
Schätzen in drei statistischen Paradigmata

Dr. Dirk Hasenclever
IMISE, Leipzig

Dirk.Hasenclever@IMISE.uni-Leipzig.de

Empirische und theoretische Verteilungsfunktion I

Seien $X_1, \dots, X_N$ unabhängig, identisch verteilte Zufallsvariablen mit theoretischer Verteilungsfunktion $F(x) := \text{pr}(X < x)$

Bezeichne $\hat{F}_N(x) := \frac{\#\{k \mid x_k < x\}}{N}$ die empirische Verteilungsfunktion zu einer Realisierung $x_1, \dots, x_N$.

$$\hat{S}_N(x) := \frac{\#\{k \mid x_k \geq x\}}{N} = 1 - \hat{F}_N(x)$$

bezeichnet man als invers - kumulierte Verteilungsfunktion)

Als Maß für die Güte der Schätzung sei

$$\Delta_N = \sup_{-\infty < x < +\infty} |\hat{F}_N(x) - F(x)| = \|\hat{F}_N - F\|$$

der Abstand zwischen empirischer und theoretischer Verteilungsfunktion in der Supremumsnorm.

Empirische und theoretische Verteilungsfunktion II

Hauptsatz der mathematischen Statistik (Gliwensko)

Es gilt : $\text{pr}[\lim_{N \rightarrow \infty} (\Delta_N) = 0] = 1$

(Wahrscheinlichkeit über virtuelle Wiederholungen von Zufallsexperimenten, in denen $N \rightarrow \infty$)

Anders gesagt : Zu vorgegebener Genauigkeit ε und angestrebter Konfidenz $1 - \delta$ gibt es eine N_0 , so dass $\text{pr}[\sup_{n \geq N_0} (\Delta_N) \leq \varepsilon] \geq 1 - \delta$

Beliebige unabhängige Wiederholbarkeit vorausgesetzt, lässt sich $F(x)$ bei hinreichenden Ressourcen (n groß) beliebig genau mit vorgegebener Sicherheit bestimmen.

Empirische und theoretische Verteilungsfunktion III

Satz von Kolmogoroff :

Genauer gilt :

$$\lim_{n \rightarrow \infty} pr[\sqrt{n}\Delta_n < y] = \begin{cases} K(y) = \sum_{k=-\infty}^{+\infty} (-1)^k \exp(-2k^2 y^2) & \text{falls } y > 0 \\ 0 & \text{falls } y \leq 0 \end{cases}$$

Die Funktion $K(y) \in [0; 1]$ und ist tabelliert.

Damit Hauptsatz auch operational :

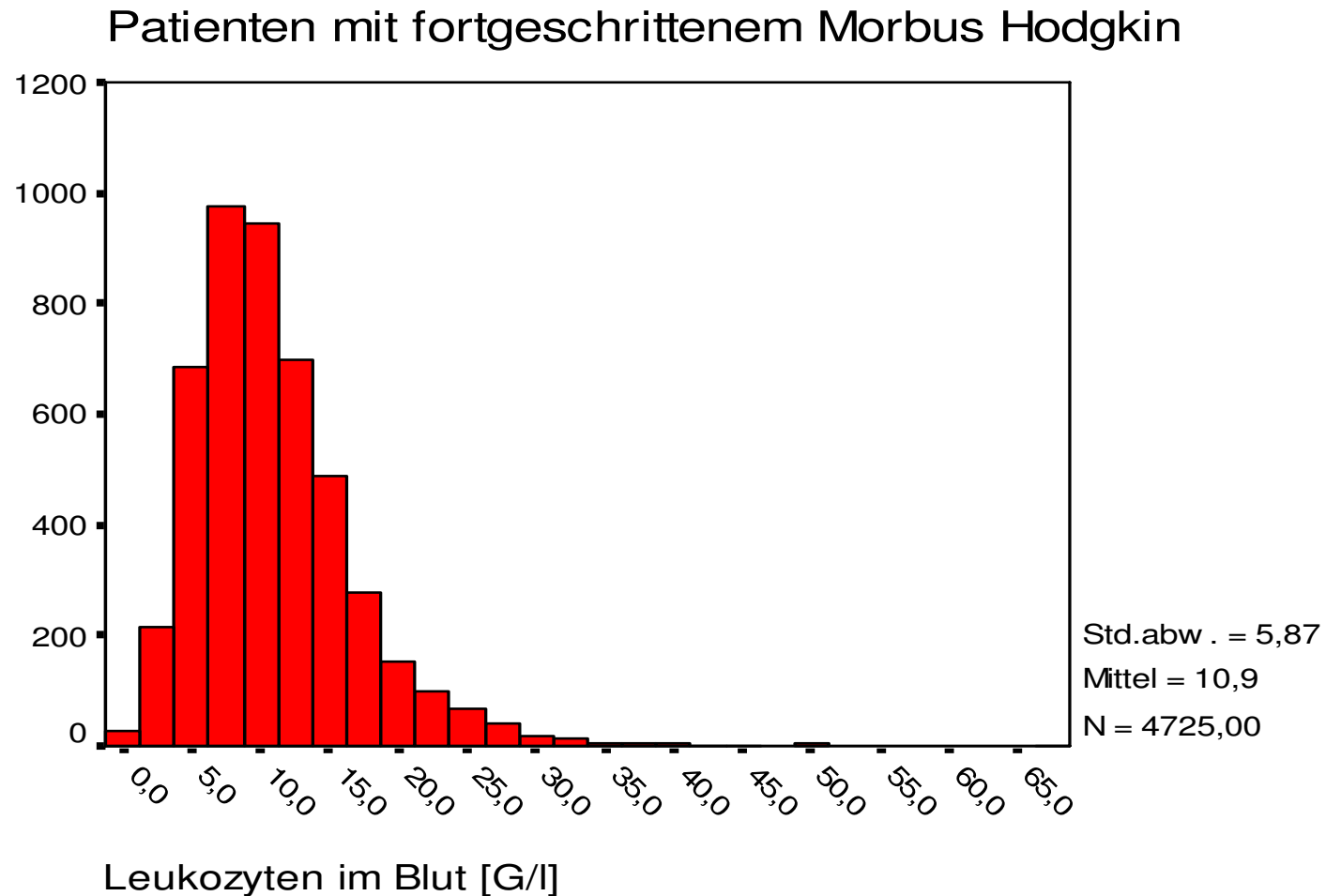
(1) zu $1 - \delta$ suche y mit $K(y) = 1 - \delta$

(2) bestimme N_0 so, dass $\frac{y}{\sqrt{n}} \leq \varepsilon$ für alle $n \geq N_0$

(NB. : $K(y)$ liegt dem Kolmogoroff - Smirnov Test auf Vorliegen eines Verteilungstyps zu Grunde.)

