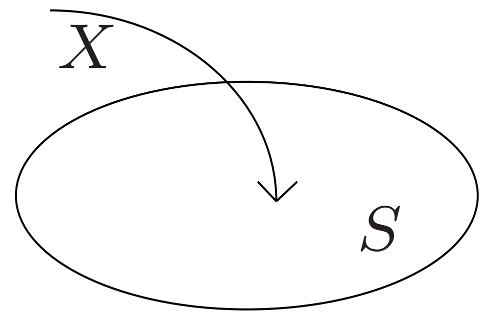
Vorlesung 12a

Schätzen von Parametern

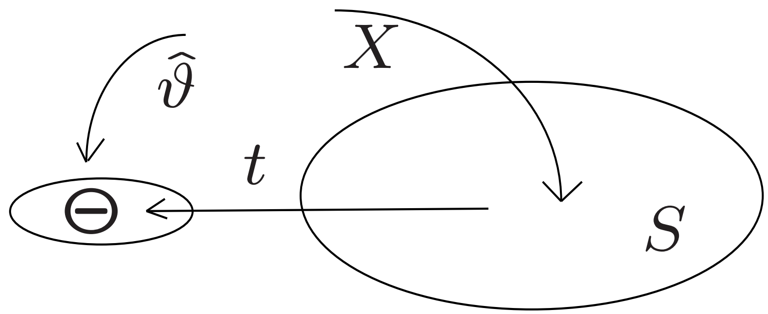
Teil 2

Unser Logo der ersten Stunde:



$$\mathbf{P}\ (X \in da) = \rho\ (da)$$

Ein Logo der Statistik:



$$\mathbf{P}_\vartheta(X \in da) = \rho_\vartheta(da)\,, \quad \vartheta \in \Theta$$

Θ ... *Parameterraum*

S ... *Beobachtungsraum*

$\hat{\vartheta} := t(X)$ ... *Schätzer* für den Parameter ϑ

Ein tragfähiger Vorschlag für die Schätzung von ϑ:

Für jedes $a \in S$ sei $t(a)$ dasjenige (oder eines von den) ϑ,

für das die W'keit, den Ausgang a zu erhalten, maximal wird.

Für diskretes X heißt dies:

$$\mathbf{P}_{t(a)}(X = a) = \max_{\vartheta \in \Theta} \mathbf{P}_\vartheta(X = a) \; .$$

M. a. W.:

$t(a)$ ist Maximalstelle von $\vartheta \mapsto \rho_\vartheta(a)$.

Die Zufallsvariable $\widehat{\vartheta} := t(X)$ nennt man dann

Maximum-Likelihood-Schätzer

für den Parameter ϑ *auf der Basis von* X.

Im Fall von Dichten:

$$\rho_\vartheta(da) = f_\vartheta(a)\,da, \quad \vartheta \in \Theta\ ,$$

nimmt man $t(a)$ als Maximalstelle von $\vartheta \mapsto f_\vartheta(a)$

Beispiel 1:

n unabhängige, N(μ, σ^2)-verteilte Zufallsvariable:

$$\vartheta := (\mu, \sigma) \mapsto f_{(\mu,\sigma^2)}(a)$$

$$= (2\pi\sigma^2)^{-n/2} e^{-\left((a_1-\mu)^2+\cdots+(a_n-\mu)^2\right)/\sigma^2}$$

hat als Minimalstelle das Zahlenpaar mit den Komponenten

$$\bar{a} := \frac{1}{n}(a_1 + \cdots + a_n),$$

$$\sqrt{\frac{1}{n}\left((a_1 - \bar{a})^2 + \cdots + (a_n - \bar{a})^2\right)}.$$

Fazit:

Seien $X_1, \ldots, X_n$ unabhängig und N(μ, σ^2)-verteilt.

Der ML-Schätzer für $\vartheta = (\mu, \sigma)$ ist dann $(\widehat{\mu}, \widehat{\sigma})$ mit

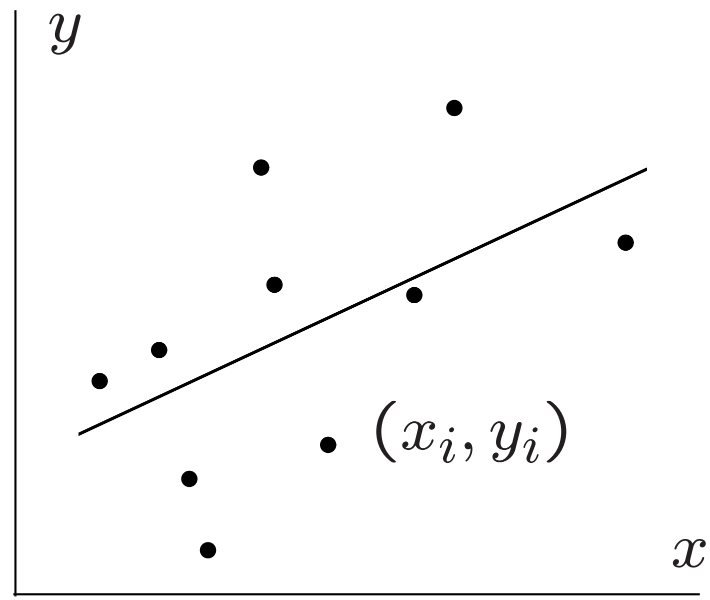
$$\widehat{\mu} := \frac{1}{n}(X_1 + \cdots + X_n),$$

$$\widehat{\sigma}^2 := \frac{1}{n}\Big((X_1 - \widehat{\mu})^2 + \cdots + (X_n - \widehat{\mu})^2\Big)\,.$$

(vgl. Buch Seite 124)

Beispiel 2:

Die einfache lineare Regression



$x_1, \ldots, x_n$ sind feste reelle Zahlen, und das Modell ist

$$Y_i = \beta_0 + \beta_1 x_i + \sigma Z_i \ , \quad i = 1, \ldots, n \ ,$$

mit $\beta_0, \beta_1 \in \mathbb{R},\ \sigma > 0$.

Die Z_i sind gedacht als unabhängige,

standard-normalverteilte ZV'e.

Das Modell lässt sich in vektorieller Form schreiben:

Sei $e := (1, \ldots, 1)$ und $x := (x_1, \ldots, x_n)$

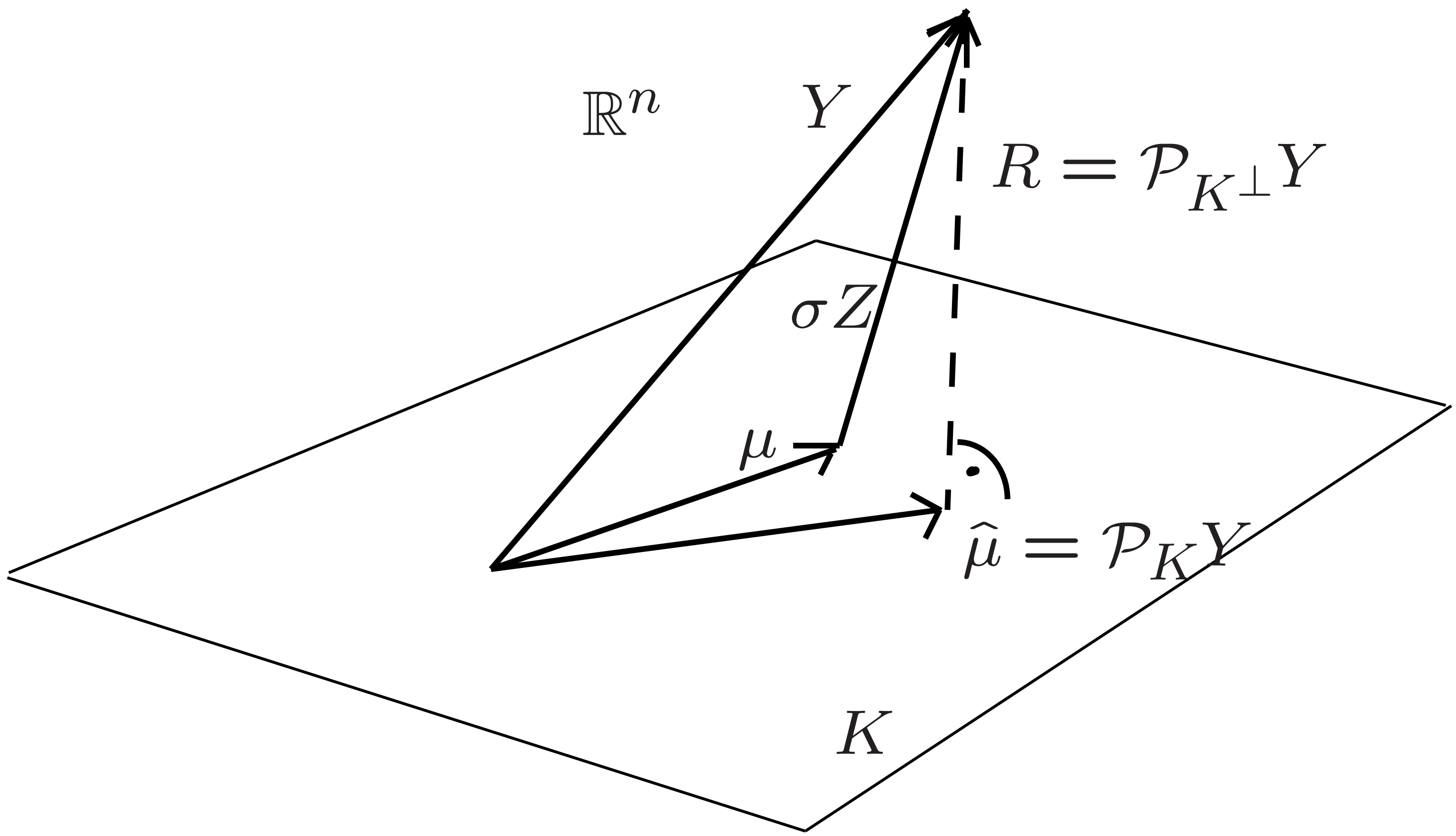
$$Y = \beta_0 e + \beta_1 x + \sigma Z$$

