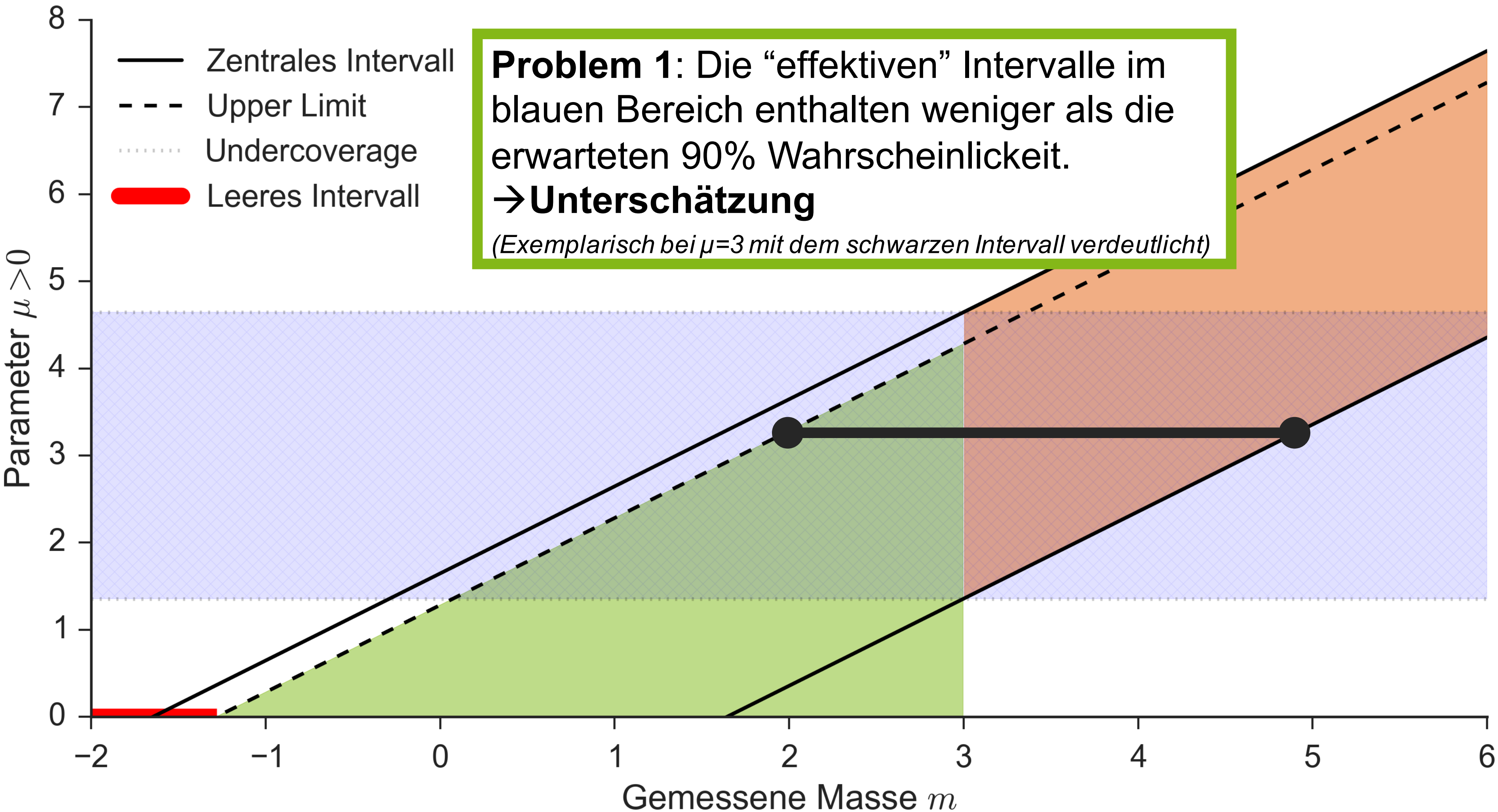
Beispiel – Flip-Flopping und leere Intervalle

- Massen-Messung → Gesuchter Parameter $\mu \geqq 0$
- Annahme: Detektorauflösung ist durch Gaußverteilung mit $\sigma=1$ beschrieben
- Folgende (**FALSCHE**) Idee:
  1. Wenn die Messung mit einer Wahrscheinlichbkeit < 3σ von 0 entfernt liegt, wird ein 90% Upper Limit veröffentlicht
  2. Ansonsten wird das 90% zentrale Konfidenzintervall benutzt
- Siehe Plot auf der nächsten Folie:
  - **Unterschätzung** der Intervalle für θ im blauen Band
  - **Leere Intervalle**, falls $m < -1{,}2$ gemessen wurde

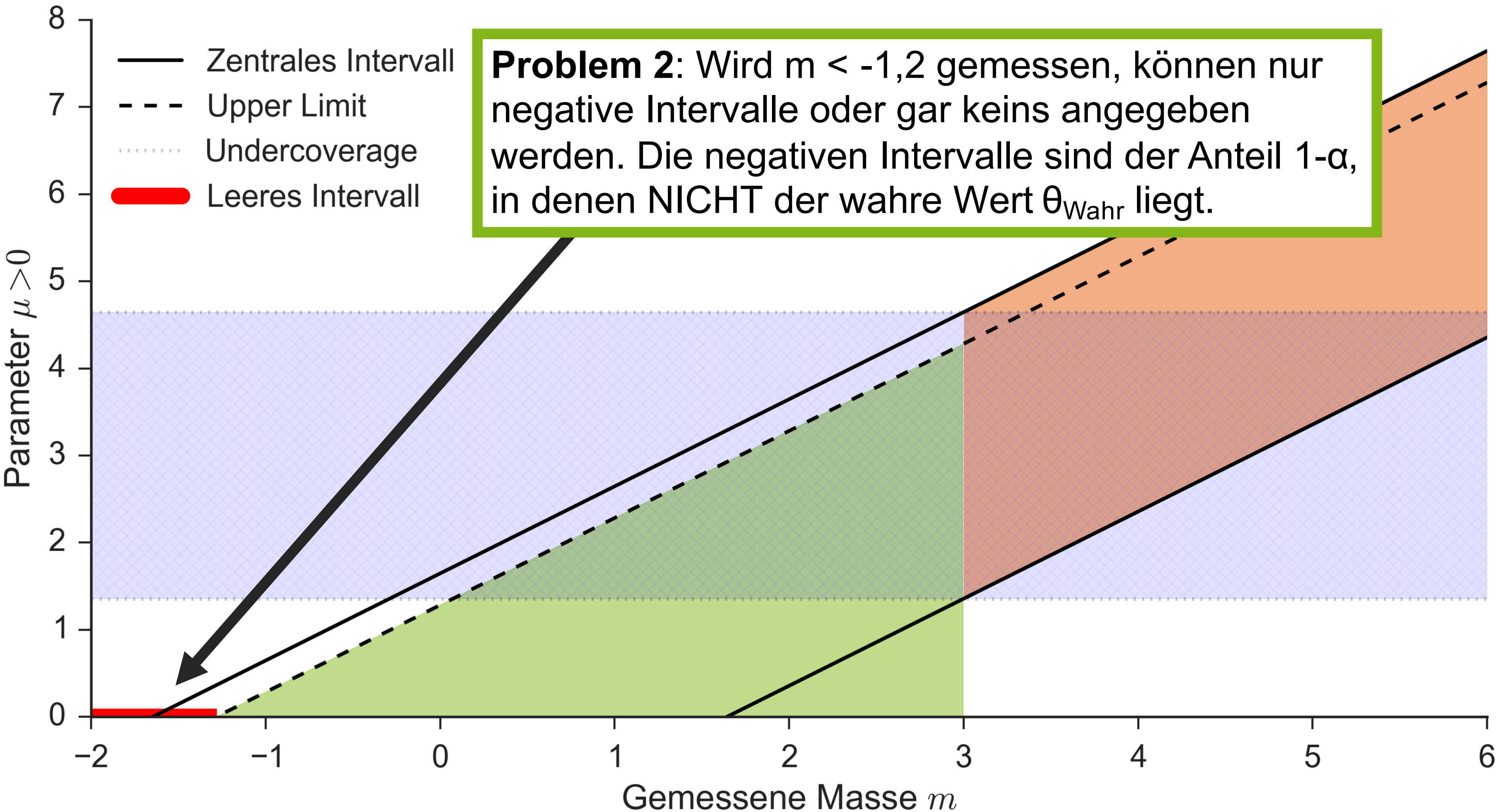
Beispiel – Flip-Flopping und leere Intervalle



Beispiel – Flip-Flopping und leere Intervalle



Beispiel – Flip-Flopping und leere Intervalle



Problem 2: Wird m < -1,2 gemessen, können nur negative Intervalle oder gar keins angegeben werden. Die negativen Intervalle sind der Anteil 1-α, in denen NICHT der wahre Wert θ_{Wahr} liegt.

