

Konzepte für Experimente an zukünftigen Hadroncollidern II

PD Dr. Oliver Kortner

27.05.2021

Wichtige Wahrscheinlichkeitsverteilung

- Die **Binomialverteilung** gibt die Wahrscheinlichkeit an, n_k Ereignisse aus insgesamt N Ereignissen zu beobachten, wenn ν_k Ereignisse erwartet werden:

$$p(n_k; \nu_k) = \binom{N}{n_k} \left(\frac{\nu_k}{N}\right)^{n_k} \left(1 - \frac{\nu_k}{N}\right)^{N-n_k}.$$

$$\langle n_k \rangle = \nu_k, \quad \text{Var}(n_k) = \nu_k \left(1 - \frac{\nu_k}{N}\right).$$

- Man erhält die **Poissonverteilung** aus der Binomialverteilung, wenn N hinreichend groß ist und $n_k \ll N$ gilt:

$$p(n_k; \nu_k) = \frac{\nu_k^{n_k}}{n_k!} e^{-\nu_k}.$$

$$\langle n_k \rangle = \nu_k, \quad \text{Var}(n_k) = \nu_k.$$

- Normalverteilung einer d -dimensionalen Zufallsgröße $x \in \mathbb{R}^d$**

$$p(x; \mu, \Sigma) = \frac{1}{(2\pi)^{d/2}} \frac{1}{\sqrt{\det(\Sigma)}} \exp\left(-\frac{1}{2}(x - \mu)^t \Sigma^{-1} (x - \mu)\right).$$

$$\Sigma \in \mathbb{R}^d, \quad \mu \in \mathbb{R}^d.$$

Eigenschaften der eindimensionalen Normalverteilung

$$p(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

$w_n :=$ Wahrscheinlichkeit, einen Wert $x \in [\mu - n\sigma, \mu + n\sigma]$ zu beobachten.

n	w_n
1	0,6827
2	0,9545
3	0,9973
4	$1 - 6,3 \cdot 10^{-5}$
5	$1 - 5,7 \cdot 10^{-7}$

w_n	n
0,900	1,645
0,950	1,960
0,990	2,576
0,999	3,290

Begriff der stochastischen Konvergenz

(t_n) sei eine Folge von Zufallsgrößen und T ebenfalls eine Zufallsgröße. Man sagt, t_n **konvergiere stochastisch gegen T** , wenn es zu jedem $p \in [0, 1[$ und $\epsilon > 0$ ein N so gibt, dass die Wahrscheinlichkeit P , dass $|t_n - T| > \epsilon$ ist, kleiner als p für alle $n > N$ ist:

$$P(|t_n - T| > \epsilon) < p \quad (n > N).$$

Mit anderen Worten: Die Wahrscheinlichkeit, einen von T verschiedenen Wert t_n zu beobachten, verschwindet für $n \rightarrow \infty$.

Gesetz der großen Zahlen. Zentraler Grenzwertsatz

Das Gesetz großer Zahlen

(x_n) sei eine Folge unabhängiger Zufallsgrößen, wobei jedes x_n derselben Verteilungsfunktion folgt. μ bezeichne den Erwartungswert von x_n . Dann konvergiert das arithmetische Mittel

$$\frac{1}{N} \sum_{n=1}^N x_n$$

stochastisch gegen μ .

Der zentrale Grenzwertsatz

(x_n) sei eine Folge identisch verteilter Zufallsgrößen mit dem Mittelwert μ und der Standardabweichung σ . Dann konvergiert für $N \rightarrow \infty$ die standardisierte Zufallsgröße

$$Z_N := \frac{\sum_{n=1}^N x_n - N\mu}{\sigma\sqrt{N}}$$

punktweise gegen eine Normalverteilung mit dem Mittelwert 0 und der Standardabweichung 1.

Punktschätzung

Es sei α ein Parameter einer Wahrscheinlichkeitsverteilung. Das Ziel der **Punktschätzung** ist es, den besten Schätzwert (die beste Messung in der Sprechweise der Physiker) von α zu finden.

x : Zufallsgröße, die den experimentellen Messwerten entspricht.