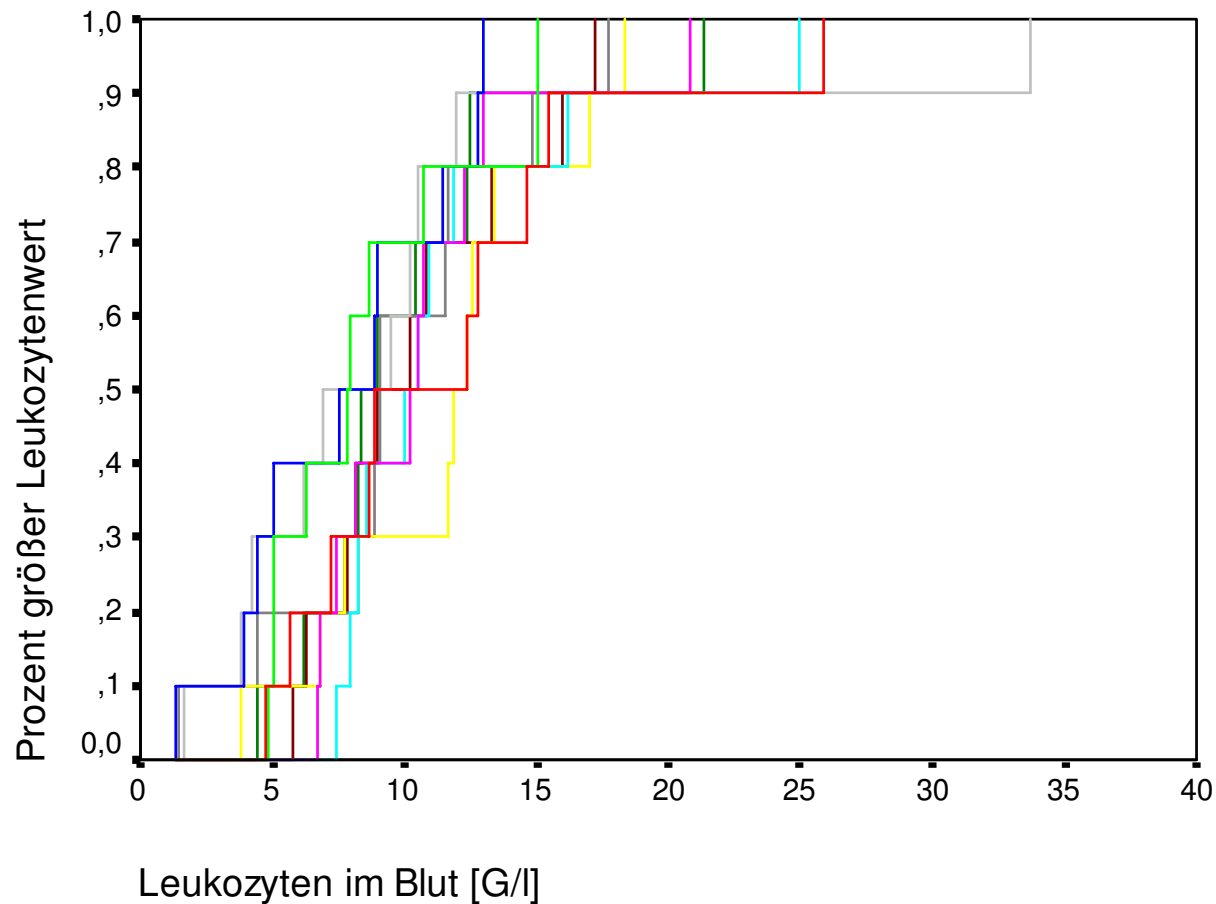
Illustration zum Hauptsatz der Statistik

Ziehe aus sehr großem Datensatz



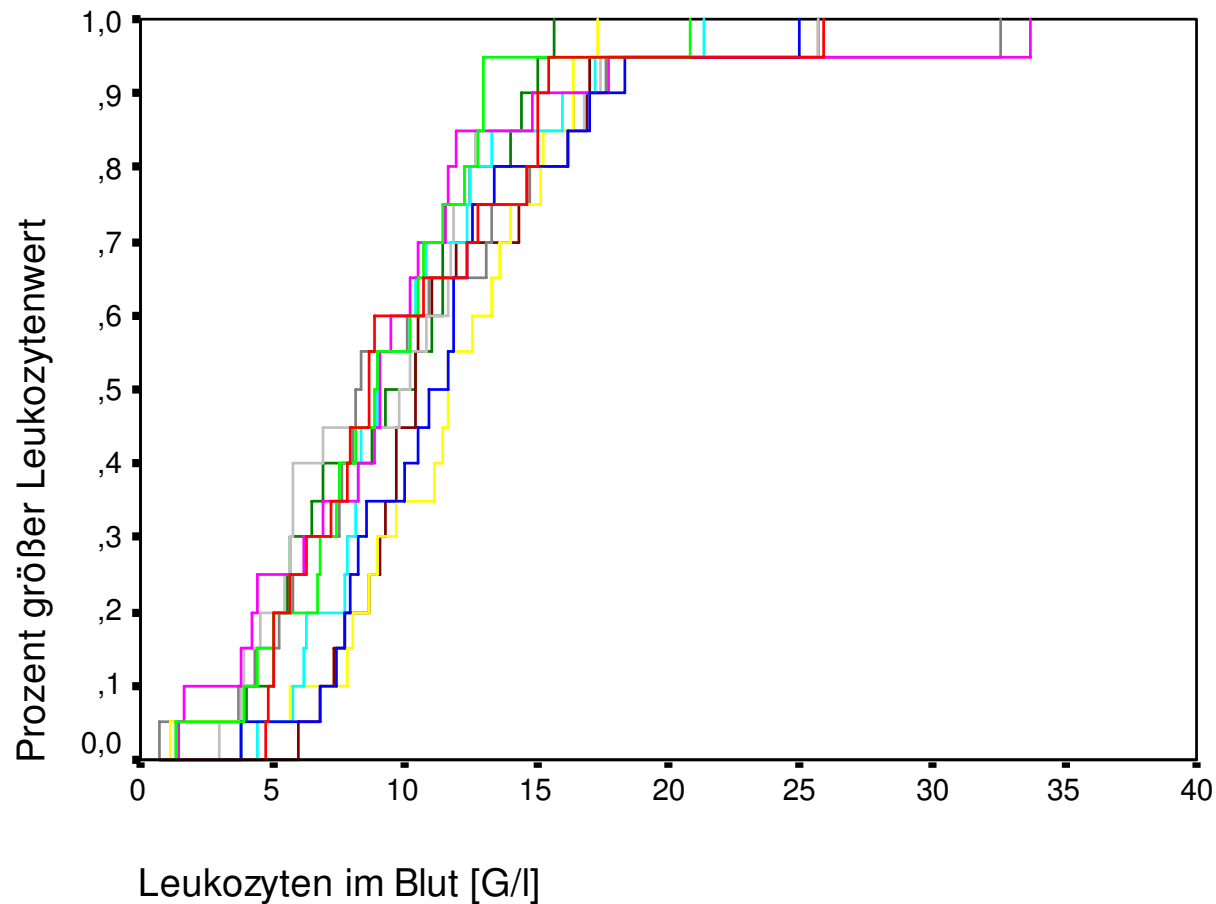
Schätzung einer Verteilung – Effekt der Fallzahl