Konfidenzintervalle – Feldman-Cousins Konstruktion

- Umgeht die Probleme im vorherigen Beispiel
  - Vermeidet Flip-Flopping und leere Intervalle
- Prinzip immer noch wie bei Neyman, aber …
- … verwende Ordnungsprinzip zum Erzeugen der Konfidenzbänder
  - Immer noch frequentistisch
  - Möglichkeit beschränkte Parameter zu berücksichtigen
  - Benutze Likelihood-Verhältnisse zum Ordnen

Konfidenzintervalle – Feldman-Cousins (Fortsetzung)

- Konstruktion des Feldman-Cousins Konfidenzbands

1. Für ein festes $\theta=\theta_0$ berechne wie vorher $P(x|\theta=\theta_0)$ für jeden Wert x
2. Berechne dann das θ_{Best}, welches die Likelihood für den aktuellen Messwert x maximiert: $\max[P(x|\theta)] = P(x|\theta_{Best})$
3. Bilde das Verhältnis R (=*Rang*) der beiden Likelihoodwerte

$$R = \frac{P(x|\theta)}{P(x|\theta_{\mathrm{Best}})}$$

4. Füge die x Werte absteigend ihrem Rang entsprechend dem horizontalen Konfidenzintervall hinzu, bis das gewünschte Konfidenzlevel α erreicht ist
5. Wiederhole für alle Werte von θ
   - Kontinuierliche θ werden diskretisiert

Konfidenzintervalle – Feldman-Cousins (Fortsetzung)

- Vorgehen für kontinuierliche x:
  - Entweder kann x ebenfalls diskretisiert werden, dann können diskrete Werte x dem Rang nach hinzugefügt werden bis α erreicht ist
  - Oder die Gleichung

$$\int_{x_-}^{x_+} P(x|\theta = \theta_0)\mathrm{d}x = \alpha$$

wird für jedes θ_0 numerisch gelöst. Dabei gilt die Nebenbedingung

$$R(x_-) = R(x_+)$$

  - → Siehe Beispiel A mit beschränkter Gaußverteilung

Beispiel A – Gauß mit beschränktem Parameter μ ≧ 0

- Kontinuierlicher Fall: Gaußverteilung mit festem σ und beschränktem μ

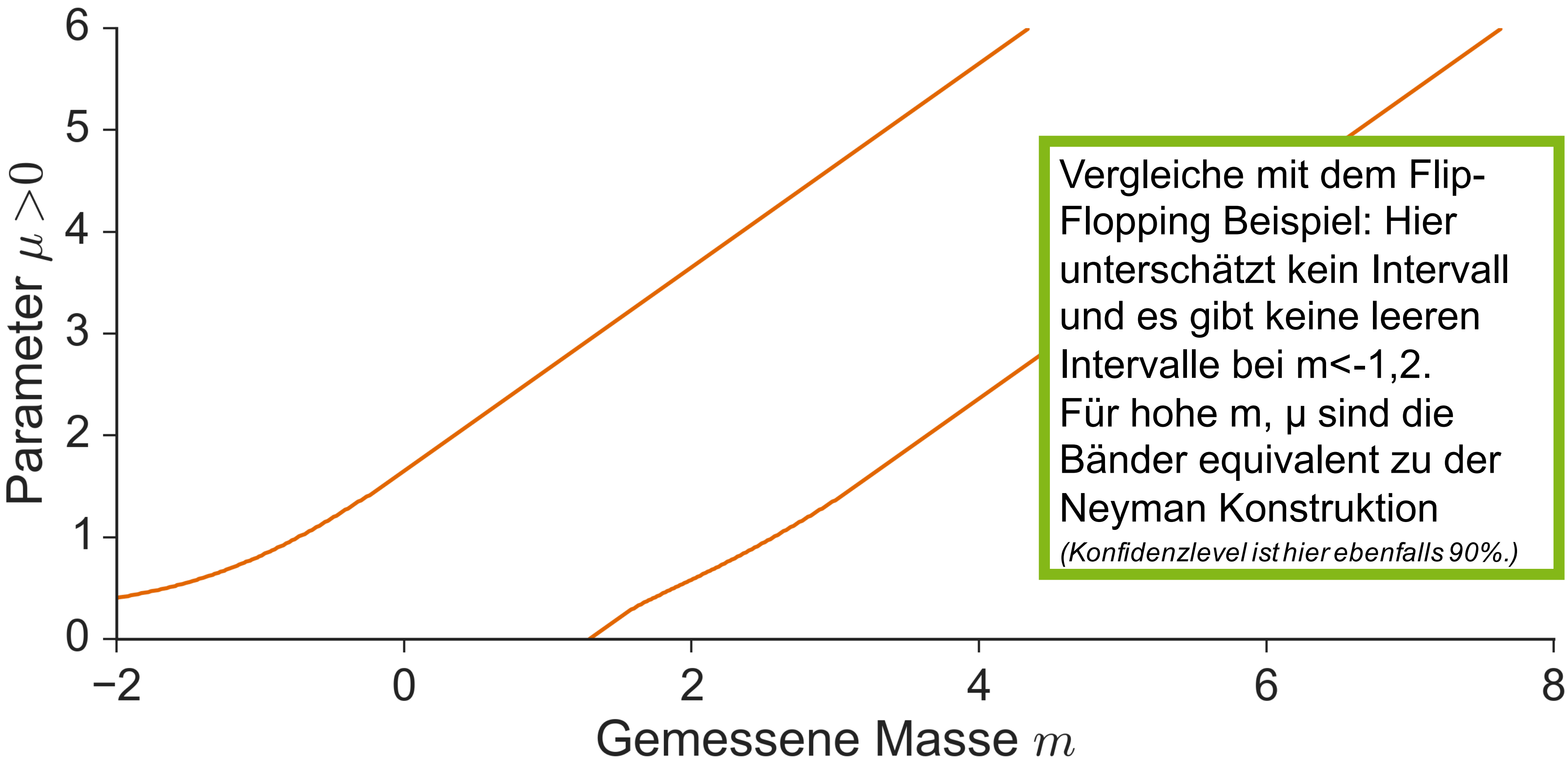
$$P(x|\mu) \propto e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- Die Likelihood wird maximal bei μ_{Best} = max[0 ; x]

$$R = \exp\left(-\frac{(x-\mu)^2}{2}\right) \quad \text{für } x \geq 0$$

$$R = \exp\left(x\mu - \frac{\mu^2}{2}\right) \quad \text{für } x < 0$$

Beispiel A – Gauß mit beschränktem Parameter μ≧0 (Forts.)



Vergleiche mit dem Flip-Flopping Beispiel: Hier unterschätzt kein Intervall und es gibt keine leeren Intervalle bei m<-1,2.
Für hohe m, μ sind die Bänder equivalent zu der Neyman Konstruktion
(Konfidenzlevel ist hier ebenfalls 90%.)

Beispiel B – Poisson mit Untergrund

- Diskreter Fall: Poissonverteilung mit b erwarteten Untergrundereignissen

$$P_n = P(n|\mu) = \frac{(\mu + b)^n e^{-(\mu+b)}}{n!}$$

- Die Likelihood wird maximal bei μ_{Best} = max[0 ; n-b]

$$R = \left(\frac{\mu + b}{b}\right)^n e^{-\mu} \quad \text{für } n \leq b$$

$$R = \left(\frac{\mu + b}{n}\right)^n e^{-(\mu+b-n)} \quad \text{für } n > b$$

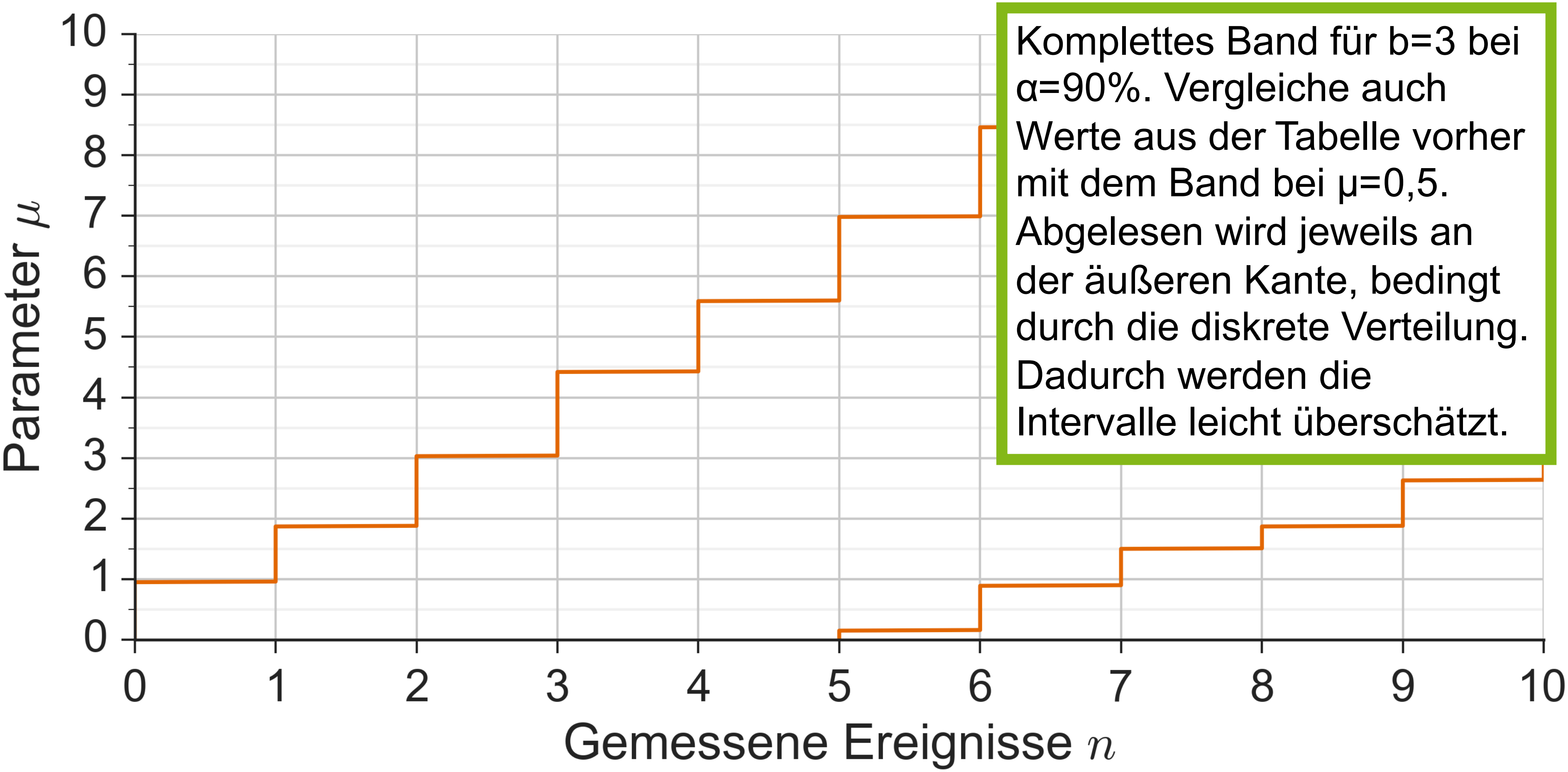
Beispiel B – Poisson mit Untergrund (Fortsetzung)

- Hier Beispielrechnung für μ_0=0,5 und b=3
  - Die Haken markieren die Werte, die für ein 90% Konfidenzlevel benutzt werden → konservatives Intervall wegen der diskreten Verteilung

n	$P(n\|\mu_0)$	μ_{Best}	$P(n\|\mu_{Best})$	R	Rang	90% CL
0	0,030	0	0,050	0.607	6	✔☐
1	0,106	0	0,149	0,708	5	✔☐
2	0,185	0	0,224	0,826	3	✔☐
3	0,216	0	0,224	0,963	2	✔☐
4	0,189	1	0,195	0,966	1	✔☐
5	0,132	2	0,175	0,753	4	✔☐
6	0,077	3	0,161	0,480	7	✔☐
7	0,039	4	0,149	0,259	8	

$$\sum_{n=0}^{6} P(n|\mu_0) = 0,935$$

Beispiel B – Poisson mit Untergrund (Fortsetzung)



Komplettes Band für b=3 bei α=90%. Vergleiche auch Werte aus der Tabelle vorher mit dem Band bei μ=0,5. Abgelesen wird jeweils an der äußeren Kante, bedingt durch die diskrete Verteilung. Dadurch werden die Intervalle leicht überschätzt.

