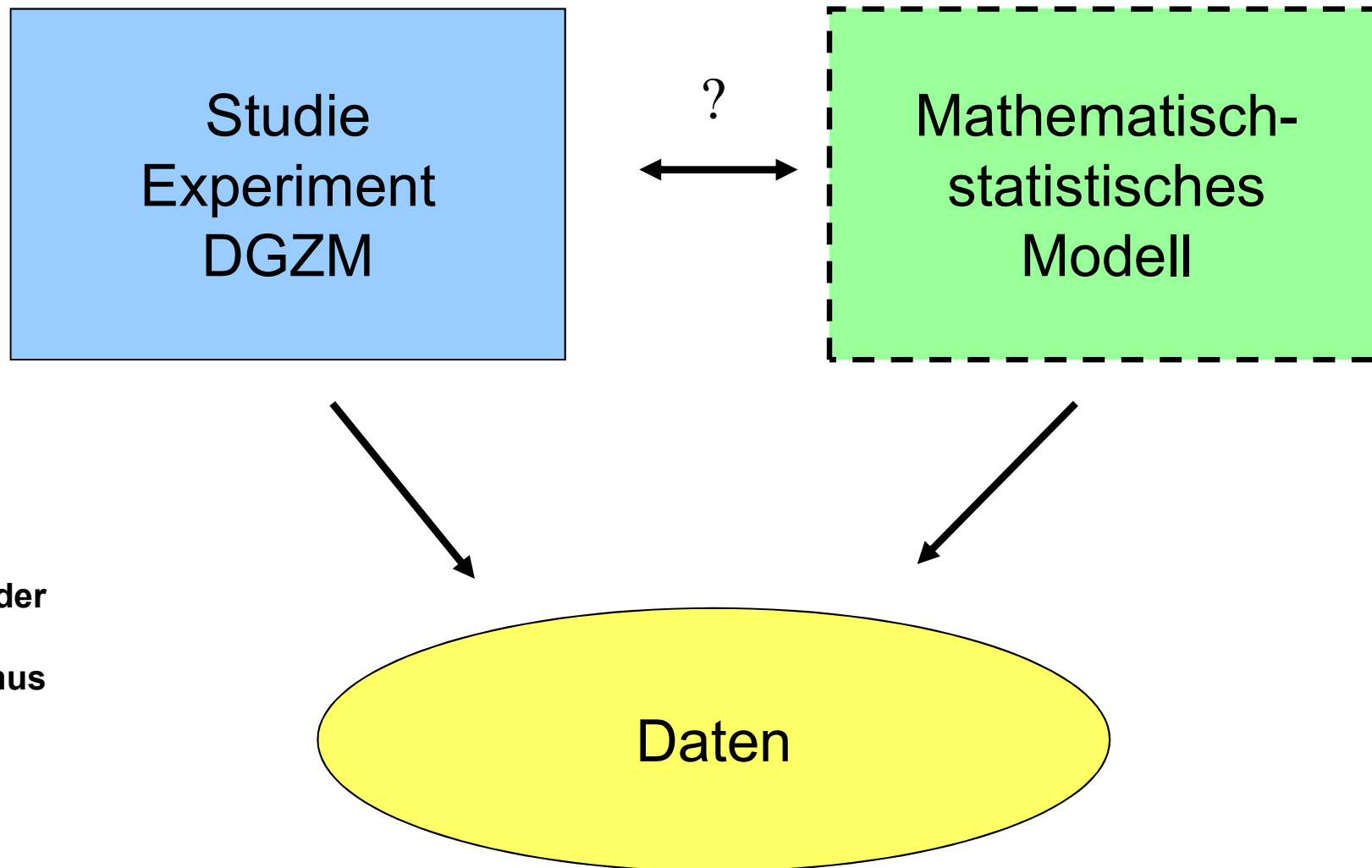

Unabhängigkeit und Grenzwertsätze

Dr. Dirk Hasenclever
University of Leipzig

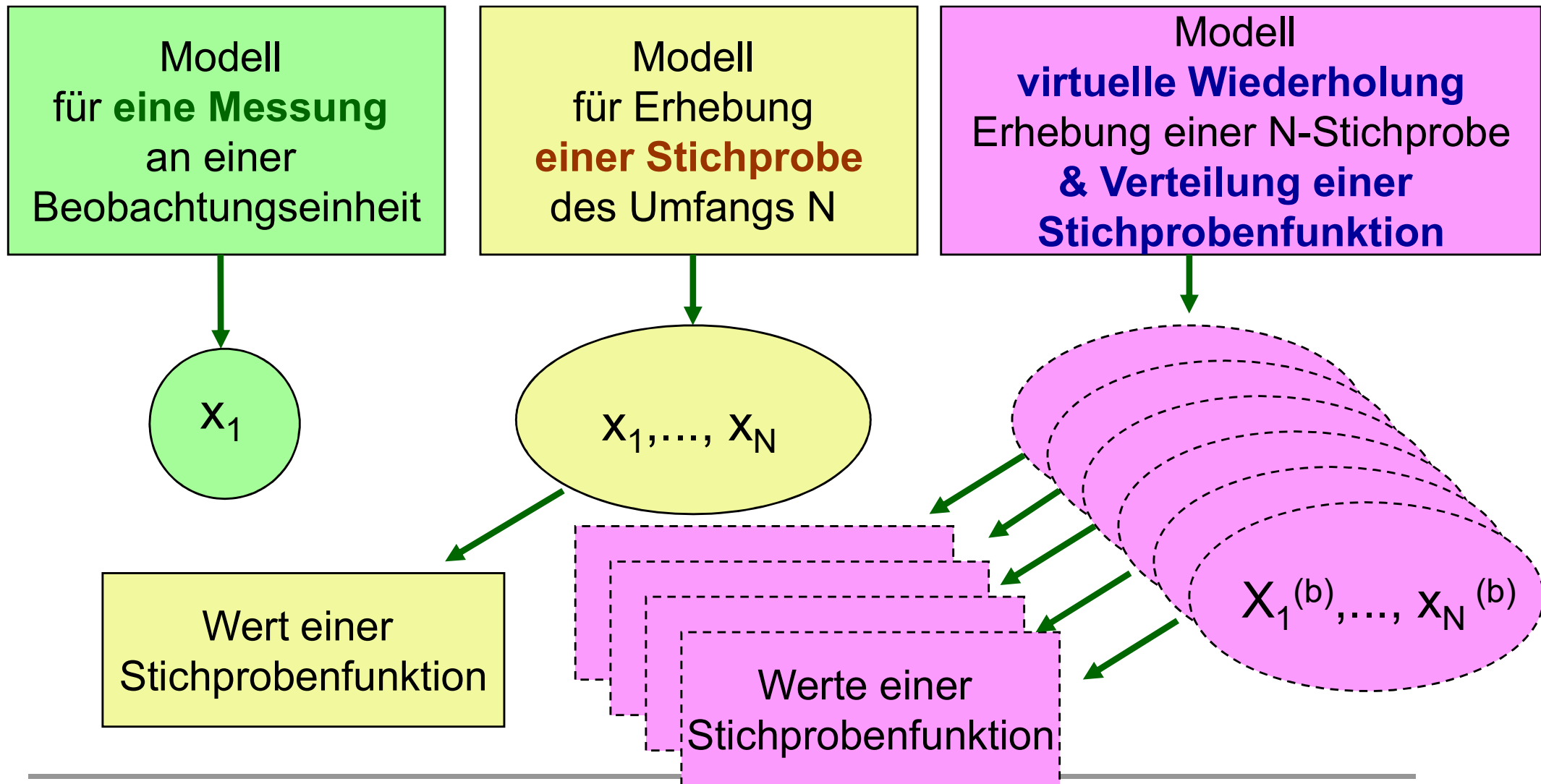
[imise.](#)

Statistische Modellierung



DGZM=
Daten
generierender
Zufalls-
mechanismus

Modellierung von Zufallsexperimenten



Einfachstes Modell für Erhebung einer Stichprobe

- Die Messwerte in der Stichprobe $x_1, \dots, x_N$ sind Realisierungen von Zufallsvariablen $X_1, \dots, X_N$, welche
 - **unabhängig** und
 - **identisch verteilt** sind.
- Der Stichprobenumfang **N ist vorab fest vorgegeben** bzw. unabhängig von den Realisierungen von $X_1, \dots, X_N$.