9 Stichproben mit 10 Patienten



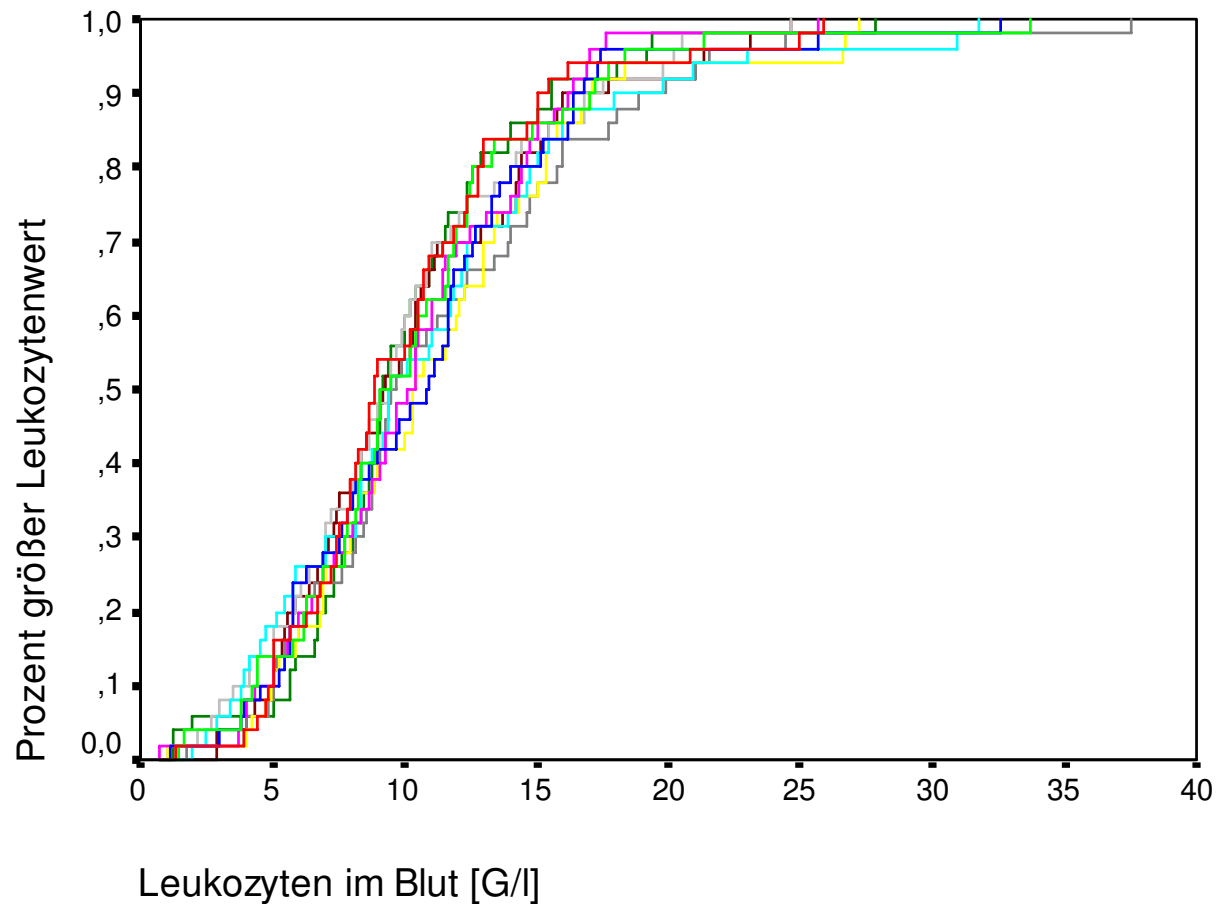
Schätzung einer Verteilung – Effekt der Fallzahl

9 Stichproben mit 20 Patienten



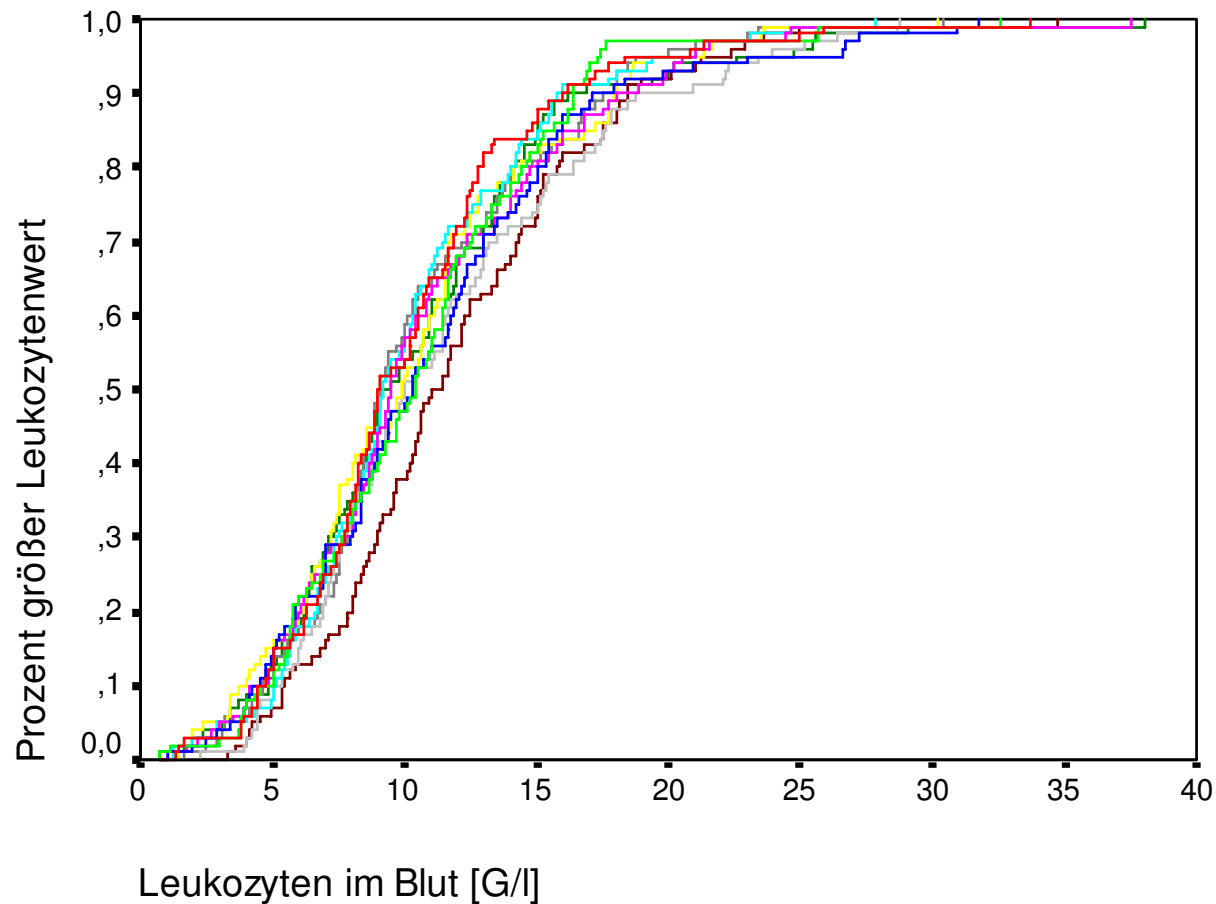
Schätzung einer Verteilung – Effekt der Fallzahl