Konfidenzintervalle

- Varianzen, Momente
  - Unabhängig von der zugrunde liegenden Verteilung
  - Wohldefinierte Fehlerfortpflanzung
  - Fehlerkombination unabhängiger Messungen
- Konfidenzintervalle
  - Alternative Methode um Unsicherheiten anzugeben
  - Abhängig von der zugrundeliegenden Verteilung
  - Fehlerpropagation und Kombination schwierig
  - Frequentistisch ⇔ **Bayesisch**

Konfidenzintervalle – Bayesische Definition

- Bayesische Konfidenzintervalle werden als Kredibilitätsintervalle bezeichnet
- Aus der Posterior p.d.f. lassen sich Intervalle bestimmen, die den Parameter θ mit der Wahrscheinlichkeit α enthalten
- Ein Intervall für das

$$\int_R \mathrm{d}\theta p(\theta|x) = \alpha$$

  gilt, wird α-Kredibilitätsintervall genannt
- Es gibt viele Möglichkeiten solche Intervalle zu konstruieren, oft wird das kürzeste Intervall gewählt

highest posteriori density region

- Das kürzeste Intervall wird auch als *highest posteriori density region* (HPD) bezeichnet und lässt sich mit der Bedingung

$$p(\theta|x) \geq p(\theta^*|x), \forall\theta \in R \wedge \forall\theta^* \notin R$$

  konstruieren
- Die HPD-Region muss nicht zwangsweise ein zusammenhängendes Intervall sein
- Interpretation von Bayesischen Konfidenzintervallen:

> Das Intervall enthält mit einer Wahrscheinlichkeit α
> den wahren Wert des Parameters θ_{t}

Wahl des Priors

- Bei der Auseinandersetzung mit einem Problem nimmt die bekannte Information über das Problem zu
- Ist bereits eine Posterior p.d.f. vorhandenen, kann diese in den meisten Fällen als neue Prior p.d.f. verwendet werden

- Liegt wenig bis keine Information vor, ist die Wahl des Priors schwieriger
  - Gleichverteilter Prior
  - Jeffreys Prior

Gleichverteilter Prior

- Wird aufgrund mangelnder Informationen eine Gleichverteilung als Prior gewählt entspricht die Posterior p.d.f. bis auf die Normierung der Likelihood

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{\int \mathrm{d}\theta\, p(x|\theta)p(\theta)} \propto p(x|\theta)$$

- Der Prior wird dann bei bekannten Grenzen zu

$$p(\theta) = \mathrm{const.} = \frac{1}{x_{\max} - x_{\min}}$$

- Probleme:
  - Die Gleichverteilung ist auf unbeschränkten Intervallen nicht normierbar
    → Problematisch beim Vergleich verschiedener Modelle
  - Die Gleichverteilung ist nicht invariant gegenüber Reparametrisierungen

Jeffreys Prior

- Der Jeffreys Prior $p(\theta) \propto 1/\theta$ ist invariant gegenüber der Transformation $\theta \to \alpha\theta$
- Sind obere und untere Grenze von θ bekannt, kann der Prior durch

  $$p(\theta) = \frac{1}{\theta \ln(\theta_{\text{max}}/\theta_{\text{min}})}$$

  normiert werden. Ohne Grenzen ist dieser Prior ebenfalls nicht normierbar.

- Der Jeffreys Prior liefert die gleiche Wahrscheinlichkeit pro logarithmischem Intervall
  - Beispiel: Eine gesuchte Größe θ, mit bekannten Grenzen $\theta_{\text{min}} = 10^{-1}$ und $\theta_{\text{max}} = 10^3$ liegt durch einen gleichverteilten Prior mit der höchsten Wahrscheinlichkeit zwischen 10^2 und 10^3. Mit einem Jeffreys Prior sind alle Dekaden gleich wahrscheinlich.

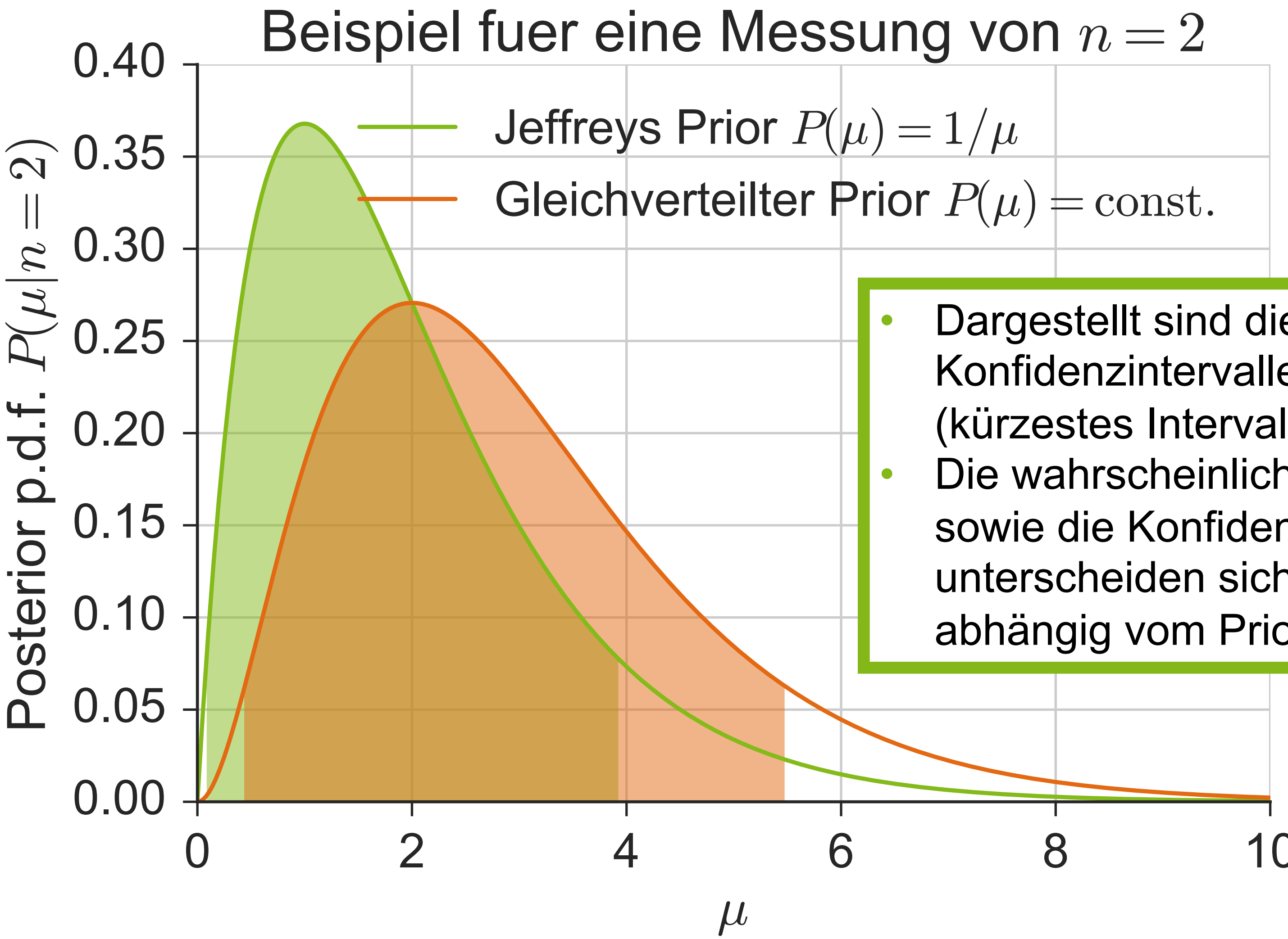
Einfluss verschiedener Priors

- Ein Zählexperiment beobachtet n Ereignisse
- → n ist eine poisson-verteilte Zufallsvariable
- Die Wahrscheinlichkeit P_n, dass n Ereignisse beobachtet werden ist dann:

$$P_n = P(n;\mu) = \exp(-\mu)\frac{\mu^n}{n!}$$

- Was kann je nach Wahl des Priors über den Erwartungswert und eine Konfidenzregion ausgesagt werden?
- Nach der Wahl der Priors lässt sich die Posterior p.d.f. wieder wie folgt bestimmen:

$$P(\mu|n) = \frac{P(n|\mu)P(\mu)}{\int \mathrm{d}\mu P(n|\mu)P(\mu)}$$



- Dargestellt sind die 90% Konfidenzintervalle (kürzestes Intervall)
- Die wahrscheinlichsten Werte sowie die Konfidenzintervalle unterscheiden sich deutlich abhängig vom Prior