= systematische Komponente + Rauschen

= $\mu + \sigma Z$

mit $\mu \in K :=$ der von e und x aufgespannte Teilraum von $\mathbb{R}^n$

(Dies ist ein Beispiel eines *normalen Linearen Modells.*)



Als Maximum-Likelihood-Schätzer für (μ, σ^2) ergibt sich

$$\hat{\mu} := \mathcal{P}_K Y \ , \quad \hat{\sigma}^2 := |\mathcal{P}_{K^\perp} Y|^2/n \ .$$

Berechnung der Koeffizienten $\widehat{\beta}_0$ und $\widehat{\beta}_1$ in der Projektion

$$\mathcal{P}_K Y = \widehat{\beta}_0 e + \widehat{\beta}_1 x :$$

Wir verwenden als Orthonormalbasis von K:

$$b := \frac{1}{\sqrt{n}} e, \quad b' := \frac{x - \bar{x}e}{|x - \bar{x}e|}.$$

$$\mathcal{P}_K Y = \langle Y, b\rangle b + \langle Y, b'\rangle b' \; = \bar{Y} e + \frac{\langle Y - \bar{Y} e, x - \bar{x} e\rangle}{|x - \bar{x} e|^2} (x - \bar{x} e).$$

Berechnung der Koeffizienten $\widehat{\beta}_0$ und $\widehat{\beta}_1$ in der Projektion

$$\mathcal{P}_K Y = \widehat{\beta}_0 e + \widehat{\beta}_1 x :$$

Wir verwenden als Orthonormalbasis von K:

$$b := \frac{1}{\sqrt{n}} e, \quad b' := \frac{x - \bar{x}e}{|x - \bar{x}e|}.$$

$$\mathcal{P}_K Y = \langle Y, b\rangle b + \langle Y, b'\rangle b' = \bar{Y} e + \frac{\langle Y - \bar{Y}e, x - \bar{x}e\rangle}{|x - \bar{x}e|^2}(x - \bar{x}e).$$

Also ergeben sich die x- und e-Koordinaten von $\mathcal{P}_K Y$ als:

$$\widehat{\beta}_1 = \frac{\langle Y - \bar{Y}e, x - \bar{x}e\rangle}{|x - \bar{x}e|^2} = \frac{\sum_{i=1}^n (Y_i - \bar{Y})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2},$$

$$\widehat{\beta}_0 = \bar{Y} - \widehat{\beta}_1 \bar{x} .$$

Berechnung der Koeffizienten $\widehat{\beta}_0$ und $\widehat{\beta}_1$ in der Projektion

$$\mathcal{P}_K Y = \widehat{\beta}_0 e + \widehat{\beta}_1 x:$$

Wir verwenden als Orthonormalbasis von K:

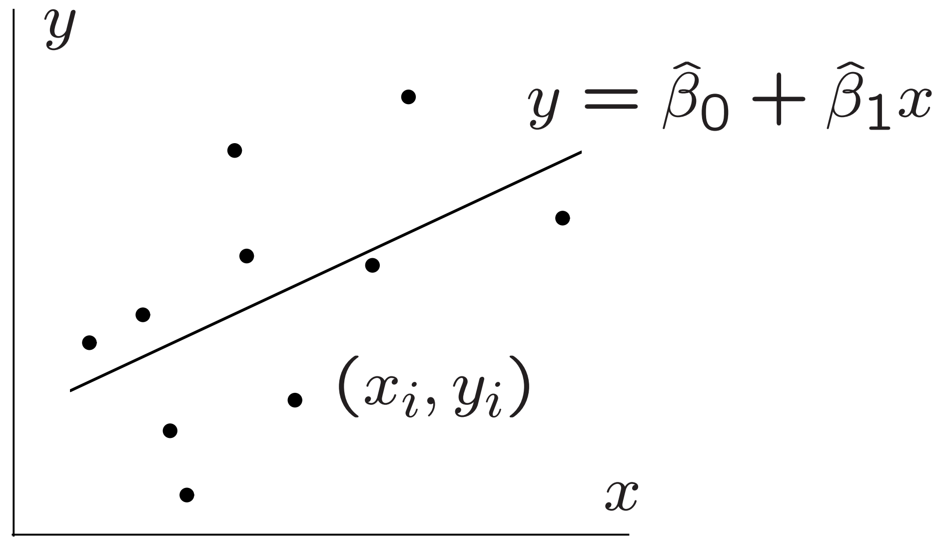
$$b := \frac{1}{\sqrt{n}} e, \quad b' := \frac{x - \bar{x}e}{|x - \bar{x}e|}.$$

$$\mathcal{P}_K Y = \langle Y, b\rangle b + \langle Y, b'\rangle b' = \bar{Y} e + \frac{\langle Y - \bar{Y}e, x - \bar{x}e\rangle}{|x - \bar{x}e|^2}(x - \bar{x}e).$$

Also ergeben sich die x- und e-Koordinaten von $\mathcal{P}_K Y$ als:

$$\widehat{\beta}_1 = \frac{\langle Y - \bar{Y}e, x - \bar{x}e\rangle}{|x - \bar{x}e|^2} = \frac{\sum_{i=1}^n (Y_i - \bar{Y})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2},$$

$$\widehat{\beta}_0 = \bar{Y} - \widehat{\beta}_1 \bar{x}.$$



Die Gerade $x \mapsto y = \widehat{\beta}_0 + \widehat{\beta}_1 x$ heißt *Regressionsgerade*.

Ihr Anstieg ist der *Regressionskoeffizient* $\widehat{\beta}_1$.

Sie geht durch den zentralen Punkt $(\bar{x}, \bar{Y})$.

Man beachte die Analogie zum Beispiel am Ende von Vorlesung 7a.

Maximum-Likelihood-Schätzer am Beispiel des Münzwurfs:

$(X_1, \ldots, X_n)$ sei p-Münzwurf mit unbekanntem p

Beobachtet wird die Realisierung $(a_1, \ldots, a_n)$

mit $k = a_1 + \ldots + a_n$.

Unter allen p ist k/n derjenige Parameter, mit dem

$$\mathbf{P}_p(X_1 = a_1, \ldots, X_n = a_n) = p^k (1-p)^{n-k}$$

maximal ist.

Man sagt:

$$\widehat{p} := \frac{1}{n}(X_1 + \cdots + X_n)$$

ist der Maximum-Likelihood-Schätzer für p.

Für $k = 0$ (kein Erfolg in n Versuchen) ergibt sich 0 als Maximum-Likelikood-Schätzung von p.

Das ist möglicherweise zu pessimistisch.

Eine Alternative bietet der sogenannte *Bayes-Schätzer* (vgl Buch S. 127).

Hier denkt man an ein zweistufiges Experiment:

1. eine auf $[0, 1]$ uniform verteilte Zufallsvariable U
2. gegeben $\{U = u\}$ einen Münzwurf mit Erfolgswahrscheinlichkeit U.

$$\tilde{p} := \mathbf{E}[U|K_n] \; .$$

Erinnerung an Vorlesung 9b:

$Z_1, Z_2, \ldots$ sei ein Münzwurf mit uniform auf $[0,1]$ verteiltem zufälligem Erfolgsparameter U,

K_n sei die Anzahl der Erfolge in den ersten n Versuchen.

Gefragt ist nach $\mathbf{E}[U \mid K_n = k]$.

Wir wissen schon:

Die bedingte Dichte von U gegeben $\{K_n = k\}$ ist

$$\mathbf{P}_k(U \in du) = \frac{1}{\frac{1}{n+1}} \binom{n}{k} u^k (1-u)^{n-k} \, du$$

$$\mathbf{E}[U \mid K_n = k] = \int \frac{1}{\frac{1}{n+1}} \binom{n}{k} u^{k+1} (1-u)^{n-k}\, du$$

$$= \frac{k+1}{n+2} \int \frac{1}{\frac{1}{n+2}} \binom{n+1}{k+1} u^{k+1} (1-u)^{(n+1)-(k+1)}\, du$$

$$= \frac{k+1}{n+2}.$$

Man nennt dies auch den

Bayes-Schätzer für die Erfolgswahrscheinlichkeit

(bei a priori uniform verteilter Erfolgswahrscheinlichkeit).