9 Stichproben von 50 Patienten



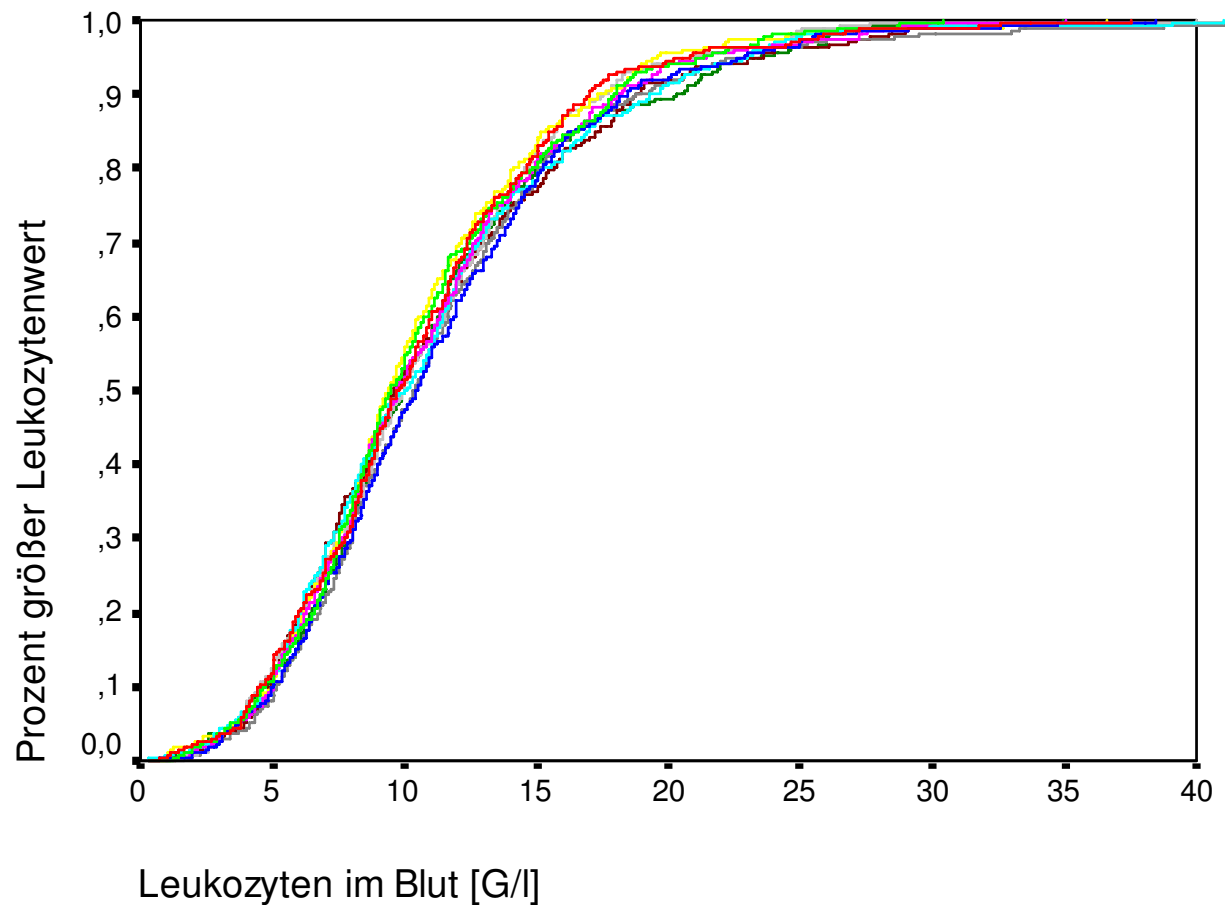
Schätzung einer Verteilung – Effekt der Fallzahl