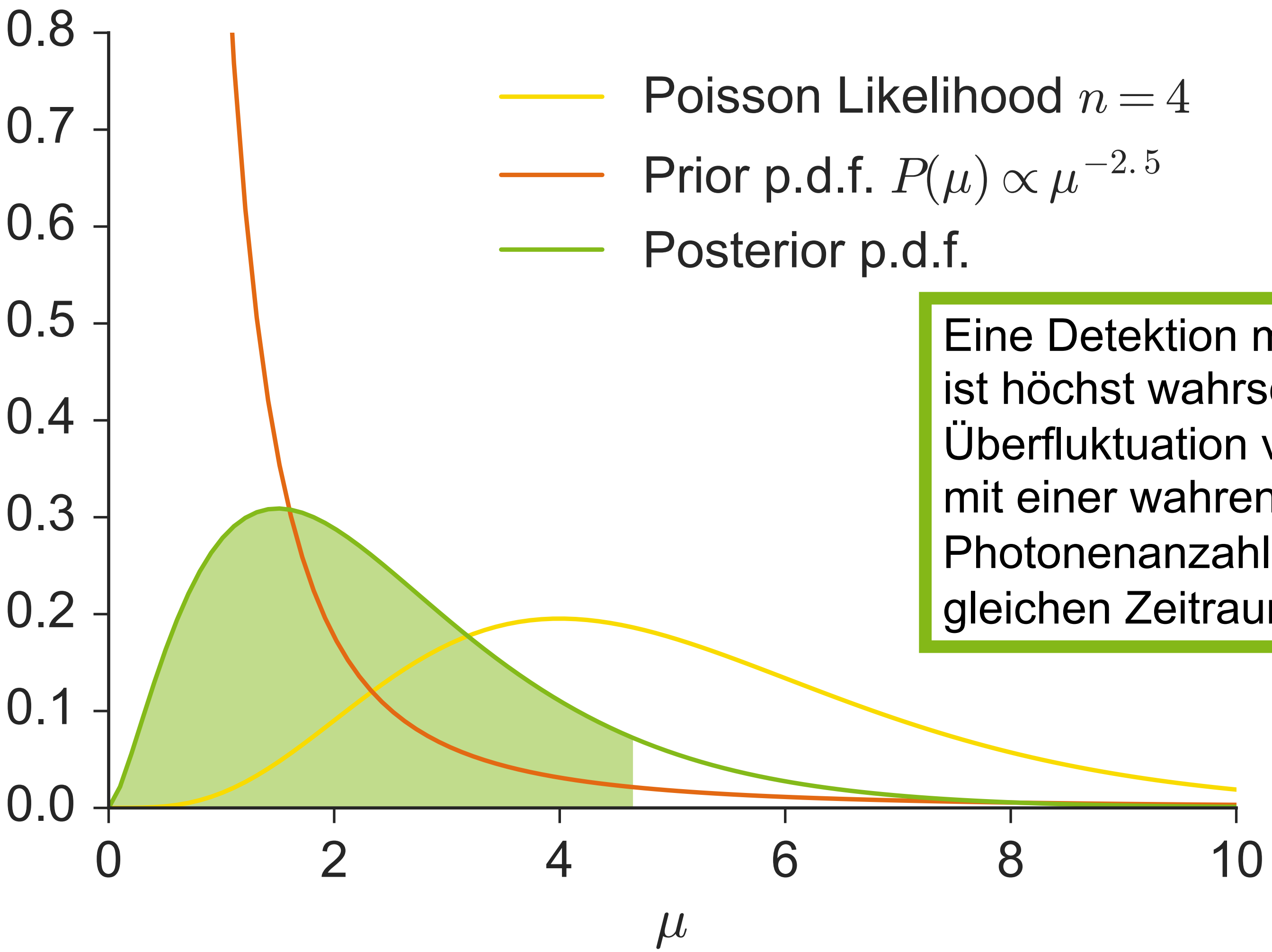
Anwendungsbeispiel – *x-ray survey*

- Mit dem *Chandra X-Ray Observatory* (Satellit mit einem Röntgenteleskop) wurden in einer Untersuchung sehr schwache Quellen mit 2-4 Photonen über einen Messzeitraum von 5000s beobachtet

  (Kenter et al. 2005, ApJS 161, 9)

- Es ist den Astronomen bekannt gewesen, dass die Anzahl von Quellen N mit einem gewissen Fluss oberhalb von S einem Potenzgesetz folgt: $N(S) \sim S^{-\beta}$ mit $\beta = 2.5$

- Die gemessenen Photonen aus einer Quelle entsprechen nicht dem besten Schätzer für den wahren Mittelwert der Quellphotonen



Eine Detektion mit 4 Photonen ist höchst wahrscheinlich eine Überfluktuation von einer Quelle mit einer wahren mittleren Photonenanzahl von 1.5 im gleichen Zeitraum

TESTEN VON HYPOTHESEN

Hypothesentests - Einleitung

- Es liegen zwei Hypothesen über die Daten vor
- Welche erklärt die Daten besser?
- Die Hypothesen werden oft mit H_0 und H_1 bezeichnet
  - H_0: Nullhypothese
  - H_1: Alternative Hypothese
- Effektiv geht es um „Ja“ / „Nein“ Entscheidungen
  - Ist das ein neues Teilchen oder nur Untergrund?
  - Ist die neue Medizin wirksam oder nicht?
  - Hat der Mensch Schuld an der globalen Erwärmung oder nicht?

Hypothesentests – Generelles Vorgehen

1. Definiere beide Testhypothesen
   - Hier: **“Simple” Hypothesen** (*siehe auch: Neyman-Pearson Test*)
     → Eine Hypothese muss komplett durch eine einzige (parameterabhängige) p.d.f. beschrieben werden können
2. Definiere eine Test-Statistik, auf der ein numerischer Test durchgeführt werden kann und eine Signifikanz des Tests
3. Wähle Verwerfungskriterien basierend auf der Test-Statistik
   - Ab einem bestimmten Wert der Teststatistik, wird die eine, oder die andere Hypothese verworfen
     - Der Wert, bei dem verworfen wird, heißt oft “*kritischer Parameter”*
   - Versuche Typ I / Typ II Fehler gering zu halten

- **Was genau ist zu tun? Wir schlüsseln das im Detail auf**

Hypothesentests – Typ I und Typ II Fehler

- Wir haben zwei Hypothesen aufgestellt, Beispiel:
  - Nullhypothese: Wir sehen keinen Peak in den Daten
  - Alternative Hypothese: Wir sehen einen Peak in den Daten
- Typ I Fehler
  - Der Test sagt, da ist ein Peak, obwohl es nicht stimmt
  - Schreibweise: Typ I Fehler treten mit der Rate α auf
  - Typ I Fehler werden auch Signifikanz (*Significance*) genannt
- Typ II Fehler
  - Der Test sagt, es gibt keinen Peak, obwohl einer da ist
  - Schreibweise: Typ II Fehler treten mit der Rate β auf
  - $(1-\beta)$ wird auch Trennkraft (*Power) genannt*
- Ziel: Minimiere sowohl α, als auch β

Hypothesentests – Typ I und Typ II Fehler (Fortsetzung)

	H_0 wahr	H_0 falsch
H_0 nicht abgelehnt	***True positive*** $P = 1\text{-}\alpha$	***Typ II Fehler*** $P = \beta$
H_0 abgelehnt	***Typ I Fehler*** $P = \alpha$ (Signifikanz)	***True Negative*** $P = 1\text{-}\beta$ (Trennkraft)

- **Vorsicht**: Signifikanz nicht verwechseln mit dem Konfidenzlevel α aus den Konfidenzintervallen.

Hypothesentests – Typ I und Typ II Fehler (Fortsetzung)

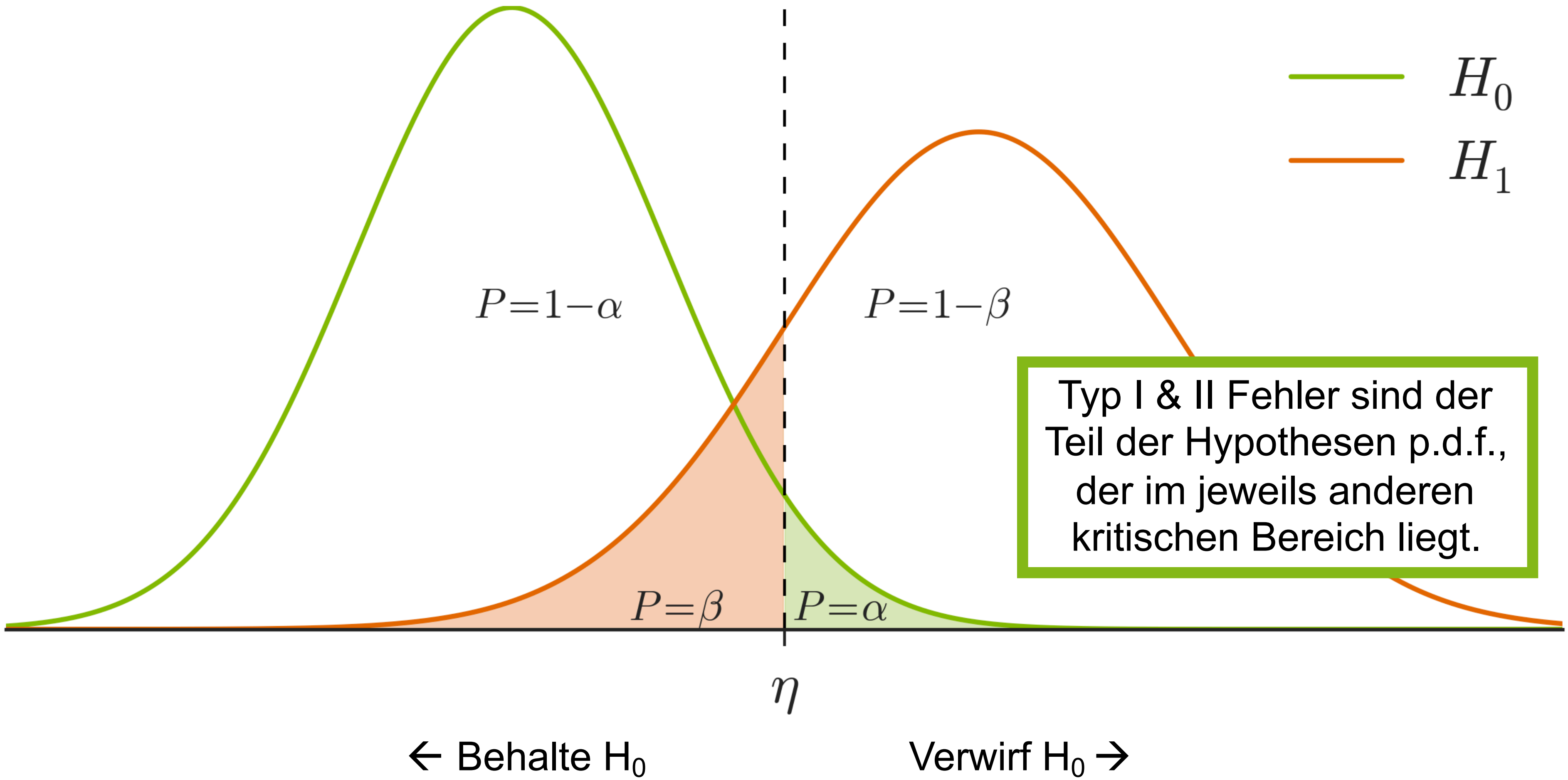
- Das Verwerfungskriteriun teilt den Hypothesenraum in zwei Bereiche
  - Region, in der H_0 verworfen wird ("*Verwurfsregion*")
  - Region, in der H_1 verworfen wird ("*Akzeptanzregion*")
- Der Typ I Fehler α ist der Teil der p.d.f. der Nullhypothese, welcher in der *Verwurfsregion* liegt