$p(x; \alpha)$: Wahrscheinlichkeitsdichte für die Messung von x in Abhängigkeit des Parameters α .

x und α können mehrdimensional sein.

Definition. Ein **Punktschätzer** \mathcal{E}_α ist eine Funktion von x , mit der der Wert des Parameters α geschätzt werden kann. $\hat{\alpha}$ bezeichne diesen Schätzwert. Es ist also $\hat{\alpha} = \mathcal{E}_\alpha(x)$.

Ziel ist es, eine Funktion \mathcal{E}_α zu finden, bei der $\hat{\alpha}$ so nah wie möglich am wahren Wert von α liegt.

Da $\hat{\alpha}$ eine Funktion von Zufallsgrößen ist, ist $\hat{\alpha}$ selbst eine Zufallsgröße.

$$p(\hat{\alpha}) = \int_D \mathcal{E}_\alpha(x) p(x; \alpha) dx,$$

wobei α den wahren Wert des Parameters bezeichnet.

Qualitätskriterien für Punktschätzer

Konsistenz

n : Anzahl der Messungen, die für die Punktschätzung verwendet werden.

$\hat{\alpha}_n$: zugehöriger Schätzwert.

α_0 : wahrer Wert von α .

Man bezeichnet \mathcal{E}_α als **konsistenten Punktschätzer**, falls $\hat{\alpha}_n$ stochastisch gegen α_0 konvergiert. D.h. die Wahrscheinlichkeit, einen von α_0 verschiedenen Wert zu schätzen, geht für $n \rightarrow \infty$ gegen 0.

Erwartungstreue

Der **Bias eines Schätzwerts** $\hat{\alpha}$ ist definiert als

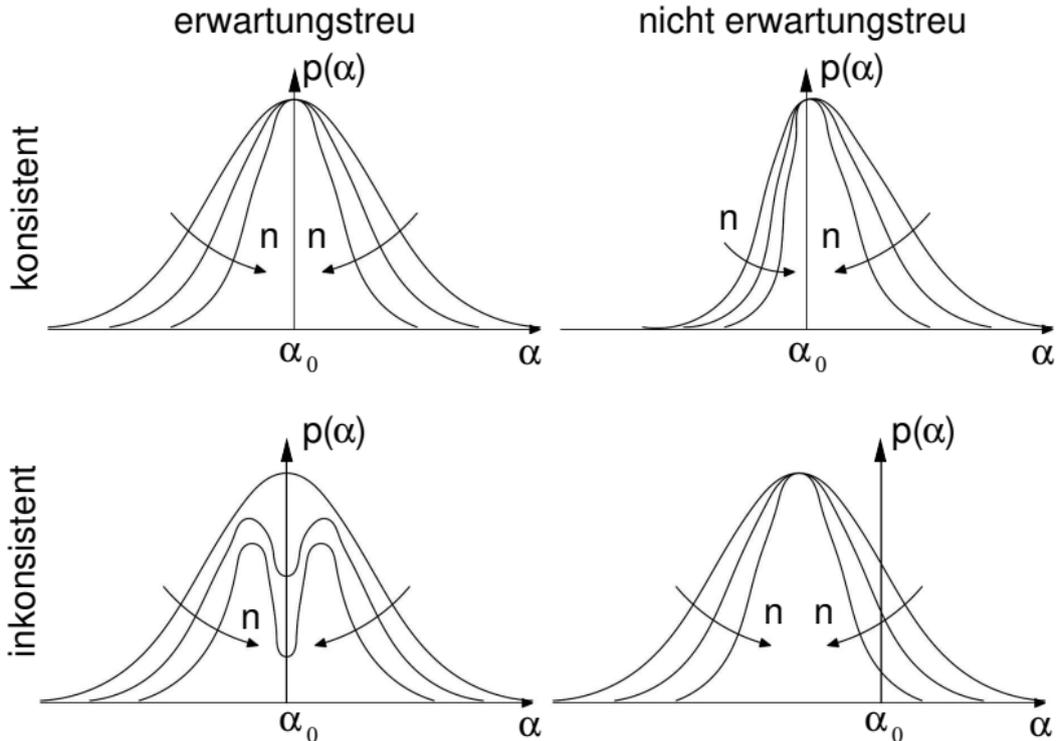
$$b_n(\hat{\alpha}) := E(\hat{\alpha}_n - \alpha_0) = E(\hat{\alpha}_n) - \alpha_0.$$