Man muss hier gar nicht rechnen. Mit der totalen W'keit gilt:

$$\mathbf{E}[U \mid K_n] = \mathbf{E}[\mathbf{P}[Z_{n+1} = 1 \mid U] \mid K_n] = \mathbf{P}[Z_{n+1} = 1 | K_n]$$

Im Buch S. 113/114 liest man nach

(vgl. dazu auch unsere Übungsaufgabe 36):

Ein Münzwurf $(Z_1, Z_2, \ldots)$

mit uniform verteilter Erfolgswahrscheinlichkeit U

ist so verteilt wie die Folge der Zuwächse in Richtung Osten

in einer Nordost-Wanderung à la Pólya. Also:

$$\mathbf{P}[Z_{n+1} = 1 | K_n = k] = \frac{k+1}{n+2}.$$

Die nächsten beiden Folien zeigen einen Vergleich

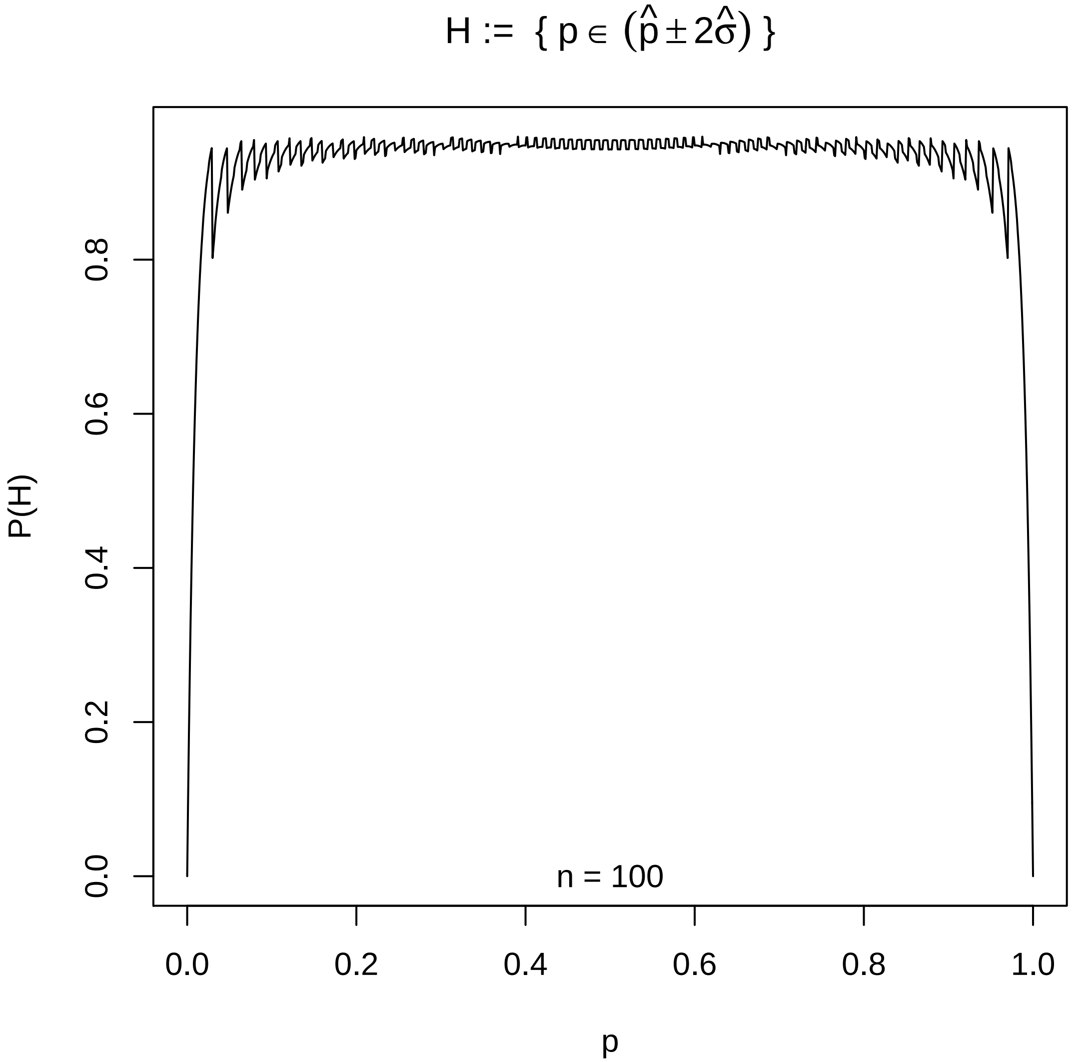
der Überdeckungsw'keiten der beiden Konfidenzintervalle

$$I := \left(\widehat{p} - 2\sqrt{\frac{1}{n}\widehat{p}(1-\widehat{p})},\, \widehat{p} + 2\sqrt{\frac{1}{n}\widehat{p}(1-\widehat{p})}\right) := (\widehat{p} \pm 2\widehat{\sigma})$$

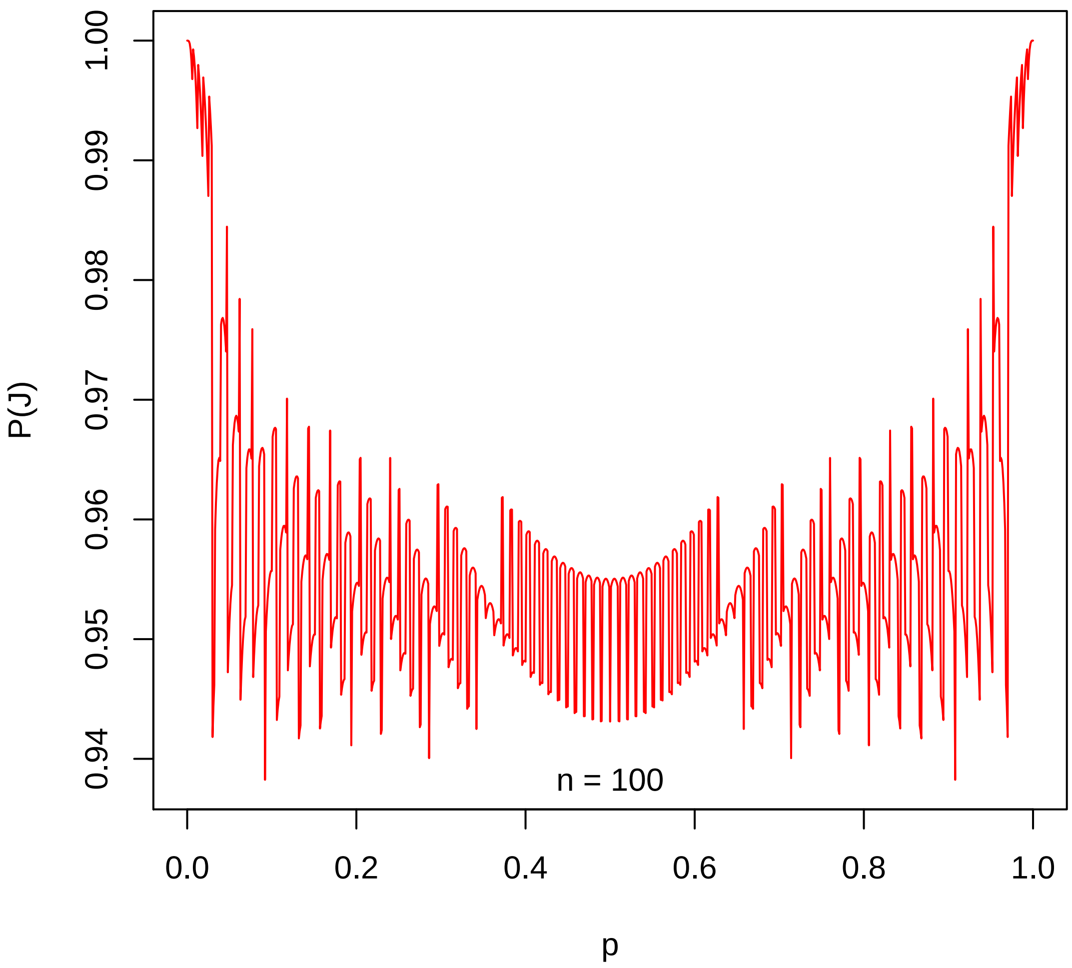
und

$$\tilde{I} := \left(\tilde{p} - 2\sqrt{\frac{1}{n}\tilde{p}(1-\tilde{p})},\, \tilde{p} + 2\sqrt{\frac{1}{n}\tilde{p}(1-\tilde{p})}\right) := (\tilde{p} \pm 2\tilde{\sigma})$$

(siehe dazu Buch Seite 129/130)