Vorsicht: Beispiel für unterschiedliche Stichprobenerhebungsregeln bei gleicher Einzelbeobachtung

- Einzelmessung: $X \sim \text{Bernoulli}(p)$
- **Binomial**: Wiederhole Versuch N mal unter identischen Bedingungen: $Y_1, \dots, Y_N$ unabhängig $\sim \text{Bernoulli}(p)$
Stichprobenfunktion: $X := Y_1 + \dots + Y_N$

$$pr(X = k) = \binom{N}{k} p^k (1-p)^{N-k}$$

- **Negativ binomial**: Wiederhole Versuch solange unter identischen Bedingungen bis **r** Erfolge beobachtet.
Sei **n** die Gesamtzahl der Versuche:
Stichprobenfunktion: $X := Y_1 + \dots + Y_n$

$$pr(X = n) = \binom{n-1}{r-1} * p^r * (1-p)^{n-r}$$

Wann sind Messungen unabhängig?

- **Intuitiv:** Zwei Messungen X und Y sind **unabhängig**, wenn die **Beobachtung $Y=y$ keinerlei Information über die Realisierung von X enthält.**
 - Beim Roulette ist bei jedem Wurf die Wahrscheinlichkeit für **“rouge”** $=18/37$, egal ob zuvor zehnmal hintereinander **“noir”** gekommen ist.
- Um dies mathematisch zu präzisieren, müssen wir Zufallsvariable mit Werten im $\mathbb{R}^n$ betrachten:

Gemeinsame Verteilung von zwei Zufallsvariablen

X, Y reelle Zufallsvariable.

Gemeinsame Verteilung bei diskreten Zufallsvariablen
ist gegeben durch die Wahrscheinlichkeitsfunktion:

$$pr(X = x \wedge Y = y)$$

Gemeinsame Verteilungsfunktion bei stetigen Zufallsvariablen:

$$F_{X,Y}(x,y) = pr(X \leq x \wedge Y \leq y)$$

mit gemeinsamer Dichtefunktion $f_{X,Y}(x,y) dx dy$

$$pr((X,Y) \in A) = \int_A f_{X,Y}(x,y) dx dy$$

$$\text{speziell: } F_{X,Y}(x,y) = \int_{-\infty}^x \int_{-\infty}^y f_{X,Y}(x,y) dx dy$$

Randverteilungen

Diskret: $f_{X,Y}(x,y) = pr(X = x \wedge Y = y)$

Randverteilung von X unter Absehung/in Unkenntnis von Y :

$$f_X(x) = pr(X = x) = \sum_{y \in \text{Träger}(Y)} f_{X,Y}(x,y)$$

Stetig: $f_{X,Y}(x,y) dx dy$ Dichtefunktion:

Randverteilung von X unter Absehung/in Unkenntnis von Y :

$$f_X(x) dx = \left(\int_{-\infty}^{\infty} f_{X,Y}(x,y) dy \right) dx$$

“Ausintegrieren!”

Additivität von Erwartungswert

Notiz:

Seien X, Y nicht notwendig unabhängige Zufallsvariable,
für die $E(X)$ und $E(Y)$ existieren.

Dann gilt: $E(X + Y)$ existiert und es gilt:

$$E(X + Y) = E(X) + E(Y)$$

Beweis:

$$E(X + Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x + y) \cdot f_{X, Y}(x, y) dx dy =$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x \cdot f_{X, Y}(x, y) dx dy + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y \cdot f_{X, Y}(x, y) dx dy =$$

$$\int_{-\infty}^{\infty} x \cdot \left(\int_{-\infty}^{\infty} f_{X, Y}(x, y) dy \right) dx + \int_{-\infty}^{\infty} y \cdot \left(\int_{-\infty}^{\infty} f_{X, Y}(x, y) dx \right) dy =$$

$$\int_{-\infty}^{\infty} x \cdot f_X(x) dx + \int_{-\infty}^{\infty} y \cdot f_Y(y) dy = E(X) + E(Y)$$

Definition: Bedingte Verteilung

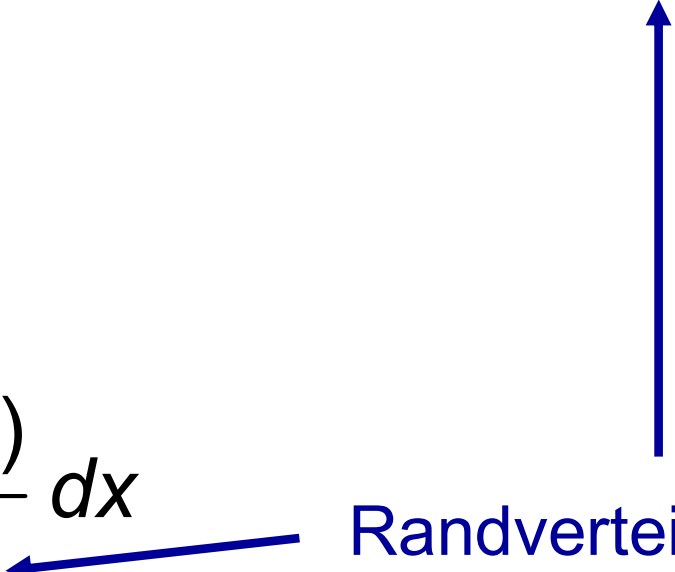
Diskret:

$$f_{X|Y}(x | y) = pr(X = x | Y = y) = \frac{pr(X = x \wedge Y = y)}{pr(Y = y)}$$

Stetig:

$$f_{X|Y}(x | y) dx = \frac{f_{X,Y}(x, y)}{f_Y(y)} dx$$

Randverteilung von Y



Definition Unabhängigkeit

Definition und Satz:

Die folgenden drei Aussagen sind äquivalent:

(1) X und Y sind unabhängig

$$(2) \quad f_{X,Y}(x,y) dx dy = f_X(x) \cdot f_Y(y) dx dy$$

(Dichte = Produkt der Randverteilungen)

$$(3) \quad f_{X|Y}(x | y) dx = f_X(x) dx$$

(Y enthält keine Information über X)

Bemerkung: Weil $f_{X,Y}(x,y) dx dy = f_{Y|X}(y | x) \cdot f_X(x) dx dy$

(nach Definition) folgt bei Unabhängigkeit auch

$$f_{Y|X}(y | x) dy = \frac{f_{X,Y}(x,y) dy}{f_X(x)} = \frac{f_X(x) \cdot f_Y(y) dy}{f_X(x)} = f_Y(y) dy.$$

Kovarianz

$$\text{cov}(X, Y) := E((X - E(X)) \cdot (Y - E(Y)))$$

Eigenschaften:

- a) $\text{cov}(X, Y) = \text{cov}(Y, X)$ (Symmetrie)
- b) $\text{cov}(a + X, Y) = \text{cov}(X, Y)$ (translationsinvariant)
- c) $\text{cov}(\alpha X + \beta Z, Y) = \alpha \cdot \text{cov}(X, Y) + \beta \cdot \text{cov}(Z, Y)$ (linear)

Folgen direkt aus Eigenschaften des Erwartungswerts.