Der Punktschätzer ist **erwartungstreu**, wenn

$$b_n(\hat{\alpha}) = 0, \text{ also } E(\hat{\alpha}_n) = \alpha_0$$

für alle n ist.

Veranschaulichung von Konsistenz und Erwartungstreue



Weitere Qualitätskriterien für Punktschätzer

Effizienz

V_{min} sei die kleinstmögliche Varianz aller Punktschätzer eines reellwertigen Parameters. Die **Effizienz** eines bestimmten Punktschätzers ist als das Verhältnis $\frac{V_{min}}{Var(\hat{\alpha})}$ gegeben, wobei $Var(\hat{\alpha})$ die Varianz von $\hat{\alpha}$ für diesen Punktschätzer ist.

Suffizienz

Jede Funktion von Messdaten x nennt man eine **Statistik**. Eine **suffiziente Statistik für α** ist eine Funktion der Messdaten, die den gesamten Informationsgehalt über α enthält.

In der Hochenergiephysik verwendete Punktschätzer

Methode des maximalen Likelihoods

$p(x; \alpha)$: Wahrscheinlichkeit, die Messwerte x bei einem gegebenem Parameter α zu erhalten.

- Wenn man nun die gemessenen Werte x in die Funktion $p(x; \alpha)$ einsetzt, erhält man eine Statistik von x , die man **Likelihood** oder **Likelihoodfunktion** $L(x; \alpha)$ nennt.
- Man verwendet den Begriff des Likelihoods, um die Verwandtschaft mit der Wahrscheinlichkeitsdichte $p(x; \alpha)$ anzudeuten und gleichzeitig deutlich zu machen, **dass L keine Wahrscheinlichkeitsfunktion ist.**

$f(x_k; \alpha)$ sei die Wahrscheinlichkeitsdichte für den Ausgang einer einzelnen Messung x_k . Bei n unabhängigen Messungen $x = (x_1, \dots, x_n)$ ist dann

$$L(x_1, \dots, x_n; \alpha) = \prod_{k=1}^n f(x_k; \alpha).$$

Bei der **Methode des maximalen Likelihoods** nimmt man als Schätzwert für α den Wert von α , für den $L(x; \alpha)$ maximal wird.

Asymptotisches Verhalten des maximalen Likelihoods

$n \rightarrow \infty$

- Der Punktschätzer ist konsistent.
- Der Punktschätzer ist effizient.
- $\hat{\alpha}$ ist normalverteilt.
- Wegen der Konsistenz ist der Punktschätzer asymptotisch erwartungstreu.

n endlich

Um das Verhalten des Punktschätzer bei begrenzter Datenmenge n zu bestimmen, bedient man sich in der experimentellen Praxis Ensembles zufällig erzeugter simulierter Daten, auf die man den Punktschätzer anwendet.

Methode der kleinsten Quadrate

n Messungen x_1, \dots, x_n .

$E(x_k; \alpha)$: Erwartungswert von x_k bei gegebenem α („theoretische Vorhersage“ für den Wert von x_k).

$V = (\text{cov}(x_k, x_\ell))$: Kovarianzmatrix. Im Allgemeinen ist auch V eine Funktion von α .

$$Q^2 := \sum_{k, \ell=1}^n [x_k - E(x_k; \alpha)] V_{k\ell}^{-1}(\alpha) [x_\ell - E(x_\ell; \alpha)].$$

Bei der **Methode der kleinsten Quadrate** wählt man als Schätzwert für α denjenigen Wert, bei dem Q^2 minimal wird.