$$\alpha = \int_{\text{Verwurfsregion}} P_{\mathrm{H}_0} \mathrm{d}V$$

- Der Typ II Fehler β ist der Teil der p.d.f. der Alternativhypothese, welcher in der *Akzeptanzregion* liegt

$$\beta = \int_{\text{Akzeptanzregion}} P_{\mathrm{H}_1} \mathrm{d}V$$

Hypothesentests – Typ I und Typ II Fehler (Fortsetzung)



Typ I & II Fehler sind der Teil der Hypothesen p.d.f., der im jeweils anderen kritischen Bereich liegt.

Hypothesentests - Neyman-Pearson Test

- **Wähle** eine Signifikanz α, **bevor** der Test durchgeführt wird
- Neyman-Pearson: Die Test-Statistik **Γ** ist definiert durch

$$\Gamma(x) = \frac{P(x|H_0)}{P(x|H_1)}$$

  - Wähle nun einen kritischen Parameter η zur Signifikanz α, sodass

$$P(\Gamma(x) \leq \eta | H_0) = \alpha$$

  - Lehne H_0 ab, wenn gilt

$$\Gamma(x) \leq \eta$$

- Dieser Test hat zu einem gegebenen α die höchste Trennkraft (1-β)

Hypothesentests - Neyman-Pearson Test (Fortsetzung)

- Die Wahrscheinlichkeiten P(x|H) sind die **Likelihood Funktionen** für die jeweiligen **simplen Hypothesen**
  - P.d.f. sind komplett beschrieben durch Parameter H_0: $\theta \in \theta_0$ und H_1: $\theta \in \theta_1$

$$\Gamma(x) = \frac{\mathcal{L}(\theta_0|X)}{\mathcal{L}(\theta_1|x)}$$

- $\Gamma(x)$ folgt je nach Hypothesen einer bestimmten Verteilung
  - Entweder analytisch bestimmen
  - Oder numerisch durch MC Pseudo-Experimente

Hypothesentests - Neyman-Pearson Test (Fortsetzung)

- Die Region im Hypotheseraum, die den Fehler zweiter Art zu gegebenem α minimiert ist eine Kontur des Likelihood Quotienten

$$\Gamma(x) = \frac{\mathcal{L}(\theta_0|X)}{\mathcal{L}(\theta_1|x)}$$

- Der Quotient folgt einer eindimensionalen Verteilung $\mathbf{\Gamma}$(x)
  - **Beachte:** Der Hypothesenraum selbst kann mehrdimensional sein
- Der kritische Parameter η, kann dann über die Verteilung vom $\mathbf{\Gamma}$(x) bestimmt werden
  - Z.B. durch tabellierte Quantile bekannter Verteilungen im analytischen Fall
- Nachdem das η zu gegebenem α gefunden ist, kann der Test abgelehnt oder akzeptiert werden

Hypothesentests - Bemerkung zu Hypothesen

- Wichtig: Hypothesen werden durch Tests **NICHT** bewiesen.
  → Es kann nur gezeigt werden, dass die Daten mit der Alternative nicht konsistent sind

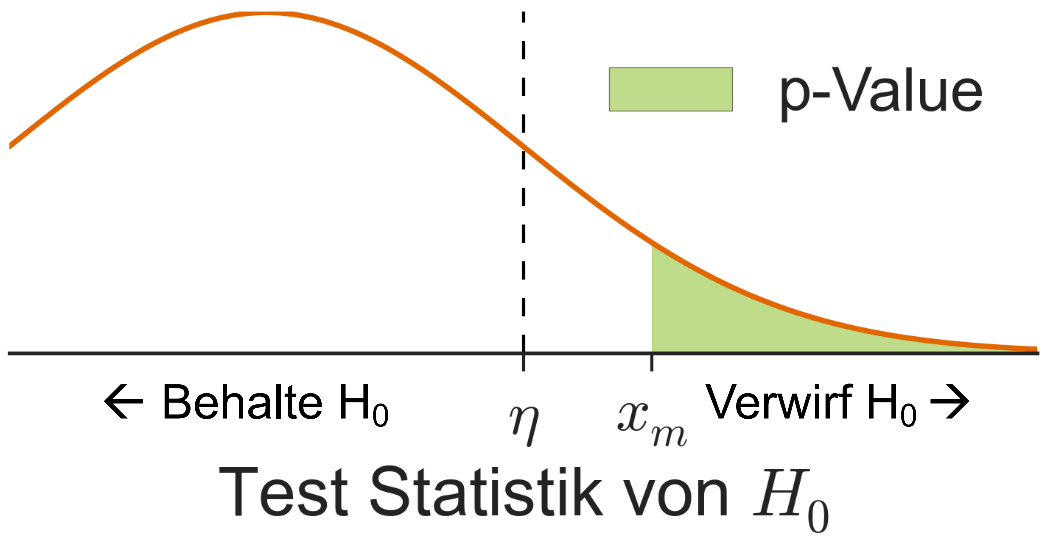
- Wie zeige ich, dass ein Effekt zu sehen ist?
  - Stelle das Gegenteil als Nullhypothese auf → H_0: Ich sehe keinen Peak
  - Die Alternative ist H_1: Ich sehe einen Peak
  - Wenn der Test fehlschlägt, ist die Nullhypothese abgelehnt
    → Das heißt **NICHT**, dass die Alternative wahr ist

- R. Barlow:
  „In statistics one cannot meaningfully accept a hypothesis: one can only reject them"

Hypothesentests - p-Values

- **Ein p-Value ist die Wahrscheinlichkeit die gemessenen Daten zu erhalten, unter der Annahme, dass die Nullhypothese richtig ist**

- Keine neuen Informationen. Entscheidend ist der kritische Wert η
  - Liegt die Teststatistik einer Messung x_m über η, dann wird H_0 abgelehnt und es ist $p < \alpha$



STATISTISCHE TESTS

Likelihood-Quotienten Test

- Neyman-Pearson Test für simple Hypothesen (siehe vorheriger Abschnitt)
- Funktioniert auch mit komplementären Hypothesen
- Beschreibe H_0 und H_1 durch Likelihood Funktionen mit Parametern θ
  - H_0: $\theta \in \Theta_0$ und H_1: $\theta \in \Theta_1 = \Theta_0^C$, wobei $\Theta_1 = \Theta_0^C$ das Komplement zu Θ_0 in Θ ist
  - Beachte: Die Parameter müssen nicht festliegen, sondern können z.B. über die Daten bestimmt werden
- Der Test ist dann wie folgt definiert

$$\Gamma(x) = \frac{\sup_{\theta \in \Theta_0} \mathcal{L}(\theta|x)}{\sup_{\theta \in \Theta} \mathcal{L}(\theta|x)}$$

  - Zähler: Maximale Likelihood für θ aus den möglichen Werten Θ_0 aus H_0
  - Nenner: Maximale Likelihood für θ aus dem gesamten Parameterraum Θ

Likelihood-Quotienten Test (Fortsetzung)

- **Anschaulich:**
  Je größer der Quotient, desto wahrscheinlicher ist die Nullhypothese. Das Maximum im gesamten Parameterraum Θ wird dann bei einem Wert ähnlich dem Maximum im Raum der Nullhypothese Θ_0 erreicht

- Der kritische Wert wird wie vorher gewählt, sodass

$$\sup_{\theta \in \Theta_0} P_\theta(\Gamma(x) < \eta) = \alpha$$

- Die Nullhypothese wird zur Signifikanz α abgelehnt, wenn gilt

$$\Gamma(x) < \eta$$

Likelihood-Quotienten Test – Wilks' Theorem

- Wenn gilt, dass
  1. sich die Nullhypothese durch eine lineare Parameter-Transformation als ein Spezialfall der Alternativ-Hypothese darstellen lässt
  2. die Anzahl der Beobachtungen gegen unendlich geht
- Dann ist die Teststatistik

$$-2\ln(\Gamma(x))$$

χ^2 verteilt

- Die Anzahl der Freiheitsgrade ist die Differenz der Dimensionalität von Θ und Θ_0