- d) $\text{cov}(X, X) \geq 0$ und $\text{cov}(X, X) = 0$ genau dann wenn $X \equiv 0$

E definiert ein Skalarprodukt auf ZV mit demselben Träger

Unabhängigkeit und Kovarianz

Notiz: Sind X und Y unabhängig, so gilt:

$$\text{cov}(X, Y) = 0$$

Beweis: $E((X - E(X)) \cdot (Y - E(Y))) =$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - E(X)) \cdot (y - E(Y)) \cdot f_{X,Y}(x, y) dx dy =$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - E(X)) \cdot (y - E(Y)) \cdot f_X(x) \cdot f_Y(y) dx dy =$$

$$\int_{-\infty}^{\infty} (y - E(Y)) \cdot f_Y(y) \left(\int_{-\infty}^{\infty} (x - E(X)) \cdot f_X(x) dx \right) dy = 0$$

Vorsicht: Umkehrung gilt nicht!

Beispiel: $\text{cov}=0$, aber nicht unabhängig

Wir betrachten die Gleichverteilung auf der Einheitskreisscheibe in kartesischen Koordinaten:

$$f_{X,Y}(x,y)dx dy = 1/\pi dx dy \quad (\text{für } 0 \leq x,y \leq 1; \text{ und } 0 \text{ außerhalb}).$$

X, Y entstehen als Randverteilungen mit Träger $[-1,1]$.

$$\begin{aligned} \text{cov}(X,Y) &= \iint_{\square} 1/\pi x \cdot y dx dy = 1/\pi \int_{-1}^1 x \left(\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} y dy \right) dx \\ &= 1/\pi \int_{-1}^1 x \left[\frac{y^2}{2} \right]_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} dx = 0 \end{aligned}$$

Offenbar sind aber X und Y nicht unabhängig, da der Träger von $Y | X$ von X abhängt.

Unabhängigkeit und Varianz

$$\begin{aligned}\text{var}(X+Y) &= \text{cov}(X+Y, X+Y) \\ &= \text{cov}(X, X) + \text{cov}(Y, Y) + 2 \cdot \text{cov}(X, Y) \\ &= \text{var}(X) + \text{var}(Y) + 2 \cdot \text{cov}(X, Y)\end{aligned}$$

Notiz: Sind X und Y unabhängig, so gilt:

$$\text{var}(X+Y) = \text{var}(X) + \text{var}(Y)$$

Korrelation

Die Kovarianz ändert sich bei Skalentransformationen

$$X \rightarrow \alpha X \text{ oder } Y \rightarrow \beta Y$$

Daher definiert man die Korrelation

$$\text{cor}(X, Y) := \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)} \cdot \sqrt{\text{var}(Y)}}$$

Eigenschaften:

$$E(X \cdot Y)^2 \leq E(XX) \cdot E(YY)$$

a) $-1 \leq \text{cor}(X, Y) \leq +1$ (Beweis mit Cauchy-Schwarz Ungleichung)

b) Die Korrelation ist invariant unter affinen Transformationen

$$X \rightarrow \tau + \alpha X \text{ oder } Y \rightarrow \tau + \beta Y.$$

c) Wenn $Y = a + bX$: $\text{cor}(X, Y) = \frac{b \cdot \text{cov}(X, X)}{\sqrt{\text{var}(X)} \cdot |b| \cdot \sqrt{\text{var}(X)}} = \text{sign}(b)$

Empirische und theoretische Verteilungsfunktion I

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

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

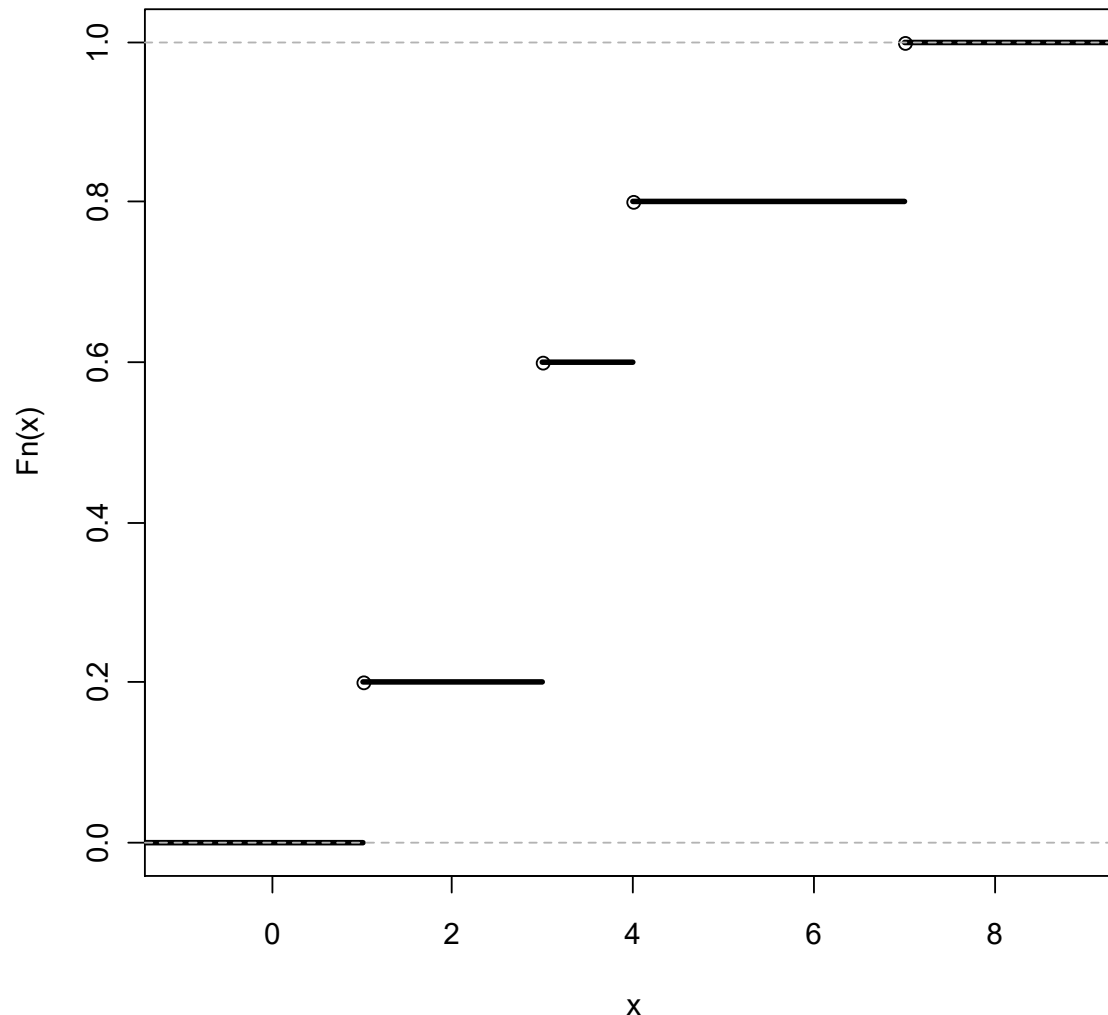
Wegen des Bernoulli'schen (schwachen) Gesetz der großen Zahl wissen wir, dass punktweise für festes x

$$\forall \varepsilon > 0 : \lim_{N \rightarrow \infty} \text{pr} \left(|\hat{F}_X^{(N)}(x) - F_X(x)| > \varepsilon \right) = 0$$

Gilt dies auch gleichmäßig auf dem ganzen Träger von X ?