Ziel: Bestimmung eines Intervalls, welches mit einer vorgegebenen Wahrscheinlichkeit den wahren Wert eines Parameters enthält.

Grenzfall der Normalverteilung

Nehmen wir an, die Größe $x \in \mathbb{R}$ sei normalverteilt, d.h.

$$p(x) = N(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}}.$$

Wenn μ und σ bekannt sind, dann ist

$$p(a < x < b) = \int_a^b N(x; \mu, \sigma) dx =: \beta.$$

Falls μ unbekannt ist, kann man $p(\mu + c < x < \mu + d)$ berechnen:

$$\begin{aligned} \beta = p(\mu + c < x < \mu + d) &= \int_{\mu+c}^{\mu+d} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}} dx = \int_c^d \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2} \frac{y^2}{\sigma^2}} dy \\ &= p(c - x < -\mu < d - x) = p(x - d < \mu < x - c). \end{aligned}$$

$$\begin{aligned}\beta = p(\mu + c < x < \mu + d) &= \int_{\mu+c}^{\mu+d} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} dx = \int_c^d \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{y^2}{\sigma^2}} dy \\ &= p(c - x < -\mu < d - x) = p(x - d < \mu < x - c).\end{aligned}$$

D.h. wenn man x gemessen hat, ist die Wahrscheinlichkeit, dass der gesuchte Wert von μ zwischen $x - d$ und $x - c$ liegt, gleich β .

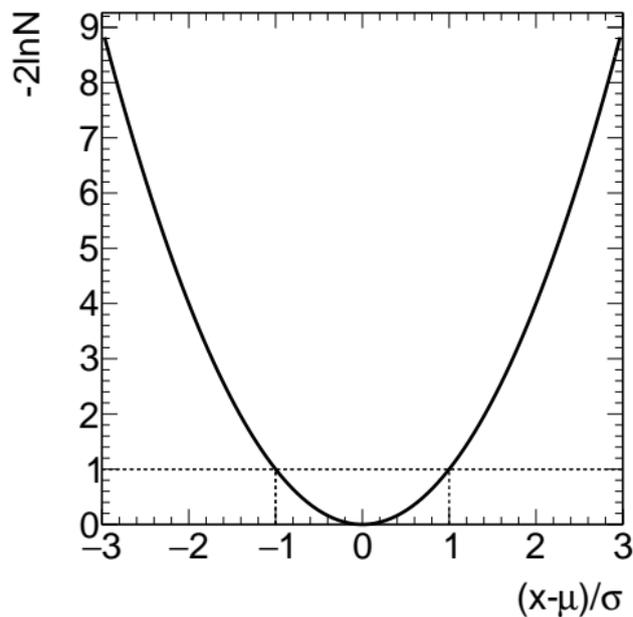
- Ist x ein Parameter $\hat{\alpha}$ aus einer Punktschätzung, die mit der Methode des maximalen Likelihoods oder der Methode der kleinsten Quadrate durchgeführt wurde, dann ist $\hat{\alpha}$ im asymptotischen Grenzfall normalverteilt, und die obigen Formeln können zur Intervallschätzung angewendet werden.
- Die Intervalle $[a, b]$ bzw. $[x - d, x - c]$ nennt man **Konfidenzintervalle**. β ist das zum Konfidenzniveau gehörige **Konfidenzniveau**.

$$Q(x; \mu, \Sigma) := (x - \mu)^t \Sigma^{-1} (x - \mu), \quad x, \mu \in \mathbb{R}^d.$$
$$p(Q) = \frac{1}{(2\pi)^{d/2}} \cdot \frac{1}{\sqrt{\det(\Sigma)}} \exp\left(-\frac{1}{2} Q(x; \mu, \Sigma)\right).$$