9 Stichproben mit 100 Patienten



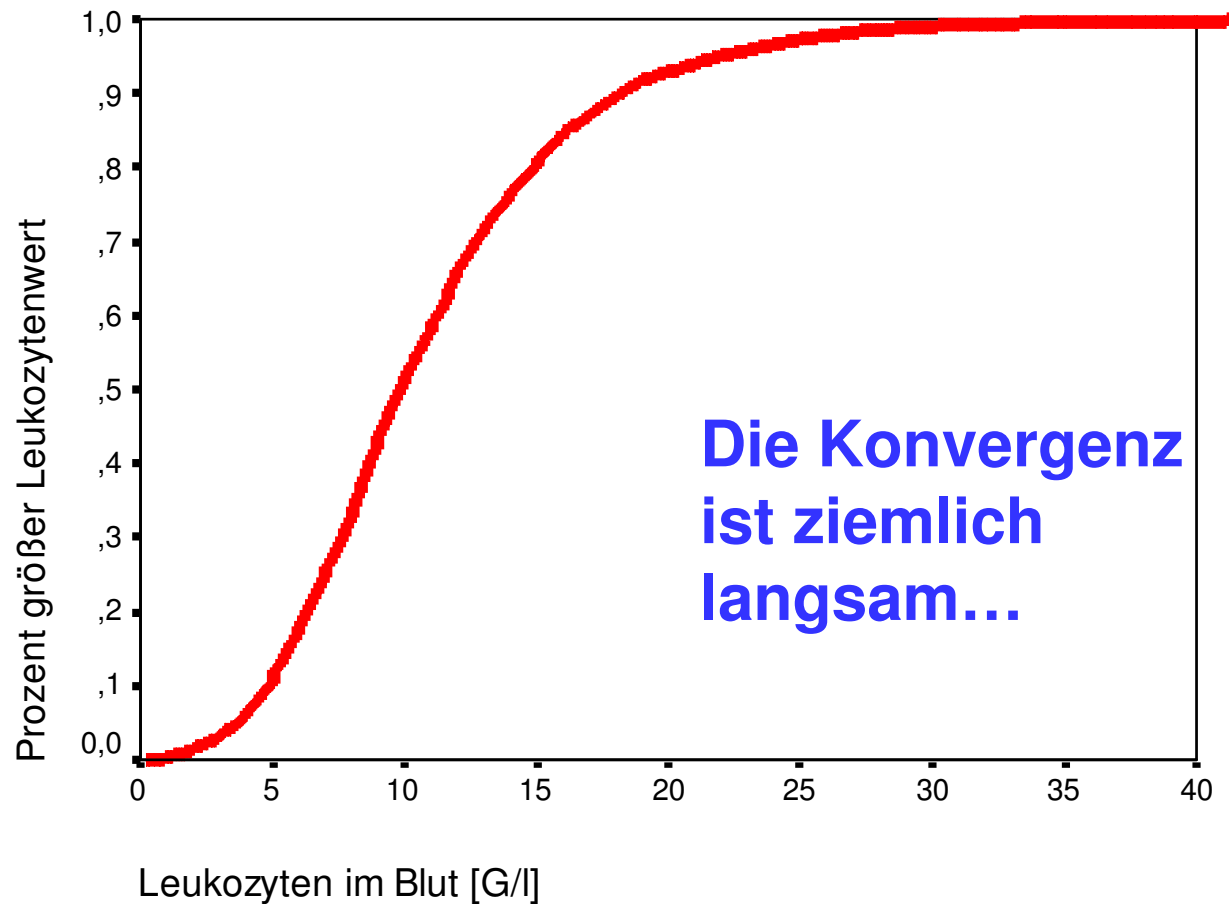
Schätzung einer Verteilung – Effekt der Fallzahl

9 Stichproben mit 400 Patienten



Schätzung einer Verteilung – Effekt der Fallzahl

Schätzung mit N=4725 Patienten



Schätzfunktionen (Schätzer, estimators)

Seien $X_1, \dots, X_n$ unabhängige, identisch - verteilte Zufallsvariable mit Verteilungsfunktion $F(x, \theta)$ sowie Verteilungsdichte $f(x, \theta)$.

Schätzproblem :

Gegeben eine Stichprobe des Umfangs n : $x_1, \dots, x_n$

Bestimme θ so gut wie möglich!

Schätzfunktion für θ :

Stichprobenfunktion $\hat{\theta}_n = \hat{\theta}_n(X_1, \dots, X_n)$

$\hat{\theta}_n$ heißt **konsistenter** Schätzer für θ , wenn für Realisierungen

$x_1, \dots, x_n, \dots$ und $n \rightarrow \infty$ die Folge

$\hat{\theta}_n(x_1, \dots, x_n)$ stochastisch gegen θ konvergiert, d.h.

$$\lim_{n \rightarrow \infty} \text{pr}(|\hat{\theta}_n(x_1, \dots, x_n) - \theta| > \varepsilon) = 0$$

Erwartungstreue Schätzer

Die Schätzfunktion $\hat{\theta}_n$ heißt *erwartungstreu* $\Leftrightarrow$

$$E(\hat{\theta}_n) = \theta$$

=

Die Schätzfunktion $\hat{\theta}_n$ heißt *asymptotisch erwartungstreu* $\Leftrightarrow$

$$\lim_{n \rightarrow \infty} E(\hat{\theta}_n) = \theta$$

Erwartungstreue Schätzer II

Beispiele :

(1) $\hat{p}_n = \frac{k}{n}$ ist ein (konsistenter,) erwartungstreuer Schätzer für die Wahrscheinlichkeit p , die einer Binomialverteilung zugrunde liegt.

(2) $\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$ ist nur asymptotisch erwartungstreu als Schätzer für die Varianz einer Normalverteilung.

Steinerscher Verschiebungssatz : $\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 - (\bar{x} - \mu)^2$

(Beweis : - $\mu + \mu$ einfügen und ausquadrieren!)

$$\begin{aligned} E(\hat{\sigma}_n^2) &= E\left(\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 - (\bar{x} - \mu)^2\right) = \frac{1}{n} \sum_{i=1}^n E((x_i - \mu)^2) - E((\bar{x} - \mu)^2) \\ &= \frac{1}{n} n \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n} \sigma^2 \quad (\text{DAHER schätzt man mit } \hat{\sigma}'_n = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 !!) \end{aligned}$$

Vergleich von Schätzfunktionen

Mittlerer quadratischer Fehler (MSE = mean square error) :

$$\begin{aligned} E[(\theta - \hat{\theta}_n)^2] &= E[(\theta - E(\hat{\theta}_n) + E(\hat{\theta}_n) - \hat{\theta}_n)^2] = \\ &= E[(\theta - E(\hat{\theta}_n))^2] + E[(E(\hat{\theta}_n) - \hat{\theta}_n)^2] = \\ &= \text{Bias}(\hat{\theta}_n)^2 + \text{var}(\hat{\theta}_n) \end{aligned}$$

systematische Verzerrung + Zufallsstreuung

Kriterien :

Konventionell sucht man unter den erwartungstreuen Schätzfunktionen

(d.h. $\text{Bias}(\hat{\theta}_n) = 0$) diejenige mit der kleinsten Varianz.

CAVE : Es gibt Fälle, in denen der Schätzer mit dem geringsten mittleren quadratischen Fehler nicht einmal asymptotisch erwartungstreu ist.

Konstruktion von Schätzfunktionen

Momentenmethode

Seien X_i unabhängig und identisch - verteilte Zufallsvariable mit Verteilungsdichte $f(x, \theta)$.

Bestimme die theoretischen k - ten Momente :

$$\mu_k(\theta) := E(X^k) = \int x^k f(x, \theta) dx$$

als Funktion von θ .

Die entsprechenden empirischen k - ten Momenten sind :

$$\hat{\mu}_k = \frac{1}{n} \sum_i x_i^k$$

Durch Gleichsetzen $\mu_k(\hat{\theta}) = \hat{\mu}_k$ (für genügend viele k) entsteht ein Gleichungssystem für $\hat{\theta}$.