$H := \{ p \in (\hat{p} \pm 2\hat{\sigma}) \}$
P(H)
0.0
0.2
0.4
0.6
0.8
n = 100
0.0
0.2
0.4
0.6
0.8
1.0
p

$$J := \{\, p \in (\tilde{p} \pm 2\tilde{\sigma}) \,\}$$

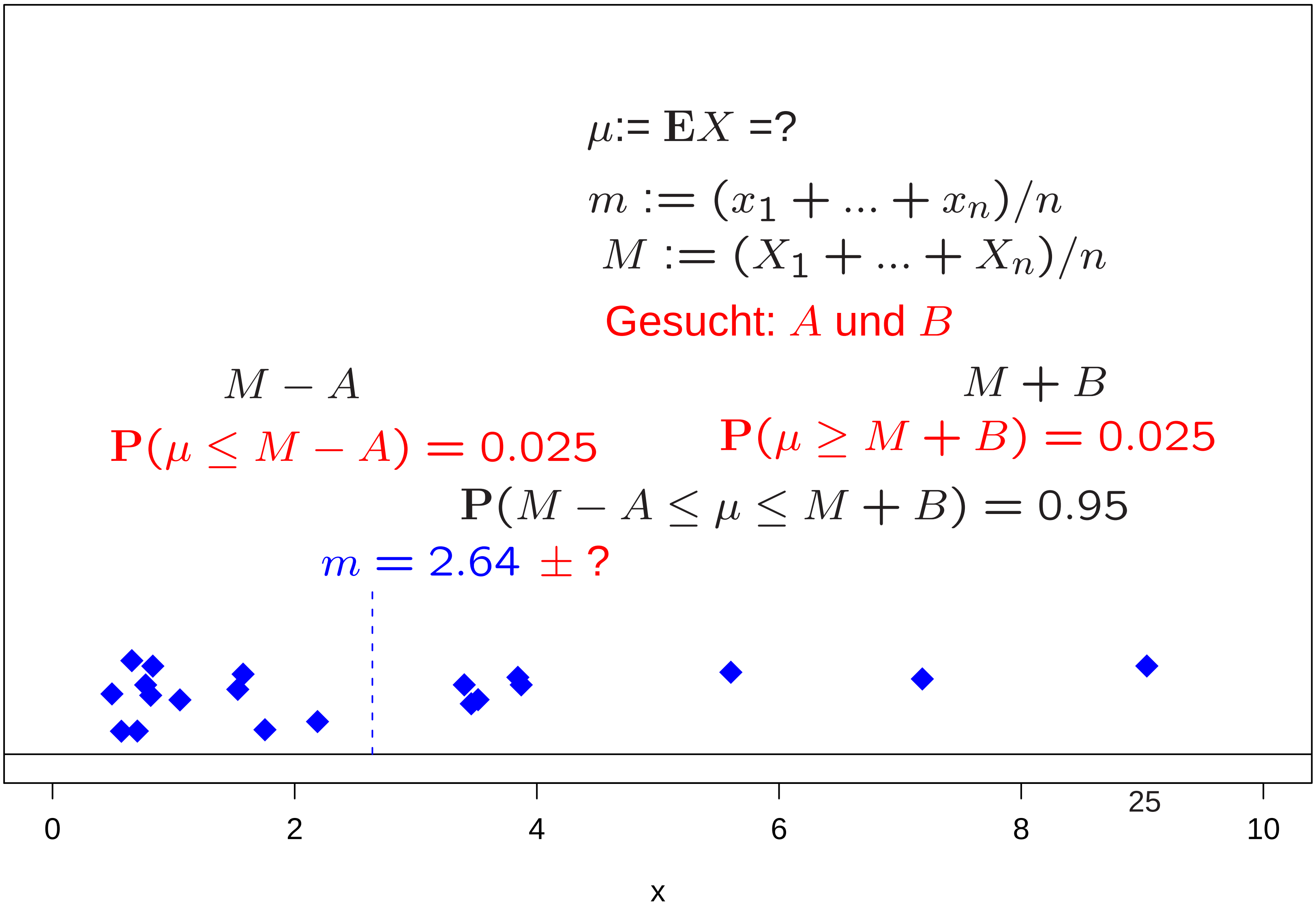


Die Idee des Bootstrap

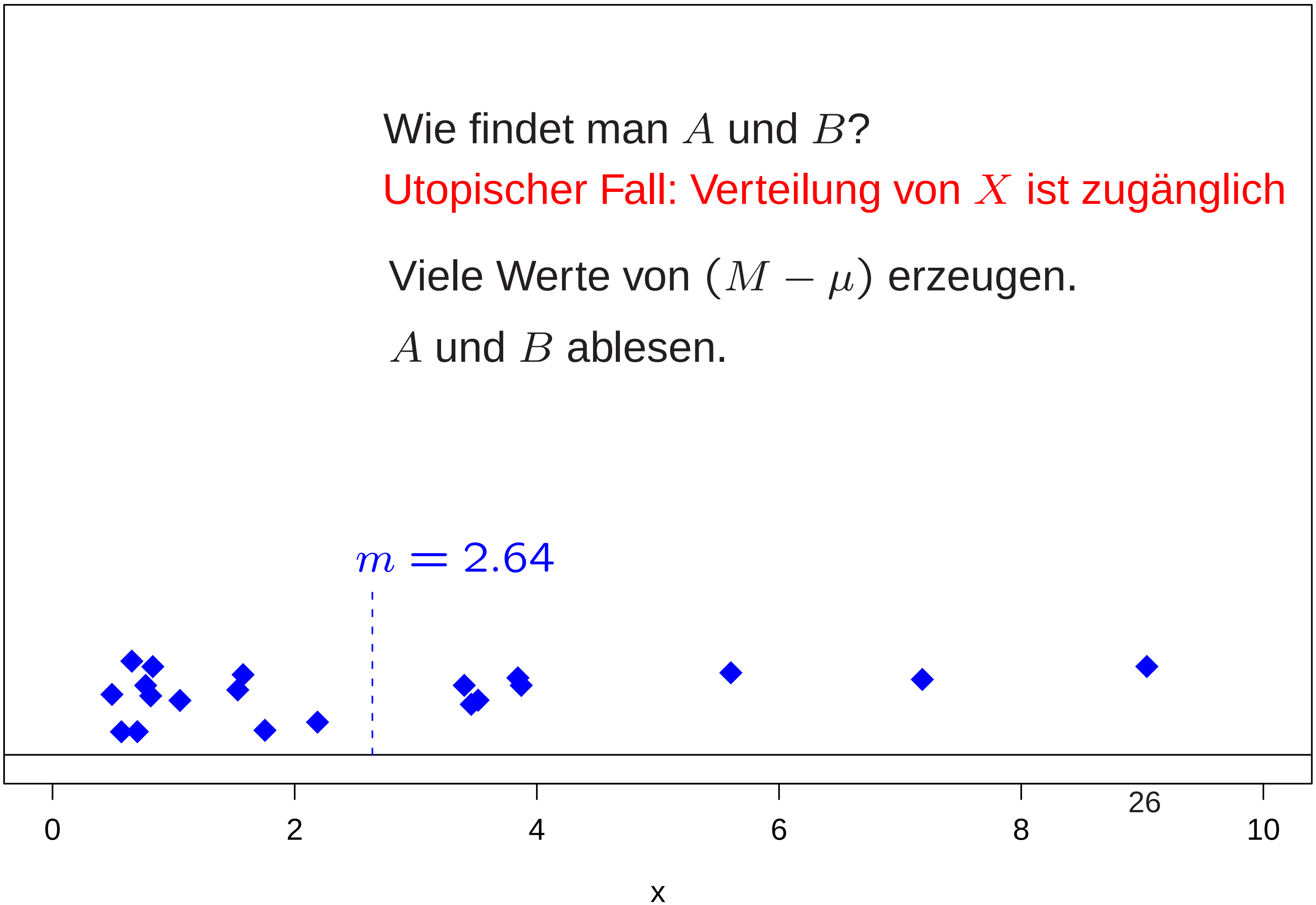
am Beispiel eines Bootstrap-Konfidenzintervalls