→ Sehr hilfreich um eine analytische Abschätzung des kritischen Wertes η aus einer bekannten Verteilung zu erhalten

Beipiel – Likelihood-Quotienten Test

- In der Mensa soll pro Schale im Mittel μ_0=10 [a.u.] Pudding gefüllt werden
- Die Abfüllmaschine teilt eine normalverteilte Menge mit bekannter Varianz σ^2 aus
- Es soll anhand von n Stichproben getestet werden, ob die Maschine ordentlich abfüllt
  - Nullhypothese: Die Maschine ist in Ordnung → H_0: $\mu = \mu_0 = 10$
  - Alternativhypothese: Die Maschine ist nicht in Ordnung → H_1: $\mu \neq \mu_0$, $\mu \in \Theta$
- P.d.f.

$$\mathcal{L}(\mu|x) = P(x|\mu) \propto e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- Test-Statistik: μ_{Best} ist das $\mu \in \Theta$, welches die Likelihood maximiert

$$\Gamma(x) = \frac{\mathcal{L}(\mu = \mu_0|x)}{\sup \mathcal{L}(\mu \in \Theta|x)} = \frac{\mathcal{L}(\mu = \mu_0|x)}{\mathcal{L}(\mu_{\text{Best}}|x)}$$

Beipiel – Likelihood-Quotienten Test (Fortsetzung)

- Umformen liefert die Test-Statistik

$$\Gamma(x) = \exp\left(-\frac{n}{2\sigma^2}(\mu_{\mathrm{Best}} - \mu_0)^2\right)$$

- Der Test wird für ein k_α zur Signifikanz α verworfen, wenn gilt

$$\Gamma(x) = \exp\left(-\frac{n}{2\sigma^2}(\mu_{\mathrm{Best}} - \mu_0)^2\right) \leq k_\alpha$$

- Umformen liefert

$$\frac{|\mu_{\mathrm{Best}} - \mu_0|}{\sqrt{\sigma^2/n}} \geq k_\alpha^*$$

Beipiel – Likelihood-Quotienten Test (Fortsetzung)

- Die Größe $z = |\mu_{Best} - \mu_0| / \sqrt{(\sigma^2/n)}$ ist standardnormalverteilt
- Damit kann der kritische Wert η zur Signifikanz α aus einer Tabelle abgelesen werden
- Z.B. wird der zweiseitige Test zur Signifikanz α = 0,05 verworfen, wenn gilt

$$\frac{|\mu_{\mathrm{Best}} - \mu_0|}{\sqrt{\sigma^2/n}} \geq z_{0,025} \geq 1,96$$

- *Bemerkung*:
  - Hier ist μ_{Best} das arithmetische Mittel der Messdaten
    → erhält man durch Ableitung der Likelihood Funktion

Gauß-, t- und F-Test

- Im Beispiel gesehen: Test-Statistik ist standardnormalverteilt
  → Likelihood-Quotienten Test liefert weitere, bekannte Tests für verschiedene Spezialfälle, z.B.
  - Gauß-Test
  - t-Test
  - F-Test

- **Siehe auch Übungsaufgabe**

Gauß-Test – Einstichproben Test

- Problem:
  - Prüfe anhand des arithmetischen Mittels x, ob gegeben Daten einer Normalverteilung mit Erwartungswert μ und bekannter Varianz σ^2 folgen
- Gegeben:
  - n unabhängige, normalverteilte Zufallszahlen $X_1, \ldots, X_n$ aus derselben Grundgesamtheit mit
    - Unbekanntem Erwartungswert μ
    - Bekannter Varianz σ^2

Gauß-Test – Einstichproben Test (Fortsetzung)

- Nullhypothese H_0: Die Daten sind normalverteilt mit Erwartungswert $\mu=\mu_0$
  - Der Wert μ_0 wird vom Tester vorgegeben
- Es können drei Fälle getestet werden
  - Zweiseitiger Test: H_0: $\mu=\mu_0$ gegen H_1: $\mu\neq\mu_0$
  - Rechtsseitiger Test: H_0: $\mu\leqq\mu_0$ gegen H_1: $\mu>\mu_0$
  - Linksseitiger Test: H_0: $\mu\geqq\mu_0$ gegen H_1: $\mu<\mu_0$
- Die Test-Statistik ist immer gleich, nur die Ablehnungsbereiche unterscheiden sich
- **Die Test-Statistik z ist unter H_0 standardnormalverteilt**

$$z = \sqrt{n}\,\frac{x-\mu_0}{\sqrt{\sigma^2}}$$

t-Test

- Sehr ähnlich zum Gauß-Test
- Einstichproben Test
  - Teste, ob die gegebenen Daten zu einer Normalverteilung mit vorgegeben Erwartungswert $\mu=\mu_0$ passen
- Zweistichproben Test
  - Test, ob die Erwartungswerte der Grundgesamtheit zweier unabhängiger Stichproben gleich sind
- Hier allerdings unbekannte Varianz der Verteilung
  - Benutze Stichprobenvarianz als Schätzwert

t-Test - Einstichproben Test

- Test-Statistik t, mit arithmetischem Mittel x ...

$$t = \sqrt{n}\,\frac{x-\mu_0}{\sqrt{S^2}}$$

- ... und Stichprobenvarianz

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - x)^2$$

- **Die Test-Statistik ist unter der Nullhypothese t-verteilt mit n-1 Freiheitsgraden**
  - Für zwei-/einseitige Test wird die selbe Test-Statistik t benutzt (siehe auch Gauß-Test)

t-Test – Zweistichproben Test

- Gegeben: Zwei Stichproben X und Y vom Umfang n bzw. m aus zwei Grundgesamtheiten mit
  - Erwartungswerten μ_x und μ_y
  - Der gleichen, unbekannten Varianz σ^2
- Test-Hypothesen
  - H_0: $\mu_x - \mu_y = w$
  - H_1: $\mu_x - \mu_y \neq w$
  - w = 0, wenn auf eine gemeinsame Grundgesamtheit getestet wird

t-Test – Zweistichproben Test (Fortsetzung)

- Test-Statistik (x, y sind die arithmetischen Mittel der Stichproben X, Y)

$$t = \sqrt{\frac{nm}{n+m}} \frac{x-y-w}{\sqrt{S^2}}$$

- Mit der gewichteten, geschätzten Varianz

$$S^2 = \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2}$$

  - Mit den Varianzen S^2_X und S^2_Y der Einzelstichproben
- **Die Test-Statistik ist unter H_0 t-verteilt mit n+m-2 Freiheitsgraden**

F-Test

- Test, ob die Varianzen zweier Stichproben X, Y mit Umfang n_X, n_Y aus unterschiedlichen, normalverteilten Grundgesamtheiten gleich sind
  - H_0: $\sigma_x^2 = \sigma_y^2$
  - H_1: $\sigma_x^2 < \sigma_y^2$
- Test-Statistik ist der Quotient der geschätzten Varianzen

$$F = \frac{S_Y^2}{S_X^2} = \frac{\frac{1}{n_Y - 1} \sum_{i=1}^{n_Y} (y_i - y)^2}{\frac{1}{n_X - 1} \sum_{i=1}^{n_X} (x_i - x)^2}$$

- **Unter H_0 ist die Test-Statistik F-verteilt mit n_Y-1 Freiheitsgraden im Zähler und n_X-1 Freiheitsgraden im Nenner**

Kolmogorow-Smirnow Test

- Einstichproben Test
  - Teste, ob eine Zufallsvariable einer im **Voraus** angenommenen Verteilung folgt
  - Nicht zulässig, eine an die Daten gefittete Verteilung zu testen
    → Kein Äquivalent zu einem „*goodness of fit*" Test
- Zweistichproben Test
  - Test, ob zwei Zufallsvariablen aus derselben Verteilung stammen
- In beiden Fällen sind die Hypothesen
  - H_0: $F_X(x) = F_0(x)$ → Die Zufallsvariable X ist verteilt wie F_0
  - H_1: $F_X(x) \neq F_0(x)$ → Die Zufallsvariable X ist NICHT verteilt wie F_0
- Test-Statistik basiert auf maximalem Abstand d zwischen den kumulierten Verteilungsfunktionen F_X und F_0