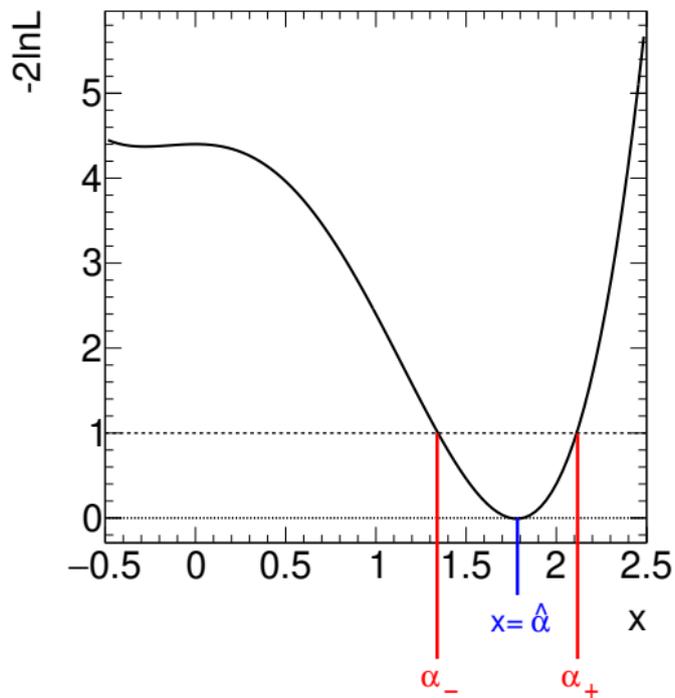
In mehreren Dimensionen wird das **Konfidenzintervall** zu einem **Konfidenzbereich**, der zum **Konfidenzniveau** β gehört:

$$p(Q(x; \mu, \Sigma) < K_\beta^2) = \beta.$$

$$-2 \ln N(x = \mu \pm \sigma; \mu, \sigma) - [-2 \ln N(x = \mu; \mu, \sigma)] = 1.$$



Verallgemeinerung



Konfidenzintervall: $[\alpha_-, \alpha_+]$.

Ziel, festzustellen, welche Hypothese (für eine Wahrscheinlichkeitsverteilung) die aufgezeichneten Messpunktverteilungen (Daten) beschreibt.

Nomenklatur. H_0 : Nullhypothese.

H_1 : alternative Hypothese.

Einfache und zusammengesetzte Hypothesen

- Wenn die Hypothesen H_0 und H_1 vollständig ohne freie Parameter gegeben sind, nennt man die Hypothesen **einfache Hypothesen**.
- Falls eine Hypothese mindestens einen freien Parameter enthält, bezeichnet man sie als **zusammengesetzte Hypothese**.

Vorgehensweise

Für den Hypothesentest muss man W so wählen, dass

$$p(\text{Daten} \in W | H_0) = \alpha$$

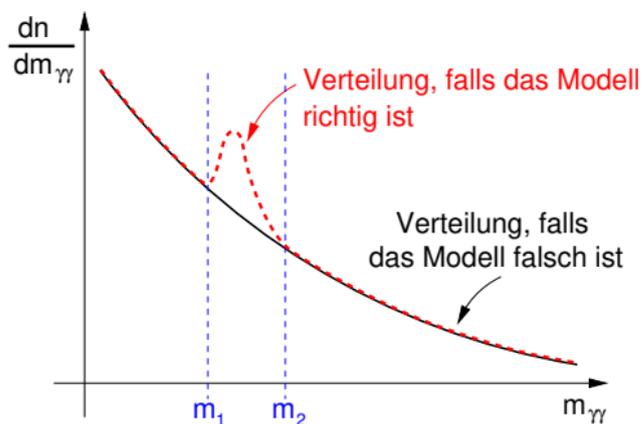
bei klein zu wählendem α und gleichzeitig

$$p(\text{Daten} \in D \setminus W | H_1) = \beta$$

mit möglichst kleinem β .

Einführendes Beispiel zum Hypothesentest

Ein theoretisches Modell sagt die Existenz eines Teilchens mit der Masse M , den Produktionswirkungsquerschnitt und die partielle Breite für den Zerfall in ein Photonenpaar voraus. Um dieses Modell zu bestätigen oder zu widerlegen, muss man sich die Verteilung von $m_{\gamma\gamma}$ ansehen.