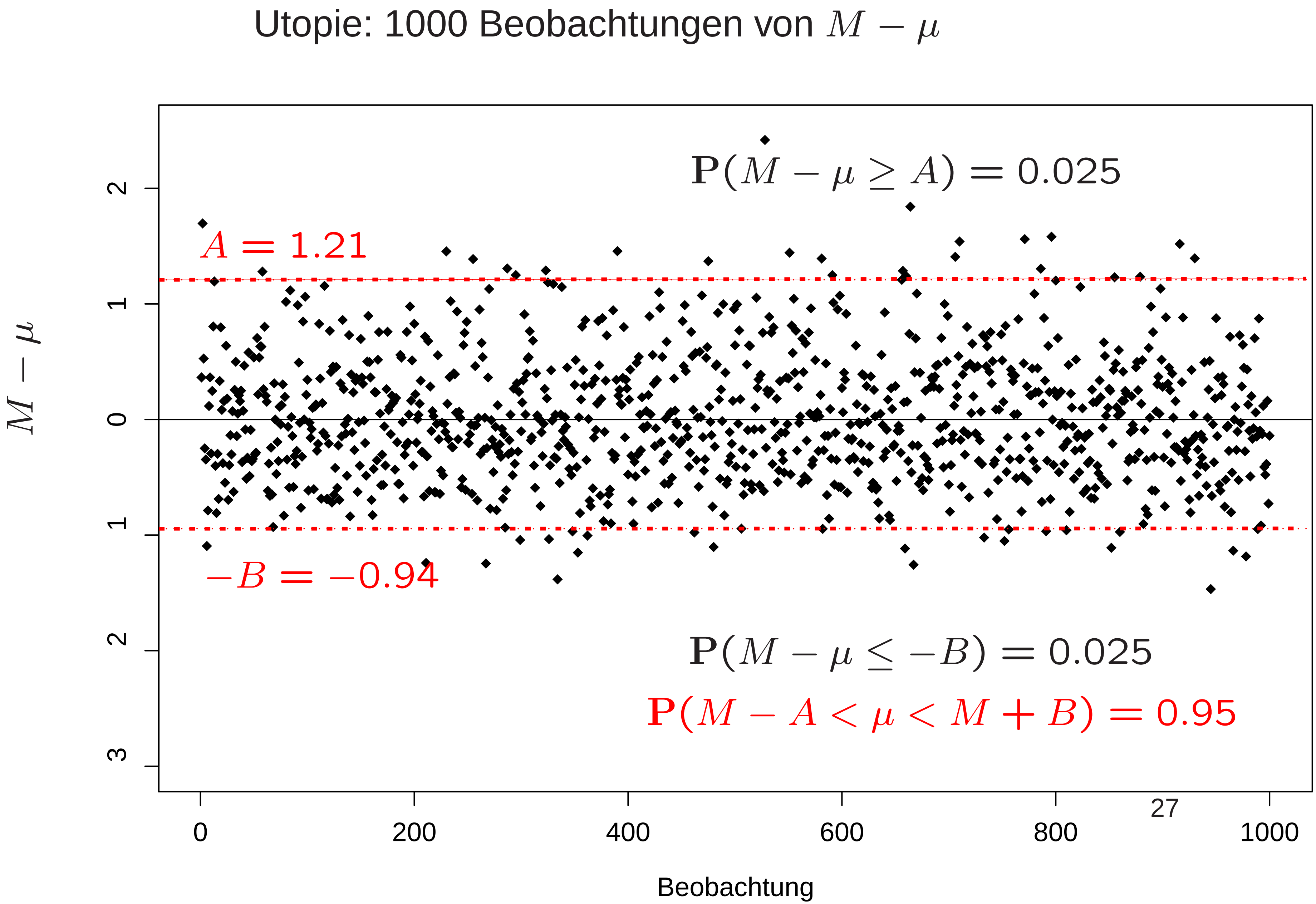
für den Erwartungswert

Beobachtungen $x_1, ..., x_n$ der Zufallsgröße X ($n = 20$)



Beobachtungen $x_1, ..., x_n$ der Zufallsgröße X ($n = 20$)

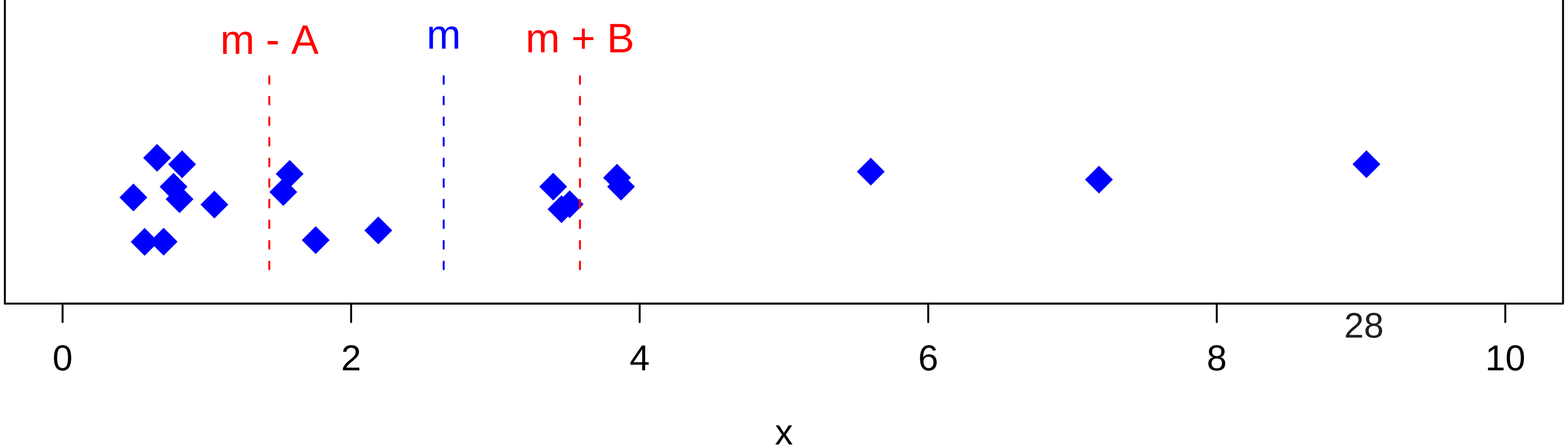



Utopie: 1000 Beobachtungen von $M - \mu$
$\mathbf{P}(M - \mu \geq A) = 0.025$
$A = 1.21$
$-B = -0.94$
$\mathbf{P}(M - \mu \leq -B) = 0.025$
$\mathbf{P}(M - A < \mu < M + B) = 0.95$
$M - \mu$
2
1
0
1
2
3
0
200
400
600
800
1000
Beobachtung

Utopie

ν:= die Verteilung von X

Mit Stichproben aus ν
könnte man A und B schätzen
ν ist unbekannt. Was tun?



Strecken nach der Decke

ν:= die Verteilung von X

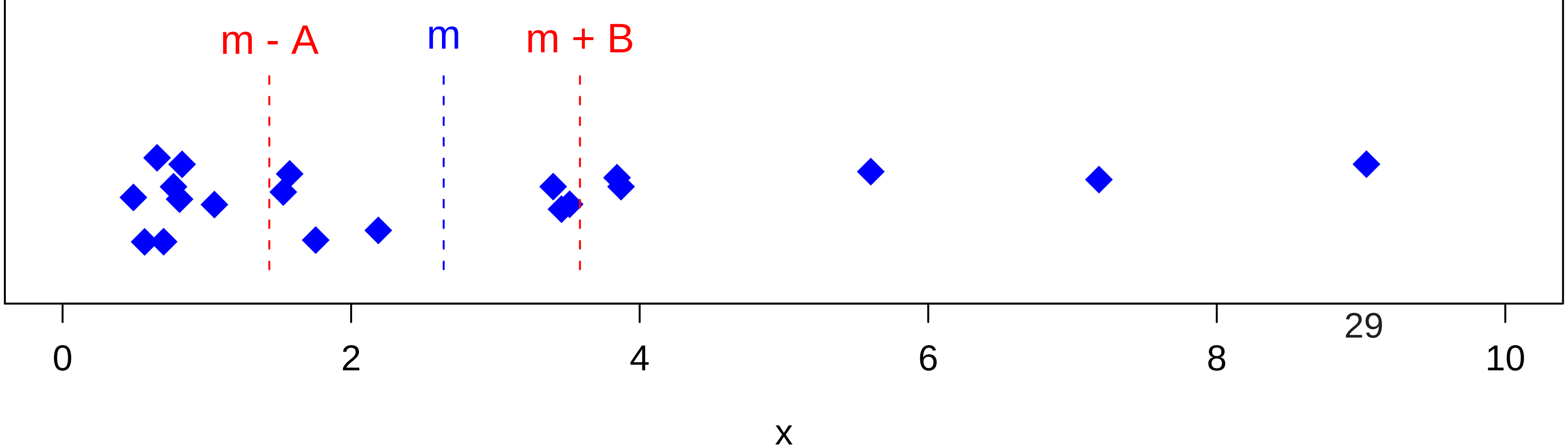
Mit Stichproben aus ν
könnte man A und B schätzen
ν ist unbekannt. Was tun?

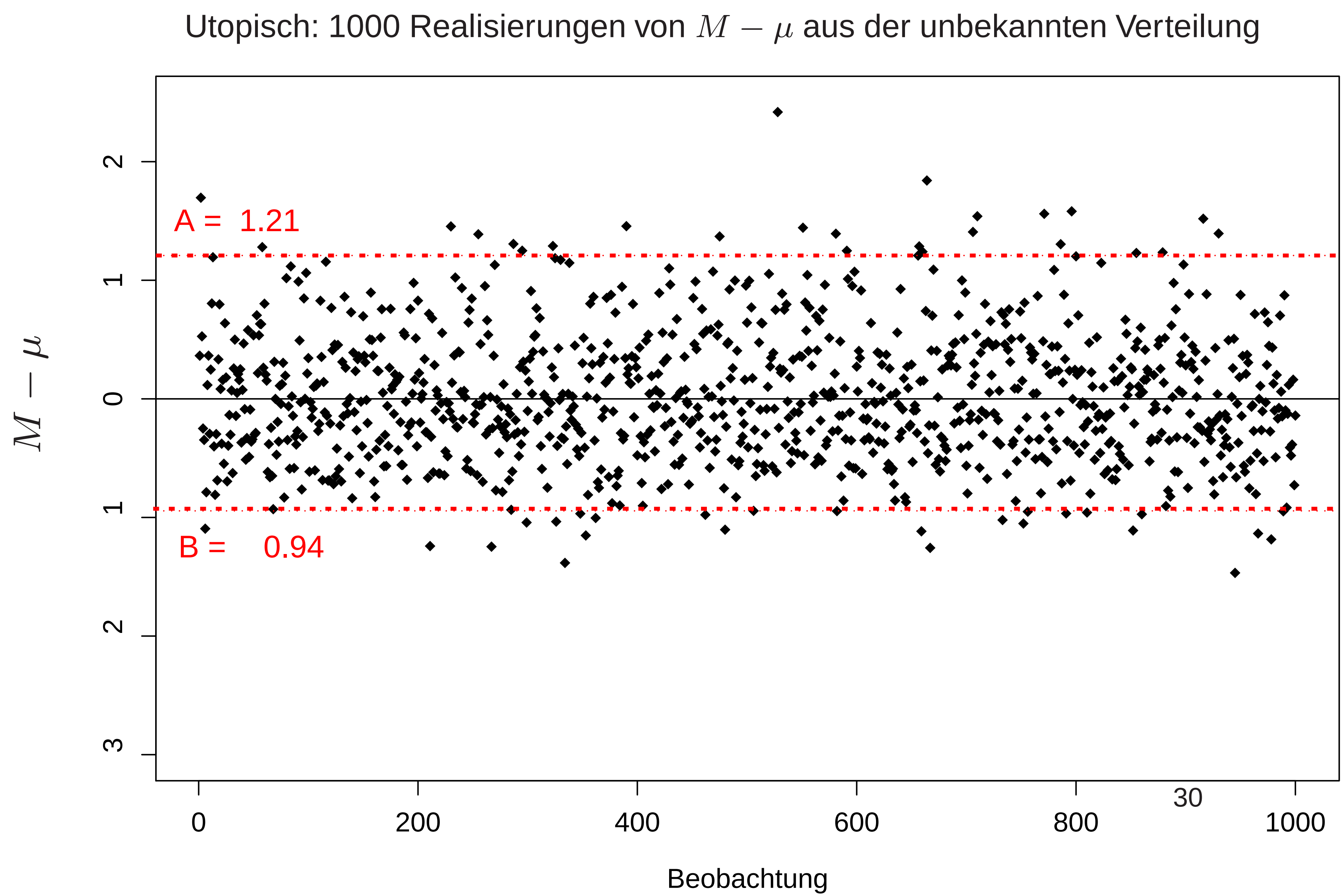
$\nu_{\vec{x}} := \frac{1}{n}(\delta_{x_1} + ... + \delta_{x_n})$
die geschätzte Verteilung von X