Durch Auflösen nach $\hat{\theta}$ ergibt sich ein Schätzfunktion (Momentenschätzer) für θ .

Konstruktion von Schätzfunktionen

Momentenmethode Beispiel

Seien $X_i \sim N(\mu, \sigma^2)$ unabhängig und identisch normalverteilt.

$$\theta = (\mu, \sigma^2)$$

$$\mu_1(\theta) = \int x \cdot \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx = \mu$$

$$\mu_2(\theta) = \int x^2 \cdot \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx = \mu^2 + \sigma^2$$

Für eine Stichprobe $x_1, \dots, x_n$

$$\hat{\mu}_1 = \frac{1}{n} \sum_i x_i$$

$$\hat{\mu}_2 = \frac{1}{n} \sum_i x_i^2$$

Gleichsetzen :

$$\hat{\mu} = \hat{\mu}_1 = \frac{1}{n} \sum_i x_i = \bar{x} \text{ arithmetisches Mittel}$$

$$\hat{\sigma}^2 = \hat{\mu}_2 - \hat{\mu}^2 = \frac{1}{n} \sum_i x_i^2 - \left(\frac{1}{n} \sum_i x_i \right)^2 = \frac{1}{n} \sum_i (x_i - \bar{x})^2 \quad (\text{Nicht erwartungstreier Varianzschätzer})$$

Konstruktion von Schätzfunktionen

Maximum Likelihood Methode

Seien X_i unabhängig und identisch - verteilte Zufallsvariable mit Verteilungsdichte $f(x, \theta)$.

Sei $x_1, \dots, x_n$ eine Stichprobe

Die Likelihood - Funktion zu dieser Stichprobe ist definiert als

$$\text{likelihood}(\theta) := \prod_{i=1}^n f(x_i, \theta)$$

Der Maximum Likelihood Schätzer für θ ist definiert als der Wert von θ , der die Likelihoodfunktion maximiert.

$$\hat{\theta}_{ML} := \operatorname{argmax}(\text{likelihood}(\theta))$$

Konstruktion von Schätzfunktionen

Maximum Likelihood Methode

Der Maximum Likelihood Schätzer $\hat{\theta}_{ML} := \operatorname{argmax}(\text{likelihood}(\theta))$

ist eine - oft nur numerisch bestimmbare - Stichprobenfunktion
und damit eine Zufallsvariable.

Unter schwachen Regularitätsannahmen gilt :

SATZ :

(1) $\hat{\theta}_{ML}$ ist konsistent und asymptotisch erwartungstreu.

(2) Asymptotisch, d.h. für große Stichproben, ist $\hat{\theta}_{ML}$ approximativ (multivariate) normalverteilt,
mit einer Varianz - Kovarianzmatrix, die sich aus der zweiten Ableitung der Likelihoodfunktion
schätzen läßt.

$$\hat{\theta}_{ML} \sim N(\theta, I^{-1}(\theta))$$

Dabei ist $I = E(\text{Matrix der zweiten Ableitungen der Likelihood funktion an der Stelle } \theta)$

I kann durch Matrix der zweiten Ableitung der Likelihoodfunktion an der Stelle $\hat{\theta}_{ML}$ geschätzt werden.

Konstruktion von Schätzern in anderen Paradigmen

- **Likelihoodians** verwenden natürlich ebenfalls den Maximum Likelihood Schätzer.
- **Bayesianer** verwenden
 - Erwartungswert der Posterior-Verteilung
–oder
 - Maximum Posterior Density estimate

Frequentistische Konfidenzintervalle I

■ Gegeben:

- ▶ Modellfamilie mit interessierendem Parameter θ
- ▶ Stichprobe (Daten) x
- ▶ Schätzer $\hat{\theta}(x)$ für θ aus der Stichprobe

■ Problemstellung:

- ▶ $\hat{\theta}(x)$ ist eine Zufallsvariable
- ▶ Beurteile die Genauigkeit der Schätzung
- ▶ Ideal wäre $|\hat{\theta}(x) - \theta|$, aber θ ist unbekannt..

Frequentistische Konfidenzintervalle II

- Ansatz:

- ▶ Statt Genauigkeit der konkreten Schätzung $\hat{\theta}(x)$ beschreibe die Genauigkeit des Schätzverfahrens $\hat{\theta}(\cdot)$ bei wiederholter Anwendung.

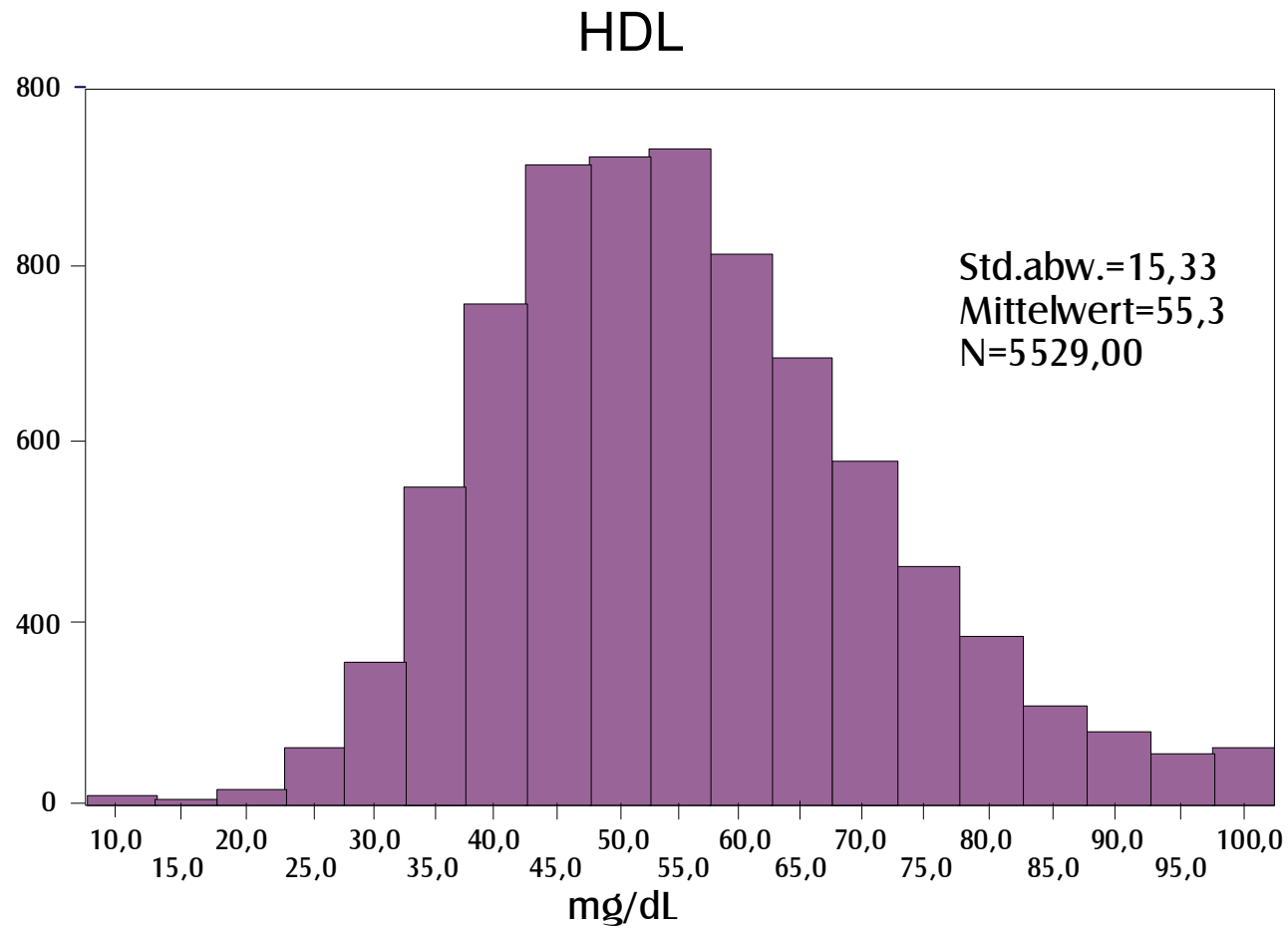
- Vorgehen: Berechne (als Stichprobenfunktion) ein zufälliges Intervall $CI^{\wedge}(x)$ mit folgenden Eigenschaften:

- ▶ $\hat{\theta}(x)$ in $CI^{\wedge}(x)$ per constructionem und
- ▶ **Überdeckungswahrscheinlichkeit $\Pr(\theta \text{ in } CI^{\wedge}(X)) \geq 1-\alpha$** für virtuelle Wiederholungen des Versuchs X und alle relevanten θ .

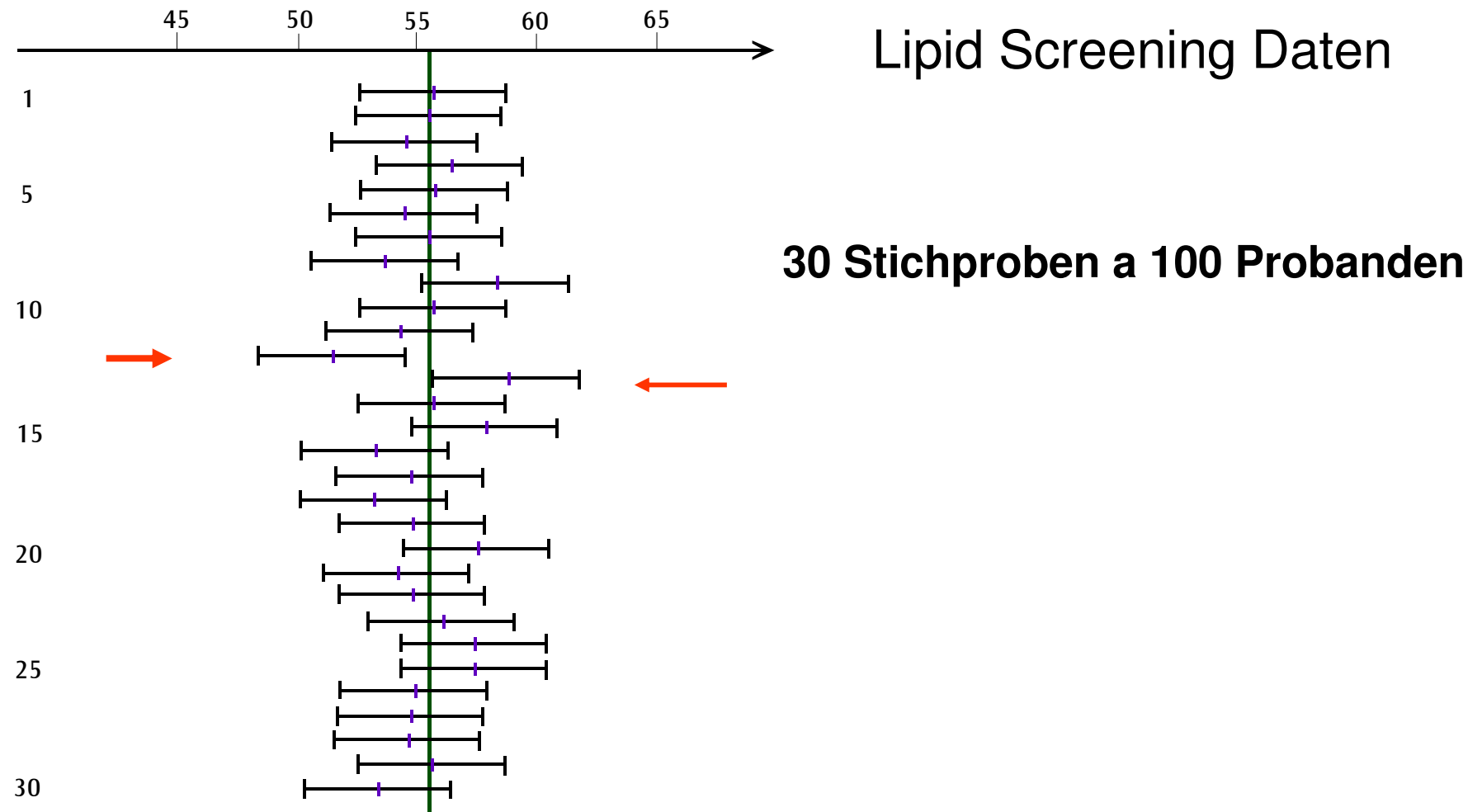
Bedeutung: 95%- Konfidenzintervall

- Bei **wiederholter unabhängiger Anwendung des Konstruktionsverfahrens** enthält das (zufällige) Konfidenzintervall in 95% aller Fälle den „wahren“ Wert.
- Im **konkret gegebenen** Fall liegt jeweils der wahre Wert
im Intervall oder auch nicht !?!
- Wahrscheinlichkeitsaussagen darüber, ob der wahre Wert in einem gegebenen Intervall liegt, können nur Bayesianer formulieren...

Lipid-Screening



Konfidenzintervall: Überdeckungswahrscheinlichkeit bei wiederholten Versuchen



Dualität: Signifikanztests – Konfidenzintervalle I

Gegeben eine Familie von Modellen $M(\theta)$ mit $\theta \in \Omega$ und Daten x .