Empirische Verteilungsfunktion ist eine Treppenfunktion

Empirische Verteilungsfunktion zu $x \leftarrow c(1,3,3,4,7)$



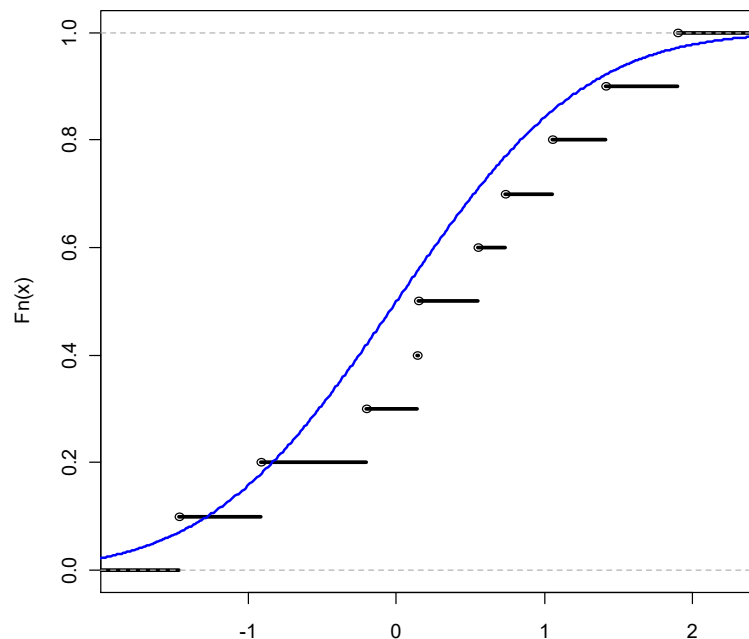
```
x<-c(1,3,3,4,7)  
plot(ecdf(x),lwd=3)
```

Empirische und theoretische Verteilungsfunktion II

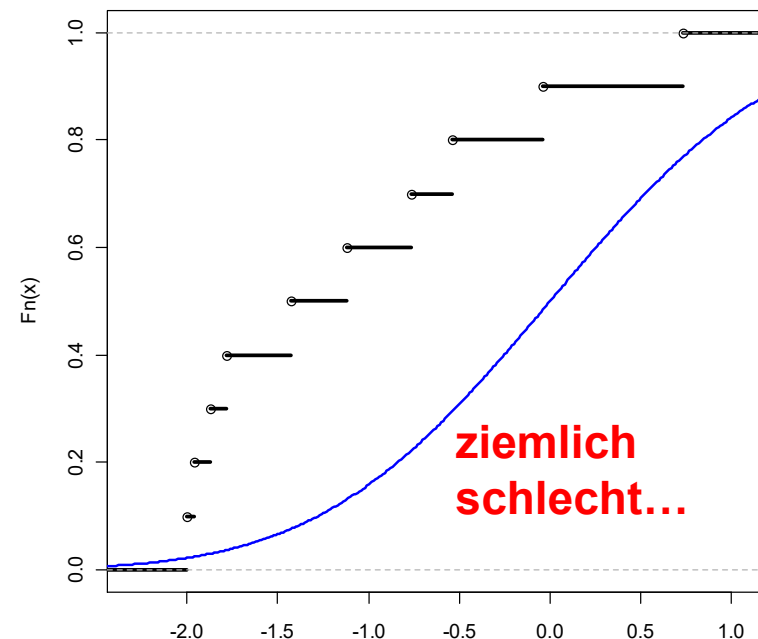
Als Maß für die gleichmäßige Approximation wählt man

$$\Delta_N := \sup_{-\infty < x < +\infty} |\hat{F}_X^{(N)}(x) - F_X(x)| =: \|\hat{F}_X^{(N)} - F_X\|_\infty$$

der Abstand zwischen empirischer und theoretischer Verteilungsfunktion in der Supremumsnorm.



N=10



Empirische und theoretische Verteilungsfunktion III

Hauptsatz der mathematischen Statistik (Gliwenko)

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

(Wahrscheinlichkeit über unendliche Serien von Einzelmessungen)

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

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

Empirische und theoretische Verteilungsfunktion IV (Zusatzinformation)

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 ist der Hauptsatz auch operational:

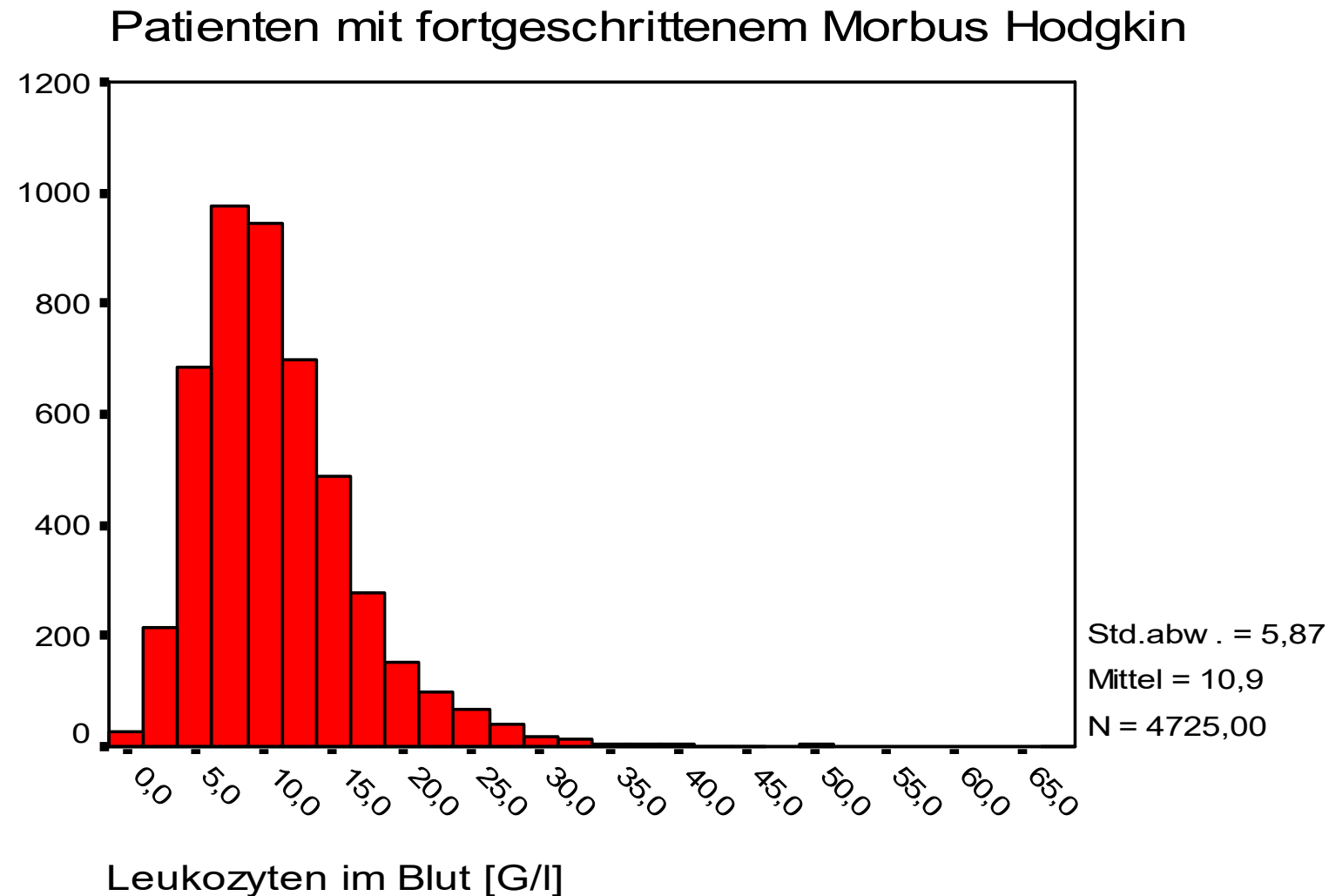
(1) zu $1-\delta$ suche y mit $K(y) = 1-\delta$