Bootstrap-Heuristik

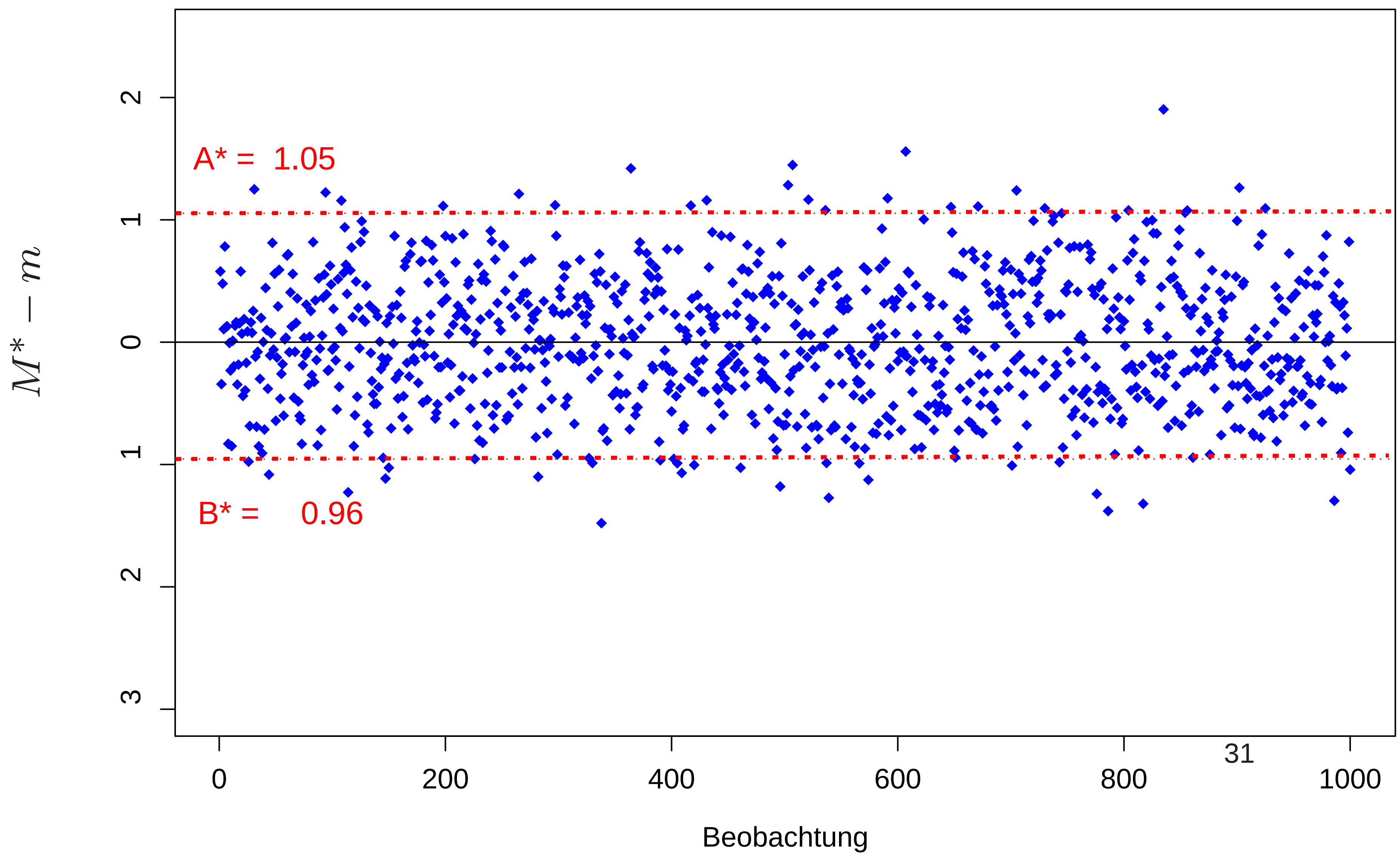
$\nu \sim \nu_{\vec{x}}$

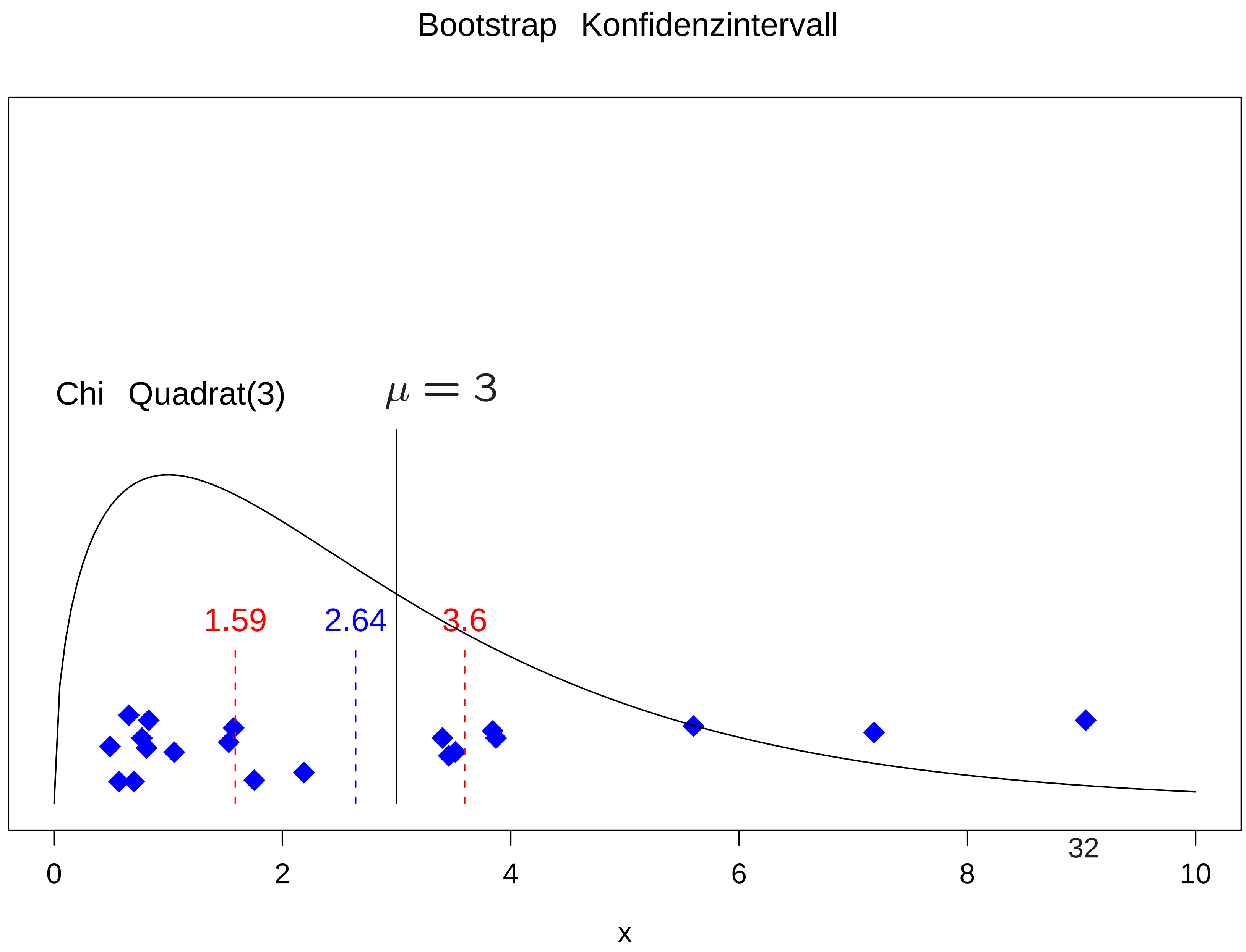
Also: Stichproben
aus $\nu_{\vec{x}}$ liefern (A^*, B^*)
Bootstrap-Konfidenzintervall
$(m - A^*, m + B^*)$




Utopisch: 1000 Realisierungen von $M - \mu$ aus der unbekannten Verteilung
A = 1.21
B = 0.94
$M - \mu$
2
1
0
1
2
3
0
200
400
600
800
1000
Beobachtung

Praktisch: 1000 Realisierungen von $(M^* - m)$ aus Bootstrap-Stichproben





Bootstrap Konfidenzintervall
Chi Quadrat(3)
$\mu = 3$
1.59
2.64
3.6
0
2
4
6
8
10
x

Überdeckungswahrscheinlichkeit

Mit welcher W'keit überdeckt das zufällige Intervall den wahren Parameter?