$$d = \sup_x \left| F_X(x) - F_0(x) \right|$$

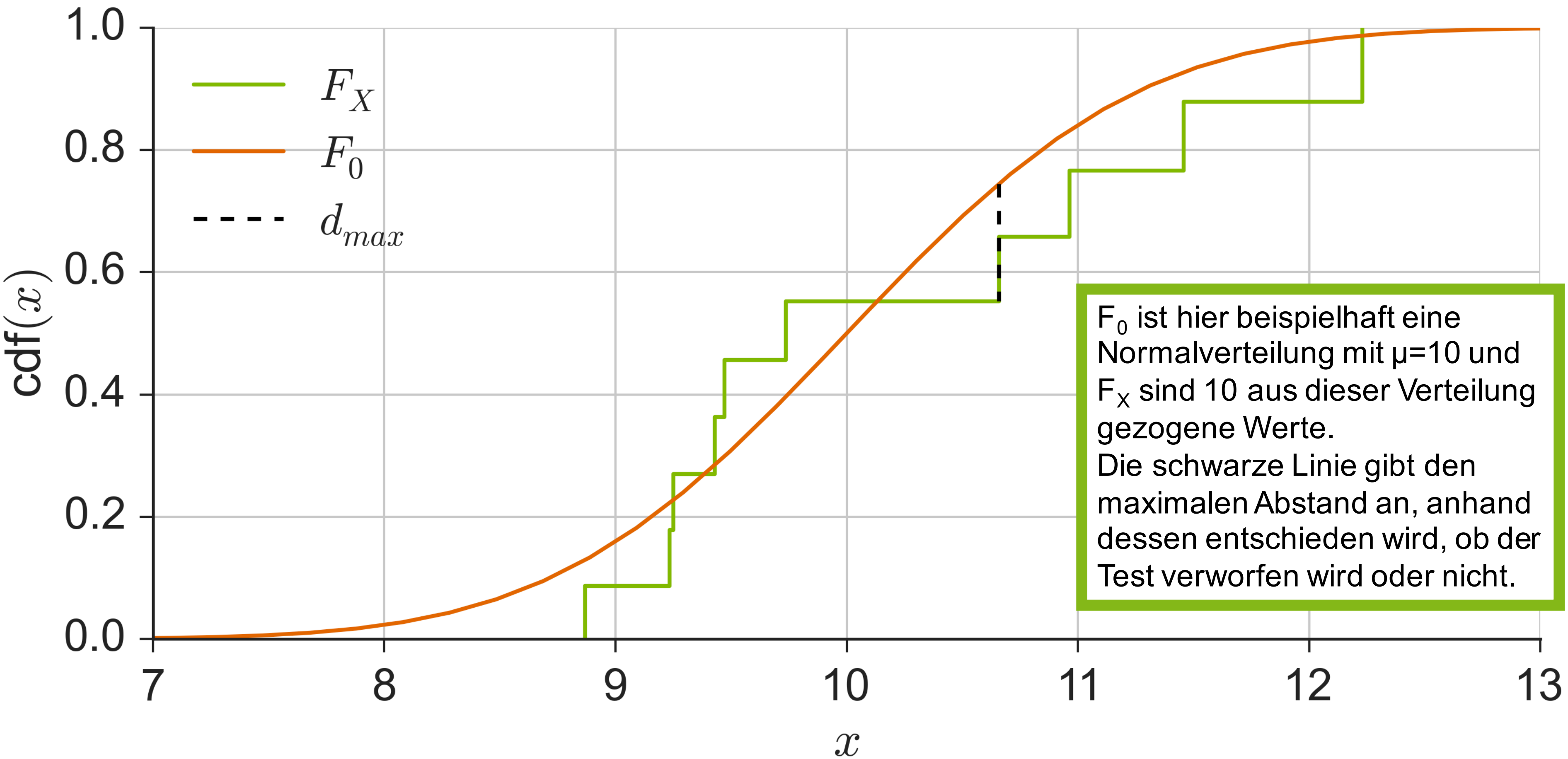
Kolmogorow-Smirnow Test – Einstichproben Test

- Teste, ob die Zufallsvariable X aus der **vorher** festgelegten Verteilung $F_0(x)$ stammt
- Aus der Stichprobe mit Umfang n von X wird die empirische Verteilungsfunktion $S(x_i)$ bestimmt
- Bestimme den maximalen Abstand aus

$$d_u = \sup_{x_i \in x} |S(x_i) - F_0(x_i)| \quad \text{und} \quad d_l = \sup_{x_i \in x} |S(x_{i-1}) - F_0(x_i)|$$

  - d_u bzw. d_l bezieht sich je auf den Abstand zwischen oberer und unterer Grenze → **das Maximum aus beiden Werten ist die Test-Statistik**
- Für n > 35 kann der kritische Wert aus $d_\alpha = \sqrt{\frac{-\ln(\alpha/2)}{2n}}$ bestimmt werden → **Test wird für $d_{max} > d_\alpha$ abgelehnt**
  - Bis n = 35 werden tabellierte Werte genutzt

Beispiel – Kolmogorow-Smirnow Test – Einstichproben Test



F_0 ist hier beispielhaft eine Normalverteilung mit μ=10 und F_X sind 10 aus dieser Verteilung gezogene Werte.
Die schwarze Linie gibt den maximalen Abstand an, anhand dessen entschieden wird, ob der Test verworfen wird oder nicht.

Kolmogorow-Smirnow Test – Zweistichproben Test

- Zufallsvariablen X und Y → Prüfe, ob X und Y aus derselben Verteilung stammen
  - Stichproben sind vom Umfang n_X bzw. n_Y
- Bilde die kumulierten Verteilungsfunktionen $S_X(x_i)$ bzw. $S_Y(y_i)$
- Die Test-Statistik ist der maximale Abstand

$$d_{\max} = \sup_z |S_X(z) - S_Y(z)|$$

- **Lehne den Test ab, wenn**

$$\sqrt{\frac{nm}{n+m}} d_{\max} > K_\alpha$$

- Für große n, m kann näherungsweise $K_\alpha = \sqrt{\frac{\ln(2/\alpha)}{2}}$ benutzt werden
  - Sonst auf tabellierte Werte zurückgreifen

Chi-Quadrat Test („*goodness of fit*“)

- Teste, ob n gemessene Daten y_i einem angenommenen Modell $f(x_i)$ folgen
  - Nullhypothese H_0: Die gemessenen Verteilungen stammen aus dem angenommenen Modell
- Test-Statistik

$$\chi^2 = \sum_{i=1}^{n} \frac{(y_i - f(x_i))^2}{(\sigma_i^{\mathrm{Modell}})^2}$$

- Test wird bei hohen Werten von χ^2 abgelehnt
  - Verglichen wird mit tabellierten χ^2 Werten mit n Freiheitsgraden am gewünschten Quantil 1-α bei einer Signifikanz α
  - Strikt einseitiger Test → Je kleiner χ^2 desto besser „passt“ das Modell
  - → **Vorsicht vor Überanpassung**

Chi-Quadrat Test („*goodness of fit*“) (Fortsetzung)

- **Vorsicht vor Überanpassung**
  - Erinnerung: χ^2 Verteilung mit n Freiheitsgraden ist die Summe von n standardnormalverteilten Zufallszahlen
  - Erwartet wird also einen Wert von $\chi^2 = n$
  - Oft auch „Chi-Quadrat über d.o.f.“ angegeben
    → χ^2 geteilt durch die Anzahl der Freiheitsgrade. Erwartung: $\chi^2 / n = 1$
- χ^2 Wert zu groß:
  - Falsches Modell gewählt
  - Fehler sind unterschätzt (zu klein)
  - Durch Zufall → Hohe Werte sind unwahrscheinlich aber immer möglich
- χ^2 Wert zu klein → „zu gutes“ Modell:
  - Fehler überschätzt
  - Zu viele gefittete Freiheitsgrade im Modell

Chi-Quadrat Test - Minimierung

- Hängt das Modell von m freien Parametern ab können diese durch Minimierung des χ^2 Werts an die Daten angepasst werden
- Dadurch reduziert sich die Anzahl der Freiheitsgrade der Test-Statistik

Anzahl Freiheitsgrade = Anzahl Datenpunkte – Anzahl an Fitparametern

  - Beachte: Auch die Gesamt-Normierung ist bereits ein Freiheitsgrad

- **Beispiel:**
  - Fitte eine Gerade ax+b an n=10 Datenpunkte
    → Test-Statistik wird mit Quantil aus einer χ^2 Verteilung mit 10 - 2 = 8 Freiheitsgraden verglichen

Chi-Quadrat Test – Fit an ein Histogramm

- Fitte Modell an ein Histogramm mit n Bins
  - Poissonverteilung in jedem Bin i mit n_i gemessenen Einträgen
  - Erwartete Varianz des Modells ist bekannt: $\sigma_i^2 = n_i^{Modell}$
- Die Test-Statistik ist

$$\chi^2 = \sum_{i=1}^{n} \frac{(n_i - n_i^{\mathrm{Modell}})^2}{n_i^{\mathrm{Modell}}}$$

- Anzahl Freiheitsgrade = Anzahl Bins – Gefittete Parameter
- Beachte:
  Die Test-Statistik ist nur bei **ausreichend hoher Statistik in den Bins** χ^2 verteilt → Poisson-Verteilung wird ausreichend gut durch eine Gauß-Verteilung beschrieben

Chi-Quadrat Test – Vergleich zweier Histogramme

- Gegeben sind zwei Histogramme mit identischem Binning mit r Bins
- Nullhypothese: Beide Histogramme repräsentieren Zufallszahlen der gleichen Verteilung
  - Es existiert für jedes Bin eine Wahrscheinlichkeit p_i dafür, dass eine Zufallszahl in i-ten Bin landet:

$$\sum_{i=1}^{r} p_i = 1$$

- Einträge im i-ten Bin des ersten Histogramms werden als n_i und des zweiten Histogramms als m_i bezeichnet

Chi-Quadrat Test – Vergleich zweier Histogramme

- Die Zählraten der einzelnen Bins folgen jeweils einer Poissonverteilung

$$\frac{e^{-Np_i}(Np_i)^{n_i}}{n_i!} \quad \text{bzw.} \quad \frac{e^{-Mp_i}(Mp_i)^{m_i}}{m_i!}$$

- Die Likelihoodfunktion eines einzelnen Bins wird somit zu

$$\mathcal{L}(p_i; n_i, m_i) = \frac{e^{-Np_i}\,(Np_i)^{n_i}}{n_i!}\,\frac{e^{-Mp_i}\,(Mp_i)^{m_i}}{m_i!}$$

- Die Likelihoodfunktion hat ein Maximum bei

$$p_i = \frac{n_i + m_i}{N + M}$$

Chi-Quadrat Test – Vergleich zweier Histogramme (Forts.)