(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.)

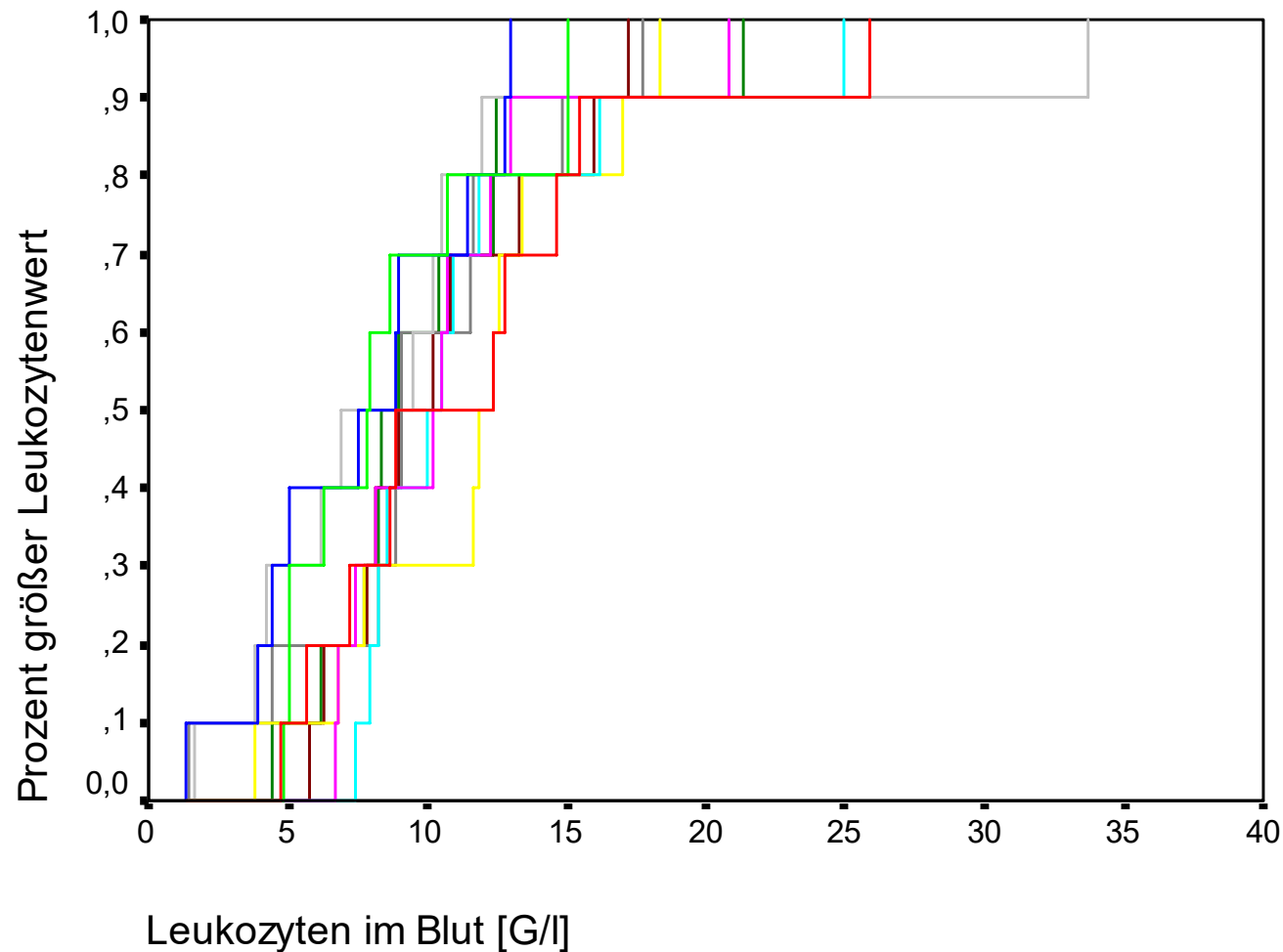
Illustration zum Hauptsatz der Statistik

Ziehe aus sehr großem Datensatz



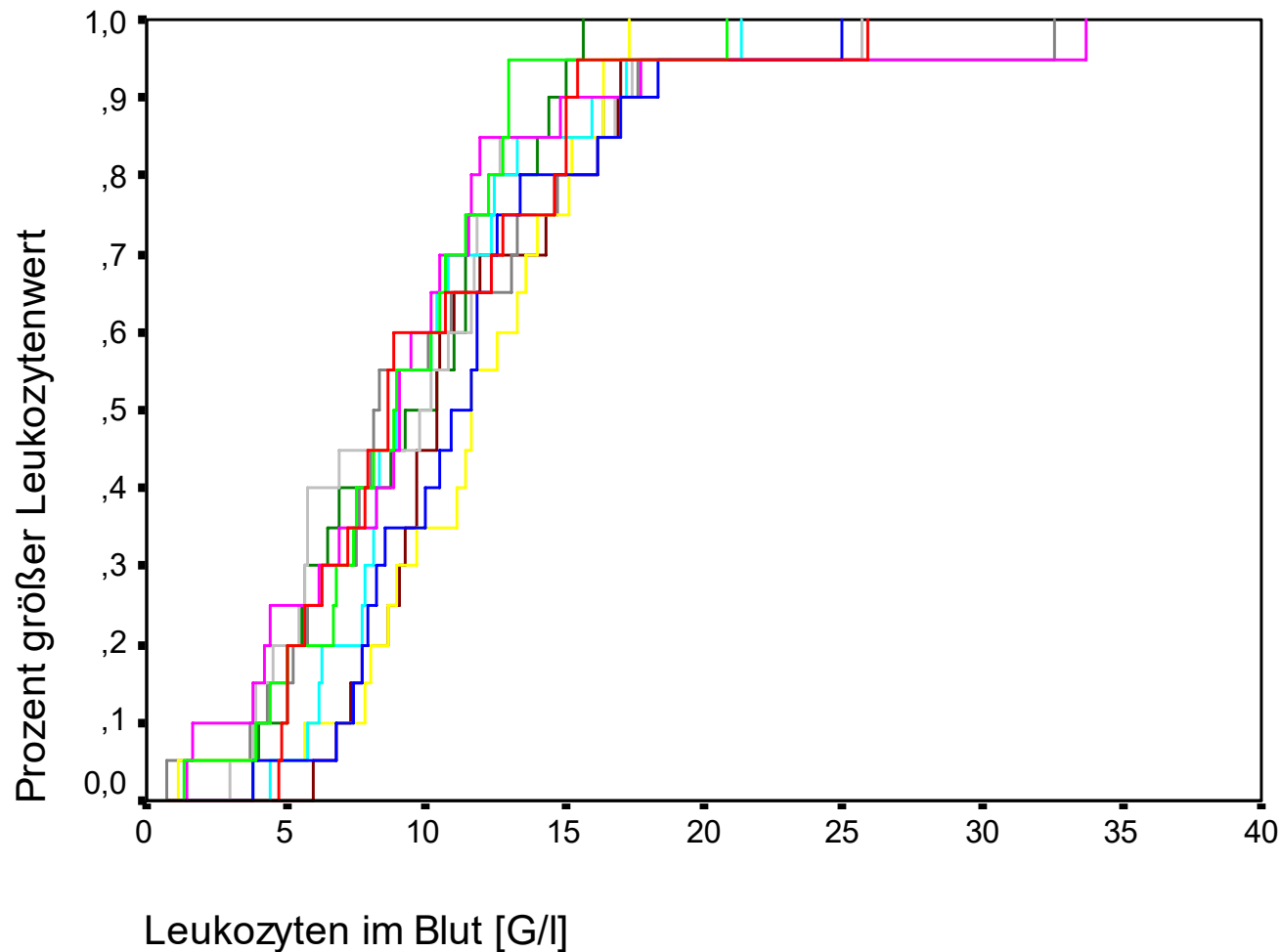
Schätzung einer Verteilung – Effekt der Fallzahl