Hoffentlich mit annähernd 95 %.

Wie können wir das prüfen?

Wir ziehen 1000 Stichproben $\vec{x}_1, ..., \vec{x}_{1000}$

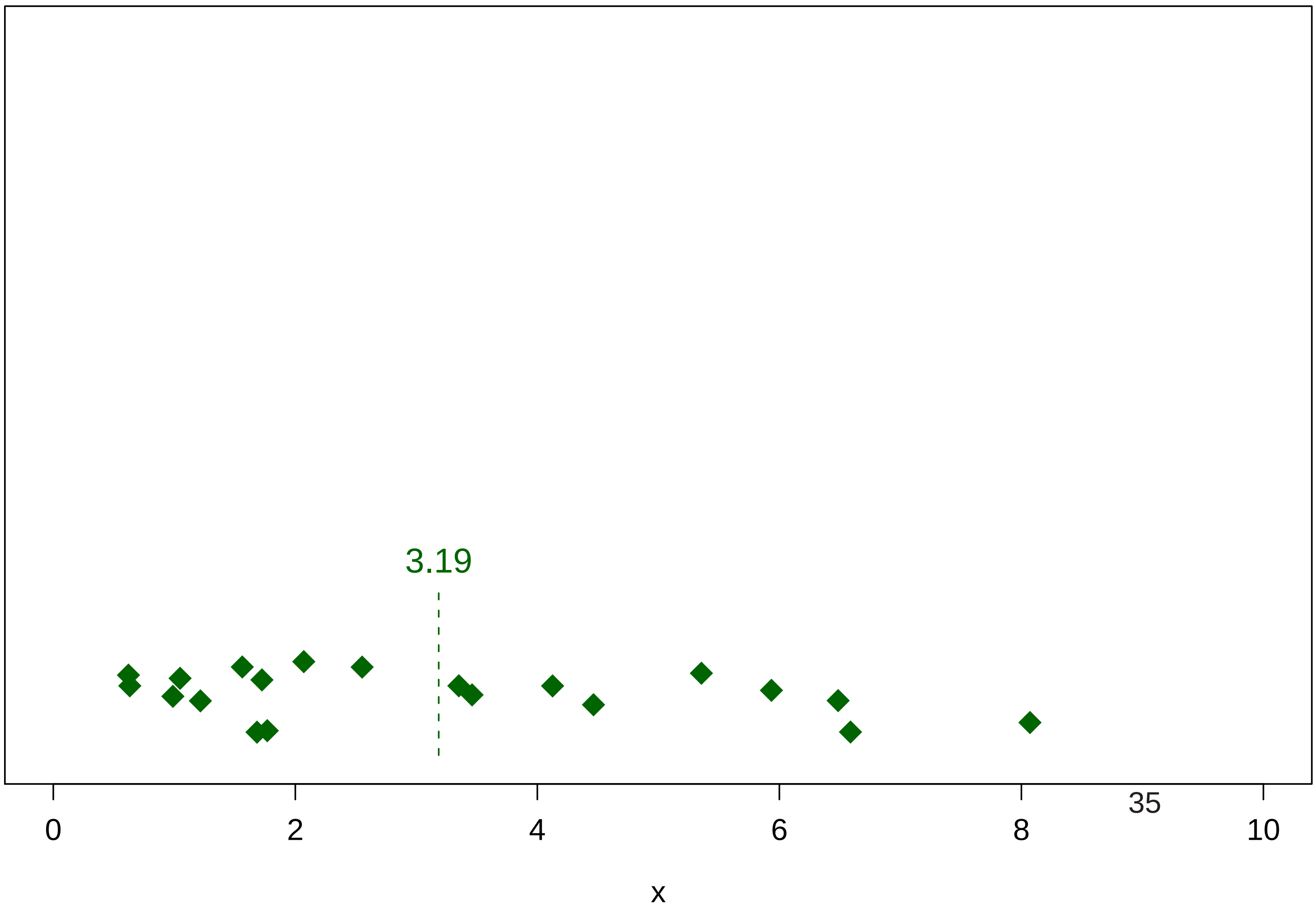
und beschaffen uns für jedes $i = 1, ..., 1000$

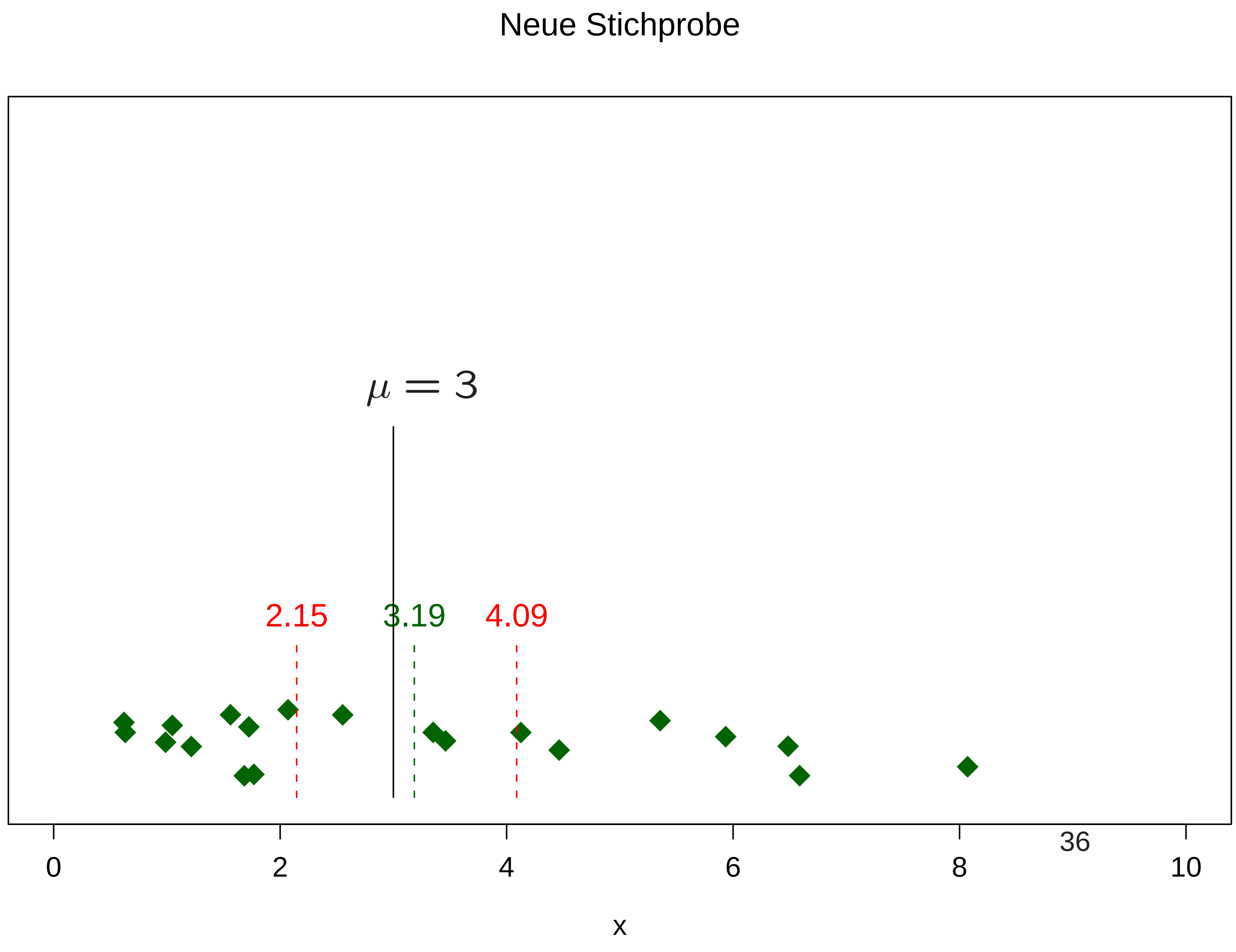
durch 1000 Zwanziger-Züge aus $\nu_{\vec{x}_i}$