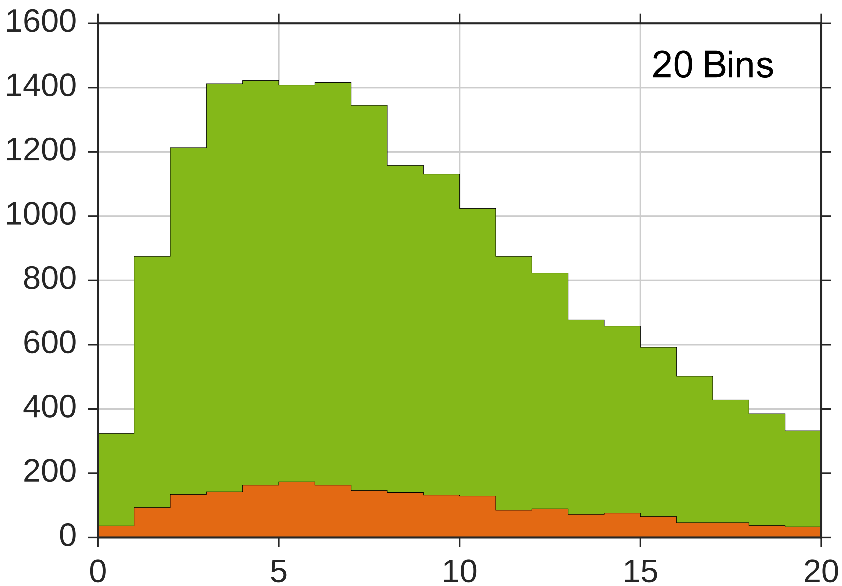
- Mit dem Likelihood Schätzer für p_i kann ein Chi-Quadrat Test aufgestellt werden:

$$\chi^2 = \sum_{i=1}^{r} \frac{(n_i - N\hat{p}_i)^2}{N\hat{p}_i} + \sum_{i=1}^{r} \frac{(m_i - M\hat{p}_i)^2}{M\hat{p}_i}$$

- Die Testgröße χ^2 folgt einer Chi-Quadrat-Verteilung mit (r-1) Freiheitsgraden
  - Es gibt 2r Summanden und geschätzt aus den Beobachtungen werden die Größen N, M, sowie (r-1) Werte für p_i
  - Ein Wert für p_i folgt aus der Bedingung, dass die Summe 1 sein muss

Chi-Quadrat Test – Vergleich zweier Histogramme

- Beispiel:



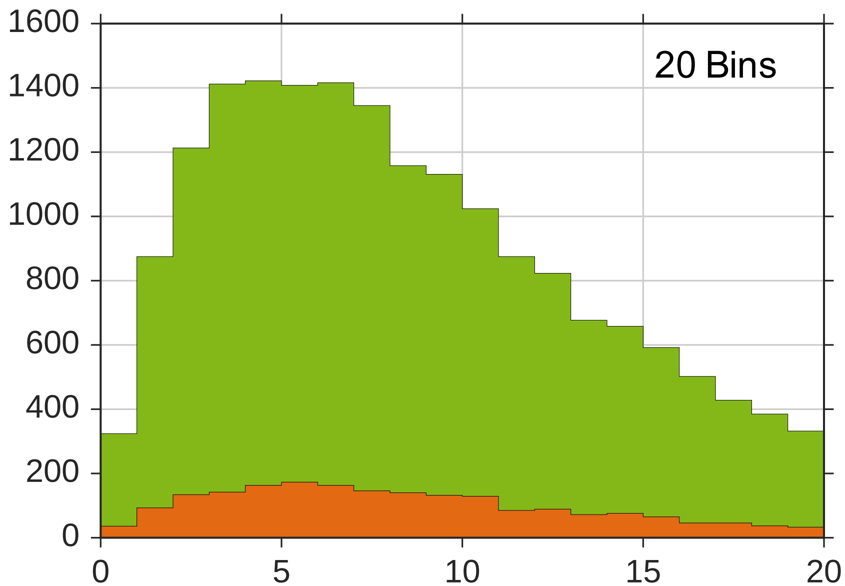
$$\boxed{\chi^2 = 10,8}$$

Percentage Points of the Chi-Square Distribution

Degrees of Freedom	Probability of a larger value of x²								
	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
1	0.000	0.004	0.016	0.102	0.455	1.32	2.71	3.84	6.63
2	0.020	0.103	0.211	0.575	1.386	2.77	4.61	5.99	9.21
3	0.115	0.352	0.584	1.212	2.366	4.11	6.25	7.81	11.34
4	0.297	0.711	1.064	1.923	3.357	5.39	7.78	9.49	13.28
5	0.554	1.145	1.610	2.675	4.351	6.63	9.24	11.07	15.09
6	0.872	1.635	2.204	3.455	5.348	7.84	10.64	12.59	16.81
7	1.239	2.167	2.833	4.255	6.346	9.04	12.02	14.07	18.48
8	1.647	2.733	3.490	5.071	7.344	10.22	13.36	15.51	20.09
9	2.088	3.325	4.168	5.899	8.343	11.39	14.68	16.92	21.67
10	2.558	3.940	4.865	6.737	9.342	12.55	15.99	18.31	23.21
11	3.053	4.575	5.578	7.584	10.341	13.70	17.28	19.68	24.72
12	3.571	5.226	6.304	8.438	11.340	14.85	18.55	21.03	26.22
13	4.107	5.892	7.042	9.299	12.340	15.98	19.81	22.36	27.69
14	4.660	6.571	7.790	10.165	13.339	17.12	21.06	23.68	29.14
15	5.229	7.261	8.547	11.037	14.339	18.25	22.31	25.00	30.58
16	5.812	7.962	9.312	11.912	15.338	19.37	23.54	26.30	32.00
17	6.408	8.672	10.085	12.792	16.338	20.49	24.77	27.59	33.41
18	7.015	9.390	10.865	13.675	17.338	21.60	25.99	28.87	34.80
19	7.633	10.117	11.651	14.562	18.338	22.72	27.20	30.14	36.19
20	8.260	10.851	12.443	15.452	19.337	23.83	28.41	31.41	37.57

Chi-Quadrat Test – Vergleich zweier Histogramme

- Beispiel:

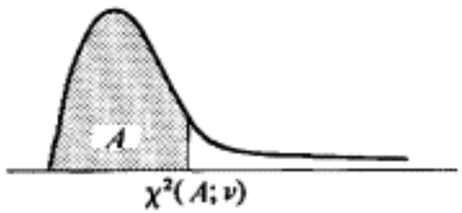


$$\boxed{\chi^2 = 10,8}$$

Percentage Points of the Chi-Square Distribution

Degrees of Freedom	Probability of a larger value of x ²								
	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
1	0.000	0.004	0.016	0.102	0.455	1.32	2.71	3.84	6.63
2	0.020	0.103	0.211	0.575	1.386	2.77	4.61	5.99	9.21
3	0.115	0.352	0.584	1.212	2.366	4.11	6.25	7.81	11.34
4	0.297	0.711	1.064	1.923	3.357	5.39	7.78	9.49	13.28
5	0.554	1.145	1.610	2.675	4.351	6.63	9.24	11.07	15.09
6	0.872	1.635	2.204	3.455	5.348	7.84	10.64	12.59	16.81
7	1.239	2.167	2.833	4.255	6.346	9.04	12.02	14.07	18.48
8	1.647	2.733	3.490	5.071	7.344	10.22	13.36	15.51	20.09
9	2.088	3.325	4.168	5.899	8.343	11.39	14.68	16.92	21.67
10	2.558	3.940	4.865	6.737	9.342	12.55	15.99	18.31	23.21
11	3.053	4.575	5.578	7.584	10.341	13.70	17.28	19.68	24.72
12	3.571	5.226	6.304	8.438	11.340	14.85	18.55	21.03	26.22
13	4.107	5.892	7.042	9.299	12.340	15.98	19.81	22.36	27.69
14	4.660	6.571	7.790	10.165	13.339	17.12	21.06	23.68	29.14
15	5.229	7.261	8.547	11.037	14.339	18.25	22.31	25.00	30.58
16	5.812	7.962	9.312	11.912	15.338	19.37	23.54	26.30	32.00
17	6.408	8.672	10.085	12.792	16.338	20.49	24.77	27.59	33.41
18	7.015	9.390	10.865	13.675	17.338	21.60	25.99	28.87	34.80
19	7.633	10.117	11.651	14.562	18.338	22.72	27.20	30.14	36.19
20	8.260	10.851	12.443	15.452	19.337	23.83	28.41	31.41	37.57

Chi-Quadrat Tabellen



Achtung beim Ablesen:
Welcher Wert ist angegeben

1-α

ν	A .005	.010	.025	.050	.100	.900	.950	.975	.990	.995
1	0.0^4393	0.0^3157	0.0^3982	0.0^2393	0.0158	2.71	3.84	5.02	6.63	7.88
2	0.0100	0.0201	0.0506	0.103	0.211	4.61	5.99	7.38	9.21	10.60
3	0.072	0.115	0.216	0.352	0.584					
4	0.207	0.297	0.484	0.711	1.064					
5	0.412	0.554	0.831	1.145	1.61					
6	0.676	0.872	1.24	1.64	2.20					
7	0.989	1.24	1.69	2.17	2.83					
8	1.34	1.65	2.18	2.73	3.49					
9	1.73	2.09	2.70	3.33	4.17					
10	2.16	2.56	3.25	3.94	4.87					
11	2.60	3.05	3.82	4.57	5.58					
12	3.07	3.57	4.40	5.23	6.30					
13	3.57	4.11	5.01	5.89	7.04					
14	4.07	4.66	5.63	6.57	7.79					
15	4.60	5.23	6.26	7.26	8.55					
16	5.14	5.81	6.91	7.96	9.31					
17	5.70	6.41	7.56	8.67	10.09					
18	6.26	7.01	8.23	9.39	10.86					
19	6.84	7.63	8.91	10.12	11.65					

α

Percentage Points of the Chi-Square Distribution

Degrees of Freedom	Probability of a larger value of x^2						
	0.99	**0.95**	**0.90**	**0.75**	**0.50**	**0.25**	**0.10**
1	0.000	0.004	0.016	0.102	0.455	1.32	2.71
2	0.020	0.103	0.211	0.575	1.386	2.77	4.61
3	0.115	0.352	0.584	1.212	2.366	4.11	6.25
4	0.297	0.711	1.064	1.923	3.357	5.39	7.78
5	0.554	1.145	1.610	2.675	4.351	6.63	9.24
6	0.872	1.635	2.204	3.455	5.348	7.84	10.64
7	1.239	2.167	2.833	4.255	6.346	9.04	12.02
8	1.647	2.733	3.490	5.071	7.344	10.22	13.36
9	2.088	3.325	4.168	5.899	8.343	11.39	14.68
10	2.558	3.940	4.865	6.737	9.342	12.55	15.99
11	3.053	4.575	5.578	7.584	10.341	13.70	17.28