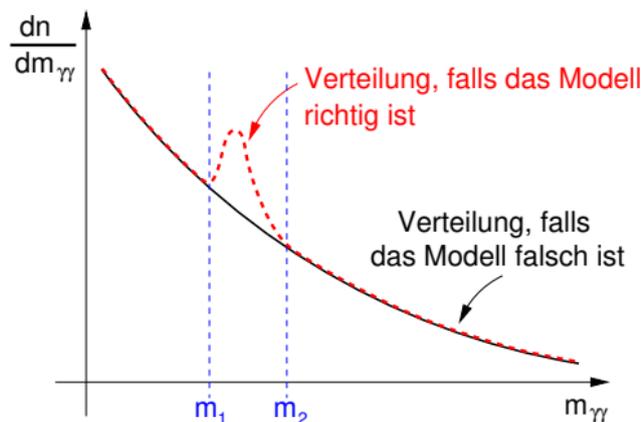
Im Intervall $[m_1, m_2]$ ist man auf die Vorhersage des Modells empfindlich. Man hat zwei Hypothesen, nämlich dass die Theorie richtig oder falsch ist.

H_0 : Nullhypothese: „Theorie ist falsch.“

H_1 : Alternativhypothese: „Theorie ist richtig.“

Bei genügend großer Datenmenge ist die Wahrscheinlichkeit, dass die gemessene $m_{\gamma\gamma}$ -Verteilung wie H_0 aussieht, klein, falls die Theorie stimmt. Gleichzeitig ist die Wahrscheinlichkeit, dass die gemessenen Massenverteilung wie H_1 aussieht, groß.

Einführendes Beispiel zum Hypothesentest



n : Anzahl der im Intervall $[m_1, m_2]$ gemessenen Ereignisse.
Man muss nun einen Schwellenwert N so wählen, dass

$$p(n > N | H_0) = \alpha$$

mit klein zu wählendem α und

$$p(n \leq N | H_1) = \beta$$

möglichst klein ist, falls die Theorie, also H_1 richtig ist.

n : Anzahl der im Intervall $[m_1, m_2]$ gemessenen Ereignisse.
Man muss nun einen Schwellenwert N so wählen, dass

$$p(n > N | H_0) = \alpha$$

mit klein zu wählendem α und

$$p(n \leq N | H_1) = \beta$$

möglichst klein ist, falls die Theorie, also H_1 richtig ist.

Experimentelle Praxis

- $\alpha = 5,7 \cdot 10^{-7}$, was 5σ einer Normalverteilung entspricht, um die Entdeckung eines Teilchens zu behaupten.
- Bei einem Wert von $\alpha = 0,3\%$, was 3σ einer Normalverteilung entspricht, sagt man, man habe Anzeichen für die Existenz eines neuen Teilchens.

Fehler erster und zweiter Art

Das Konfidenzniveau α ist als die Wahrscheinlichkeit definiert, dass $x \in W$ liegt, falls die Nullhypothese H_0 richtig ist:

$$p(x \in W | H_0) = \alpha.$$

Die Wahrscheinlichkeit β gibt die Wahrscheinlichkeit an, dass man die Alternativhypothese H_1 fälschlicherweise verwirft:

$$p(x \in D \setminus W | H_1) = \beta.$$

	H_0 richtig	H_1 richtig
Vorgehensweise		
$x \notin W \Rightarrow H_0$ wird als richtig betrachtet	Gute Akzeptanz, da $p(x \in D \setminus W H_0) = 1 - \alpha$ groß ist	Verunreinigung Fehler zweiter Art $p(x \in D \setminus W H_1) = \beta.$
$x \in W \Rightarrow H_0$ wird verworfen, H_1 als richtig betrachte	Fehlentscheidung Fehler erster Art $p(x \in W H_0) = \alpha$ ist klein	Verwerfen von H_0 gut, da $p(x \in W H_1) = 1 - \beta$ ist groß.