9 Stichproben mit 10 Patienten



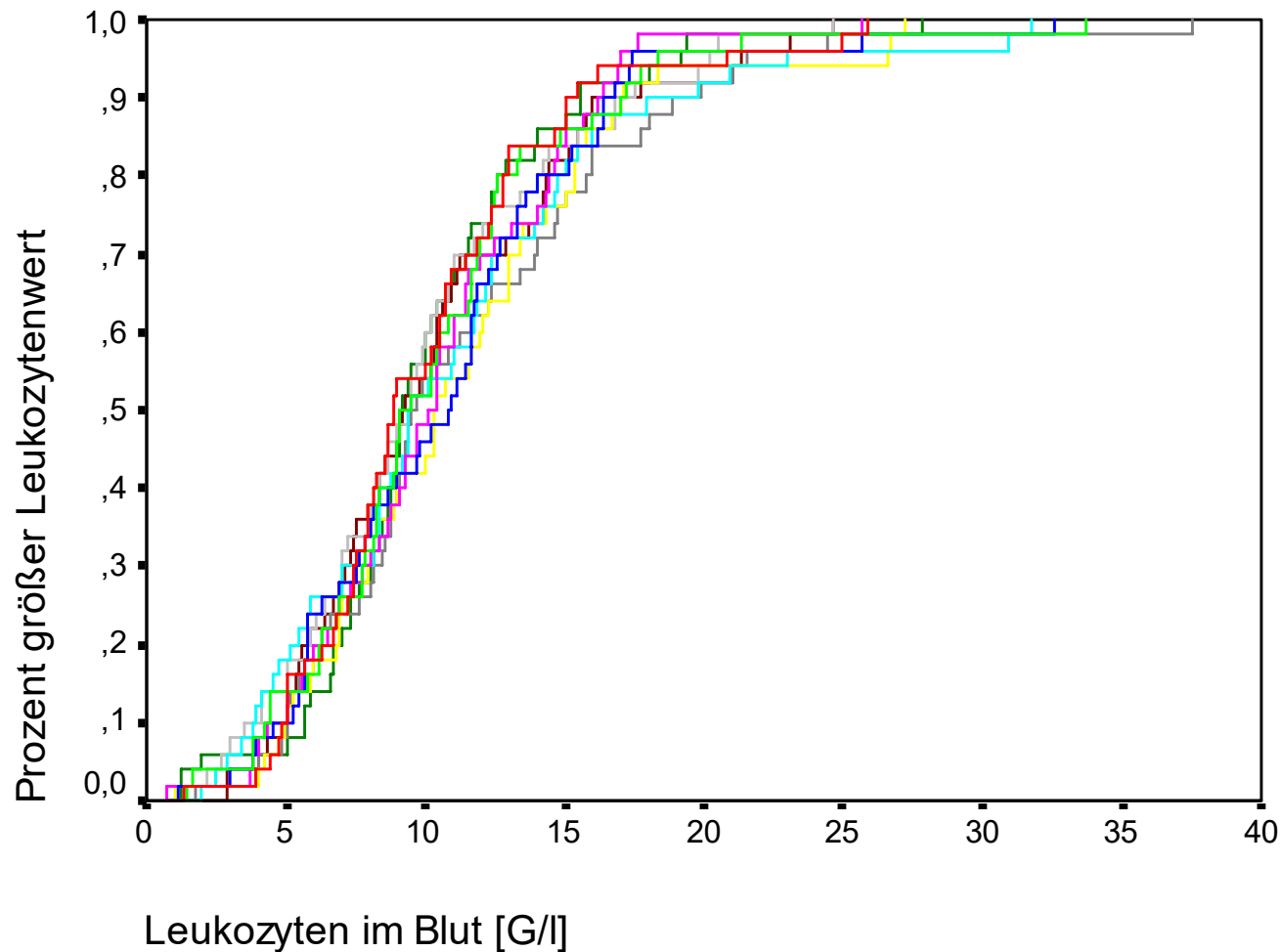
Schätzung einer Verteilung – Effekt der Fallzahl