(1) Sei $\Psi(\theta_0)$ eine Familie $\theta_0 \in \Omega$ von Signifikanz - Tests zum Niveau α der Hypothesen $H(\theta_0) : \theta = \theta_0$

$$\Psi(\theta_0)(x) = \begin{cases} 1 : \text{Ablehnen} \\ 0 : \text{Beibehalten} \end{cases}$$

Dann definiert $CI(x) = \{\theta \in \Omega \mid \Psi(\theta)(x) = 0\}$ eine $1 - \alpha$ Konfidenzmenge.

Beweis : $\text{pr}(\theta \in CI(x)) = \text{pr}(\Psi(\theta)(x) = 0) = 1 - \alpha$

Dualität: Signifikanztests – Konfidenzintervalle II

(2) Sei CI ein Verfahren zur Bestimmung von $1 - \alpha$ Konfidenzmengen,
d.h. $\text{pr}(\theta \in \text{CI}(x)) = 1 - \alpha$

Dann definiert

$$\Psi(\theta)(x) = \begin{cases} 1 & \text{falls } \theta \notin \text{CI}(x) \\ 0 & \text{falls } \theta \in \text{CI}(x) \end{cases}$$

eine Familie von Niveau α Signifikanztests.

Likelihood Präzisionsintervalle I

■ Gegeben:

- ▶ Modellfamilie mit interessierendem Parameter θ
- ▶ Stichprobe (Daten) x

■ Problemstellung:

- ▶ Maximum Likelihood Schätzer $\theta_{ML}^{\wedge}$
- ▶ Beschreibe eine Menge von Werten θ , die immer noch plausible sind gegeben die Daten, d.h. gegen welche die Daten keine starke Evidenz darstellen.

Likelihood Präzisionsintervalle II

- Normierte Likelihoodfunktion:
 - ▶ $R(\theta|x) = L(\theta|x) / L(\theta^{\wedge}_{ML}|x) \leq 1$

- Präzisionsintervall (Menge = plausibler Modelle)
 - ▶ $PI(x) = \{\theta | R(\theta|x) \geq r\}$
 - ▶ Z.B. $r = 1/8, 1/32, ..$

Normierte Likelihoodfunktion

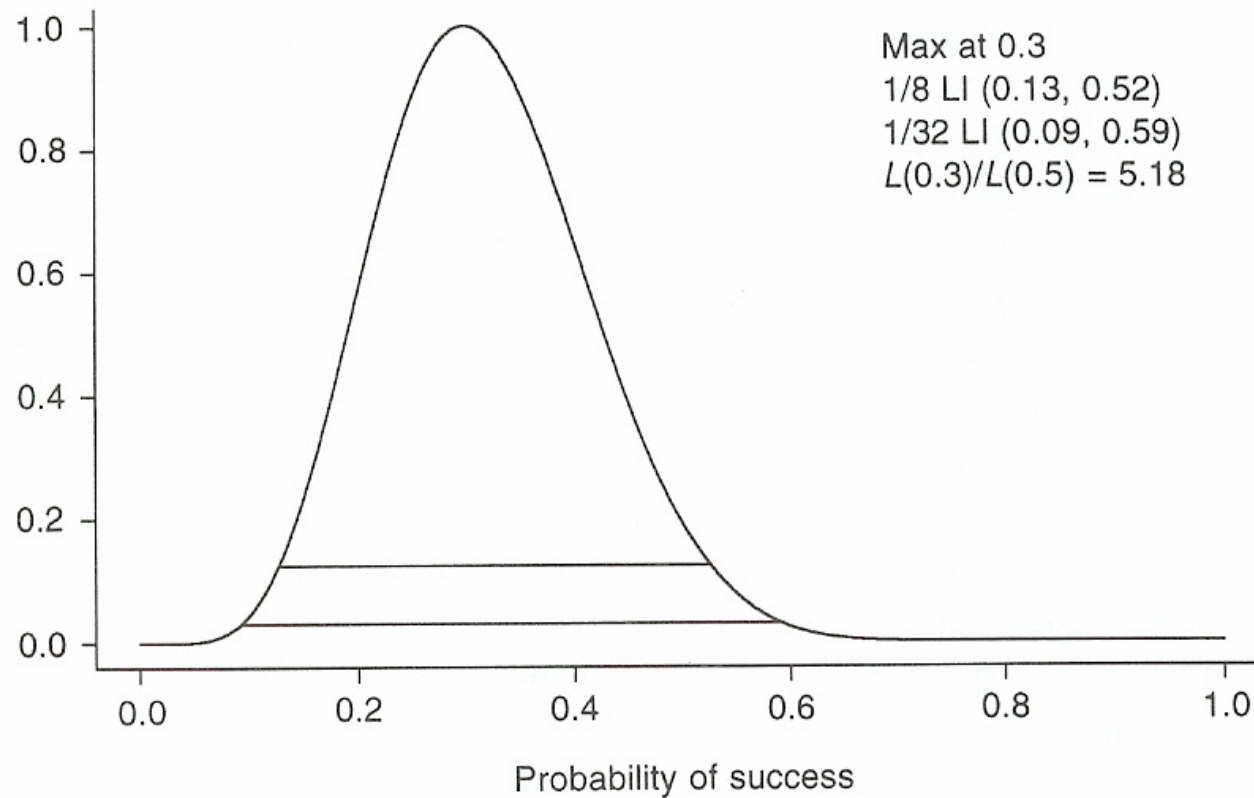


Figure 1.1 *Likelihood for probability of success: six successes observed in 20 trials.*

Bayesianische Credible intervalls I

- Gegeben:

- ▶ Modellfamilie mit interessierendem Parameter θ
- ▶ Stichprobe (Daten) x

- Problemstellung:

- ▶ Bestimme Intervall (Menge) in dem mit vorgegebener (Bayes'scher) Wahrscheinlichkeit $1-\alpha$ der wahre Parameter θ liegt.

Bayesianische Credible intervalls II

Bayes'sches Lernen :

$$\text{Posterior } pr(\theta | x) = \frac{\text{Likelihood * Prior } pr(x | \theta) * pr(\theta)}{\text{Normierungskonstante } \int_{\Omega} pr(x | \theta) * pr(\theta) d\theta}$$

$pr(\theta | x)$ beschreibt die Bayes'sche Wahrscheinlichkeit der Modelle $M(\theta)$ nach Beobachtung von x .

Ein $1 - \alpha$ Credible Intervall ist ein Intervall $Crl(x)$ mit

$$\int_{Crl(x)} pr(\theta | x) d\theta = 1 - \alpha$$

Sprechweise : "Mit 95% Bayes'scher Wahrscheinlichkeit liegt $\theta \in Crl(x)$."

Achtung : Nicht frequentistischer Wahrscheinlichkeitsbegriff!