eine Realisierung des Konfidenzintervalls

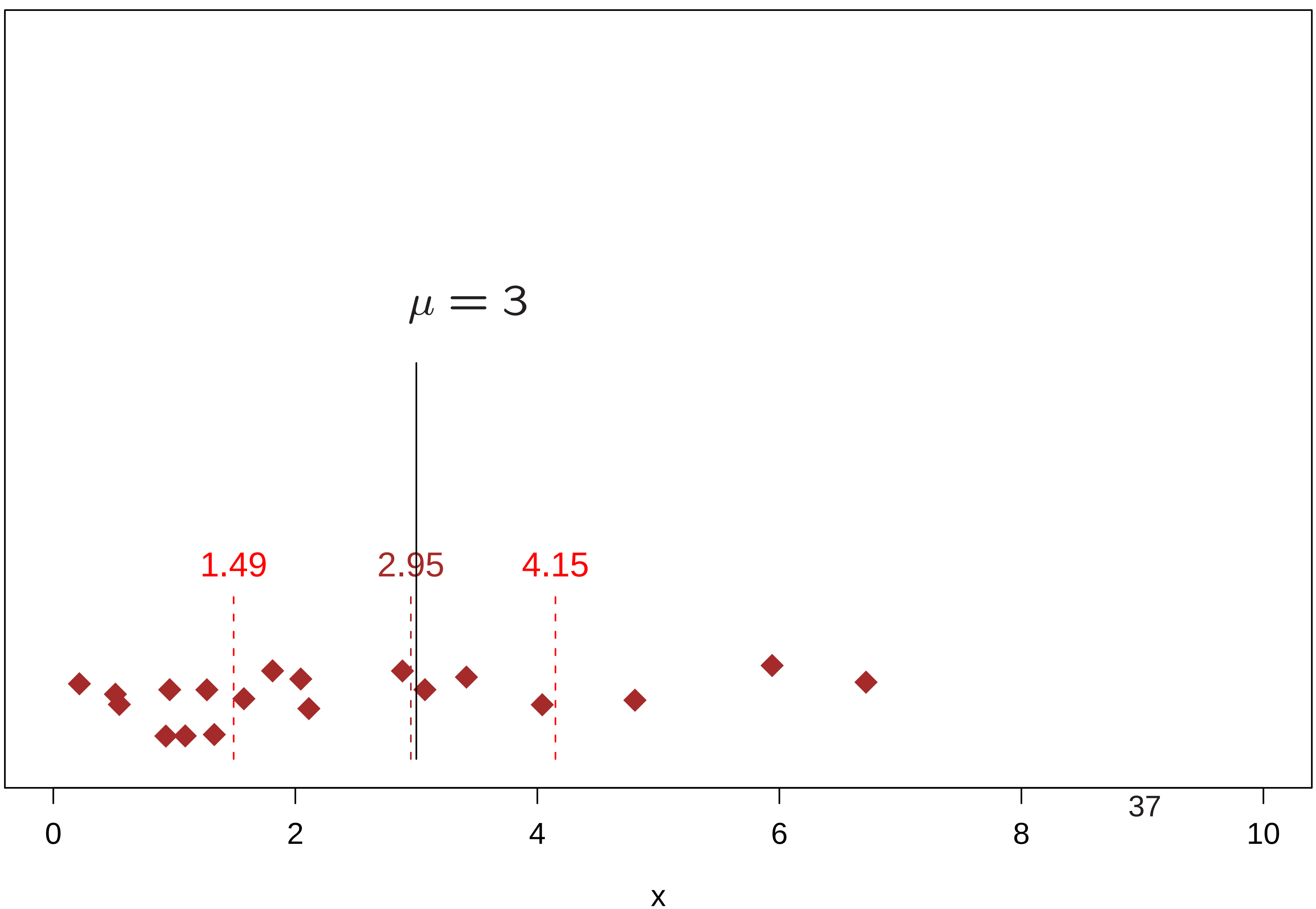
$$[U_i, O_i] := [M_i - A_i^*, M_i + B_i^*]$$

Neue Stichprobe




Neue Stichprobe
$\mu = 3$
2.15
3.19
4.09
0
2
4
6
8
10
x

Neue Stichprobe



Wir ziehen 1000 Stichproben $\vec{x}_1, ..., \vec{x}_{1000}$

und beschaffen uns für jedes $i = 1, ..., 1000$

durch 1000 Zwanziger-Züge aus $\nu_{\vec{x}_i}$

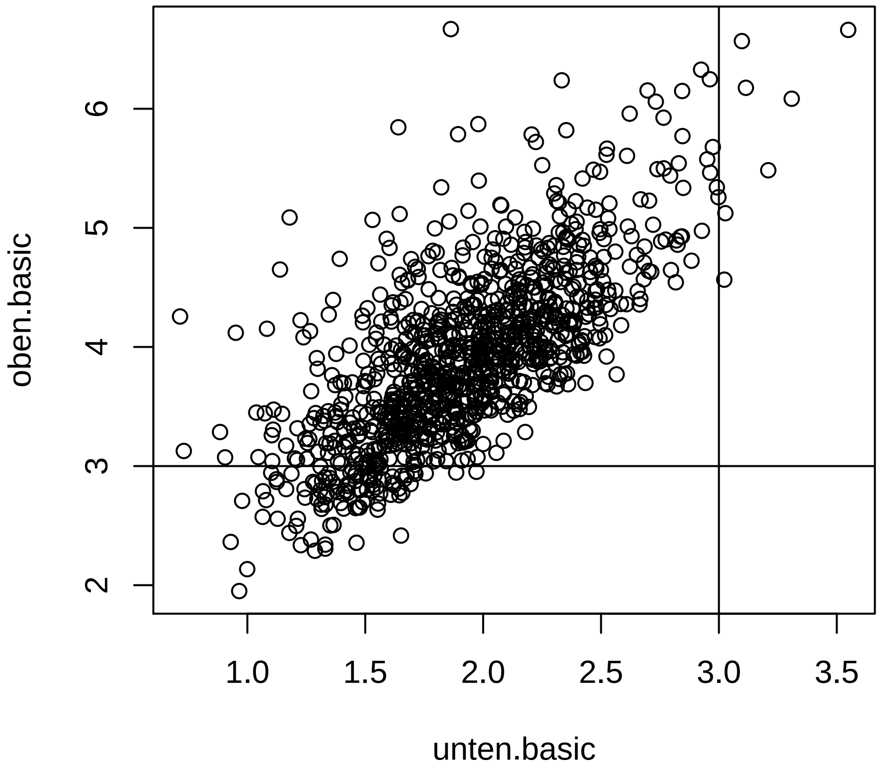
eine Realisierung des Konfidenzintervalls

$$[U_i, O_i] := [M_i - A_i^*, M_i + B_i^*]$$

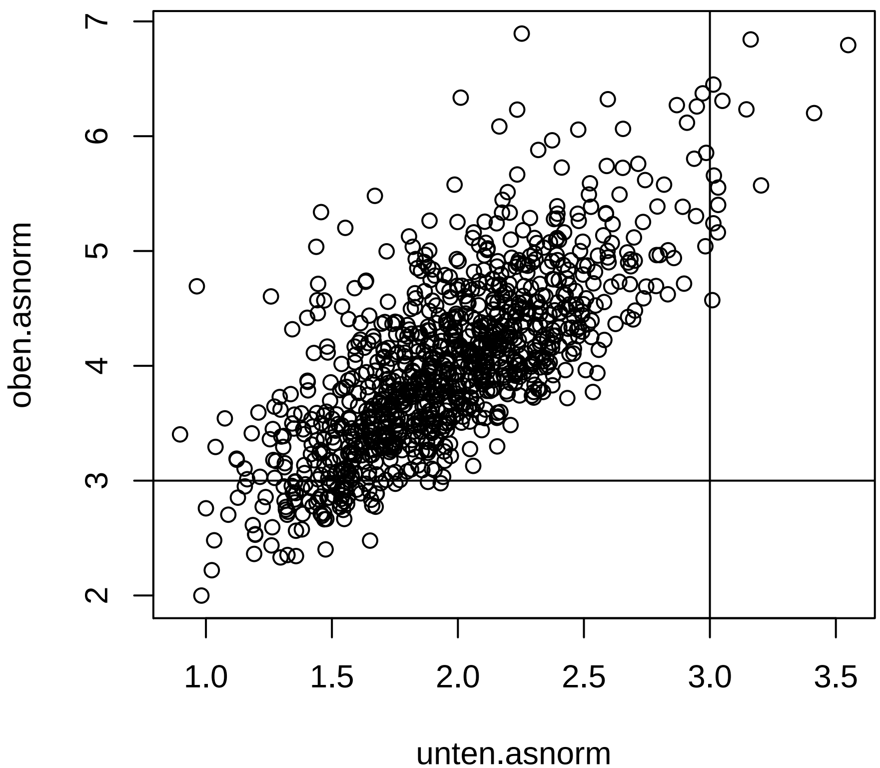
Jedes der Paare (U_i, O_i) wird dargestellt als Punkt im $\mathbb{R}^2$.

Das Intervall $[U_i, O_i]$ überdeckt $\mu = 3 \iff$

der Punkt (U_i, O_i) liegt im Quadranten $\{(x, y)|x \leq 3, y \geq 3\}$



897 von 1000 Realisierungen des *einfachen Bootstrap-Konfidenzintervalls* $(m - A^*, m + B^*)$ haben μ überdeckt



904 von 1000 Realisierungen des *normalen Konfidenzintervalls*
$[M - (1.96/\sqrt{n})s, M + (1.96/\sqrt{n})s]$ haben μ überdeckt