9 Stichproben mit 20 Patienten



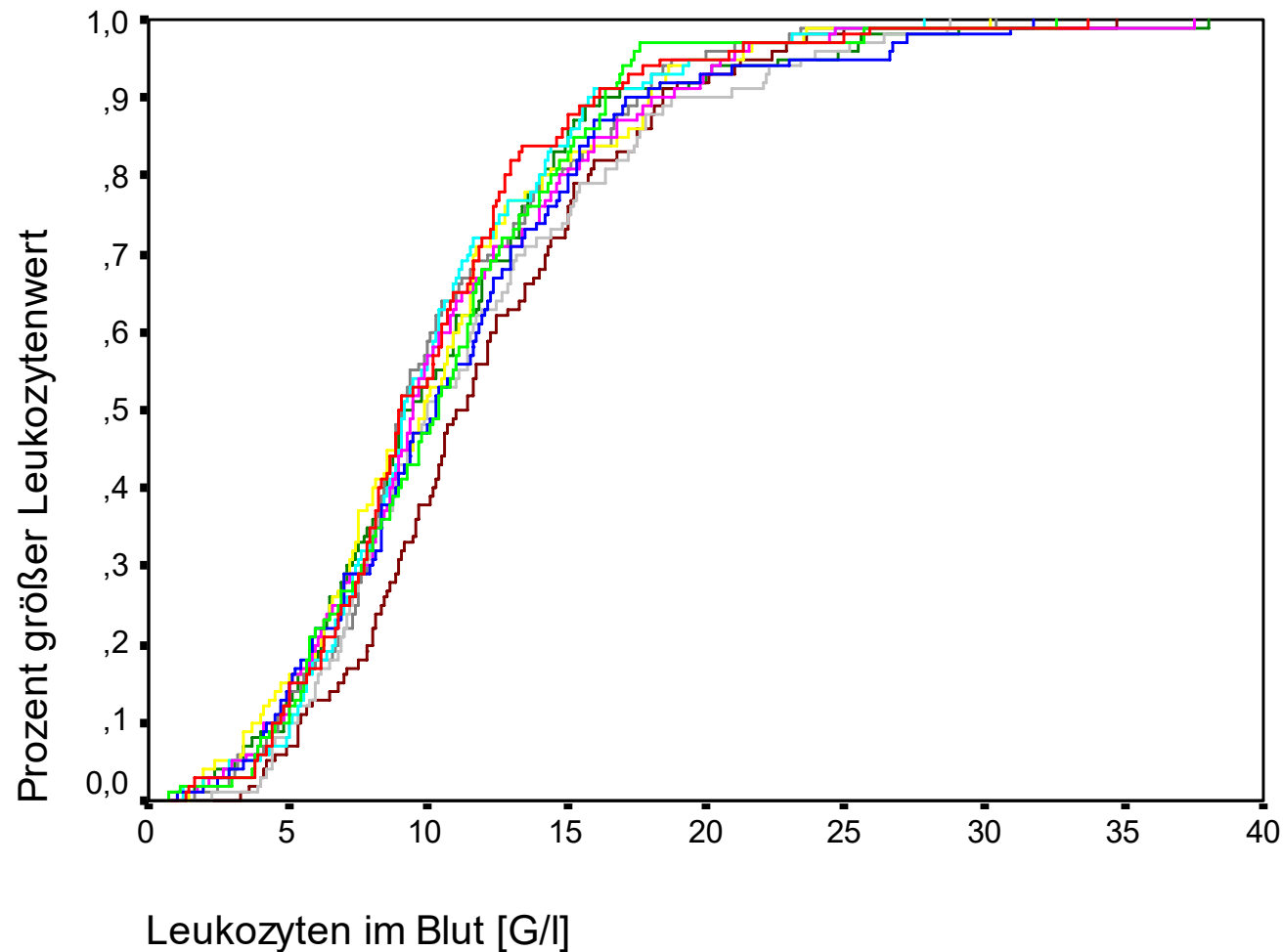
Schätzung einer Verteilung – Effekt der Fallzahl

9 Stichproben von 50 Patienten



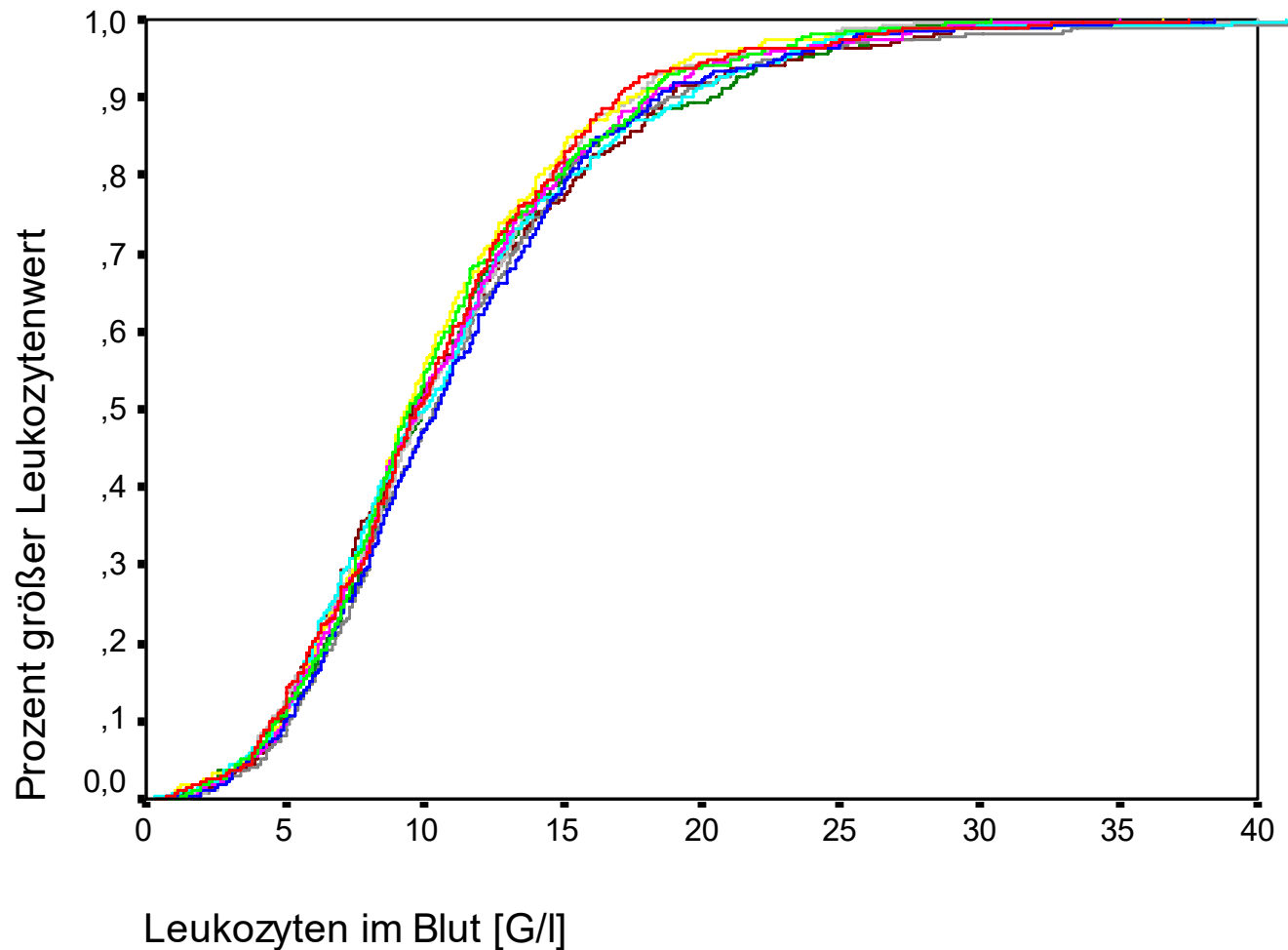
Schätzung einer Verteilung – Effekt der Fallzahl

9 Stichproben mit 100 Patienten



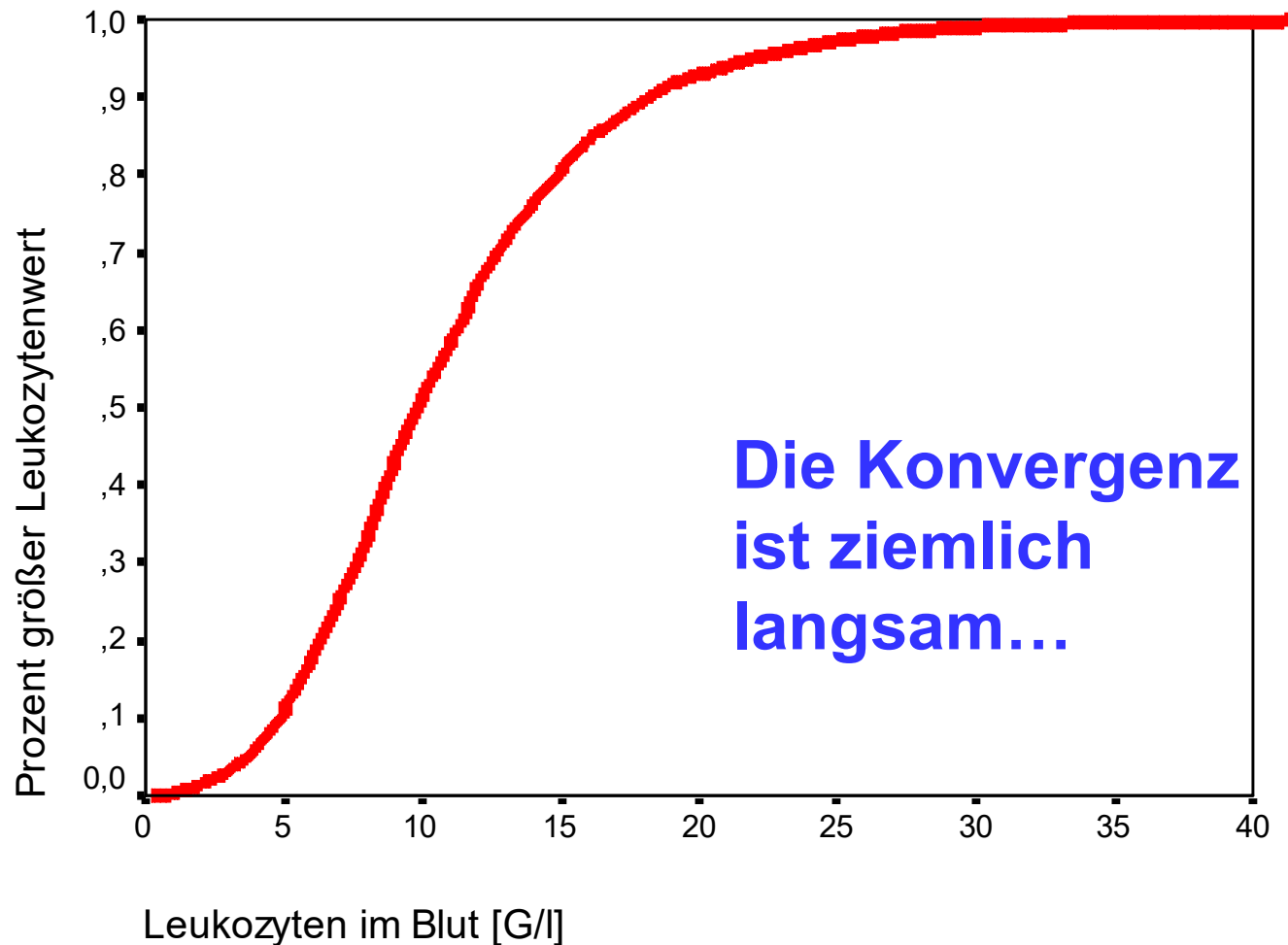
Schätzung einer Verteilung – Effekt der Fallzahl

9 Stichproben mit 400 Patienten



Schätzung einer Verteilung – Effekt der Fallzahl

Schätzung mit N=4725 Patienten



TSCHEBYSCHEFFsche Ungleichung

Satz: Sei X eine reelle Zufallsvariable.
 $E(X)$ und $\text{var}(X)$ existiere.

$$\text{pr}(|X - E(X)| > \lambda \sqrt{\text{var}(X)}) < \frac{1}{\lambda^2} \quad (\forall \lambda > 0)$$

Allgemeines (schwaches) Gesetz der großen Zahl

Sei X eine reelle Zufallsvariable mit Verteilungsfunktion F_X
und $E(X) < \infty$ und $\text{var}(X) < \infty$.

Seien $X_1, \dots, X_n$ unabhängig und identisch $\sim F_X$.

Betrachte den arithmetischen Mittelwert als neue Zufallsgröße
(Stichprobenfunktion).

$$\bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i. \quad \text{Es gilt:}$$

$$\text{a) } E(\bar{X}_n) = \frac{1}{n} \sum_{i=1}^n E(X_i) = E(X)$$

$$\text{b) } \text{var}(\bar{X}_n) = \frac{1}{n^2} \sum_{i=1}^n \text{var}(X_i) = \frac{\text{var}(X)}{n}$$

$$\text{c) } \forall \varepsilon > 0: \quad \lim_{n \rightarrow \infty} \text{pr}(|\bar{X}_n - E(X)| > \varepsilon) = 0$$

(Beweis mit Tschebyscheffscher Ungleichung. Setze $\lambda = \frac{\varepsilon \sqrt{n}}{\sqrt{\text{var}(X)}}$.)

NB: Es gilt sogar noch stärker:

$$\text{pr}\left(\lim_{n \rightarrow \infty} |\bar{X}_n - E(X)| = 0\right) = 1$$

Momentenschätzer

Mit X ist natürlich auch X^r eine reelle Zufallsvariable. Falls $E(X^r)$ und $\text{var}(X^r)$ existieren, so kann man $E(X^r)$ durch das entsprechende arithmetische Mittel

$$\frac{1}{n} \sum_{i=1}^n (x_i)^r \text{ schätzen.}$$

Insbesondere die Varianz durch $\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2$.

Zusätzliche Ergänzungen

Verteilung einer Funktion $g(X,Y)$ I

Notiz:

Seien X, Y Zufallsvariable mit gemeinsamer Dichte $f_{X,Y}(x,y) dx dy$.

Sei $Z := g(X,Y)$.

Die Gleichung $z = g(x,y)$ lasse sich eindeutig nach y auflösen:

$y = \gamma(x,z)$, mit γ stetig differenzierbar.

$$\text{Dann gilt: } f_Z(z) dz = \left(\int_{\text{Träger}(X)} f_{X,Y}(x, \gamma(x,z)) \left| \frac{\partial \gamma}{\partial z} \right| dx \right) dz$$

Beweis:

Führe neue Koordinaten (x,z) ein.

Nach Transformationsregel für mehrdimensionale Integrale:

$$f_{X,Y}(x,y) dx dy = f_{X,Y}(x, \gamma(x,z)) \left| \det \begin{pmatrix} 1 & \frac{\partial \gamma}{\partial x} \\ 0 & \frac{\partial \gamma}{\partial z} \end{pmatrix} \right| dx dz$$

$$\int_{\varphi(U)} f(\mathbf{v}) d\mathbf{v} = \int_U f(\varphi(\mathbf{u})) |\det(D\varphi)(\mathbf{u})| d\mathbf{u}.$$

Dann bestimme Randverteilung von Z .

Anwendungsbeispiele

$$Z = X + Y : \quad f_Z(z) dz = \left(\int_{\text{Träger}(X)} f_{X,Y}(x, z-x) dx \right) dz$$

$$\text{Falls } X, Y \text{ unabhängig: } f_Z(z) dz = \left(\int_{\text{Träger}(X)} f_X(x) \cdot f_Y(z-x) dx \right) dz \text{ (Faltung)}$$

$$Z = X \cdot Y : \quad f_Z(z) dz = \left(\int_{\text{Träger}(X)} f_{X,Y}\left(x, \frac{z}{x}\right) \frac{1}{|x|} dx \right) dz \quad (X \neq 0)$$