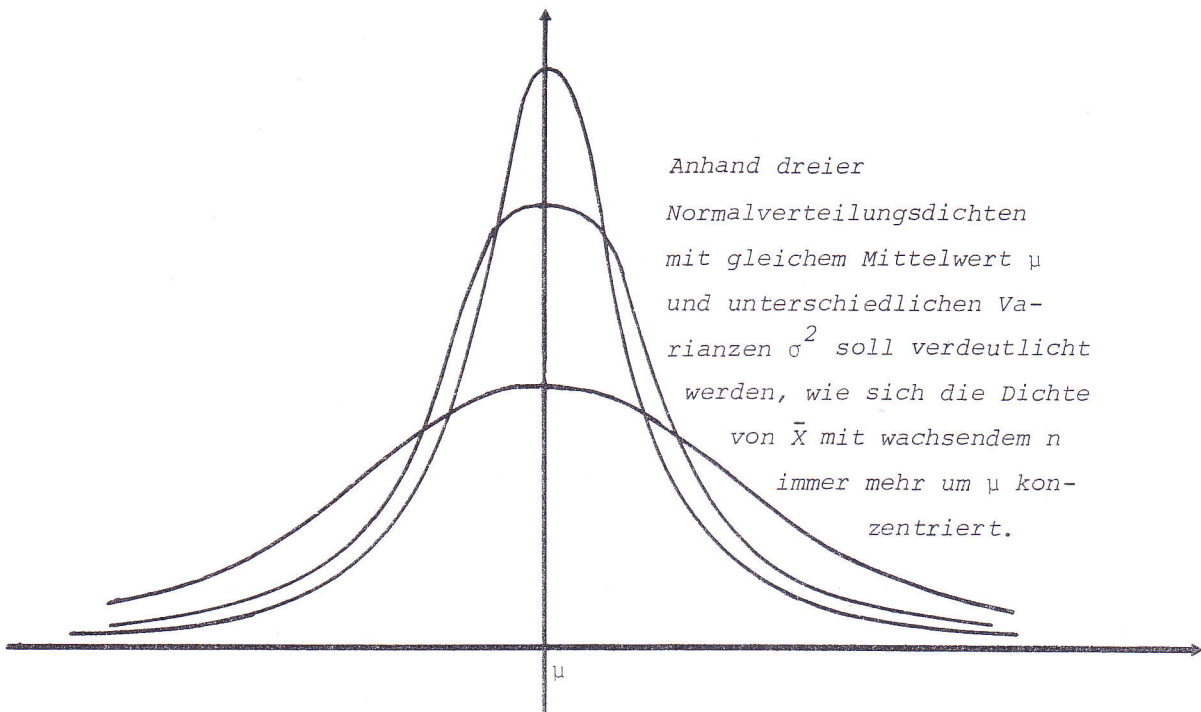
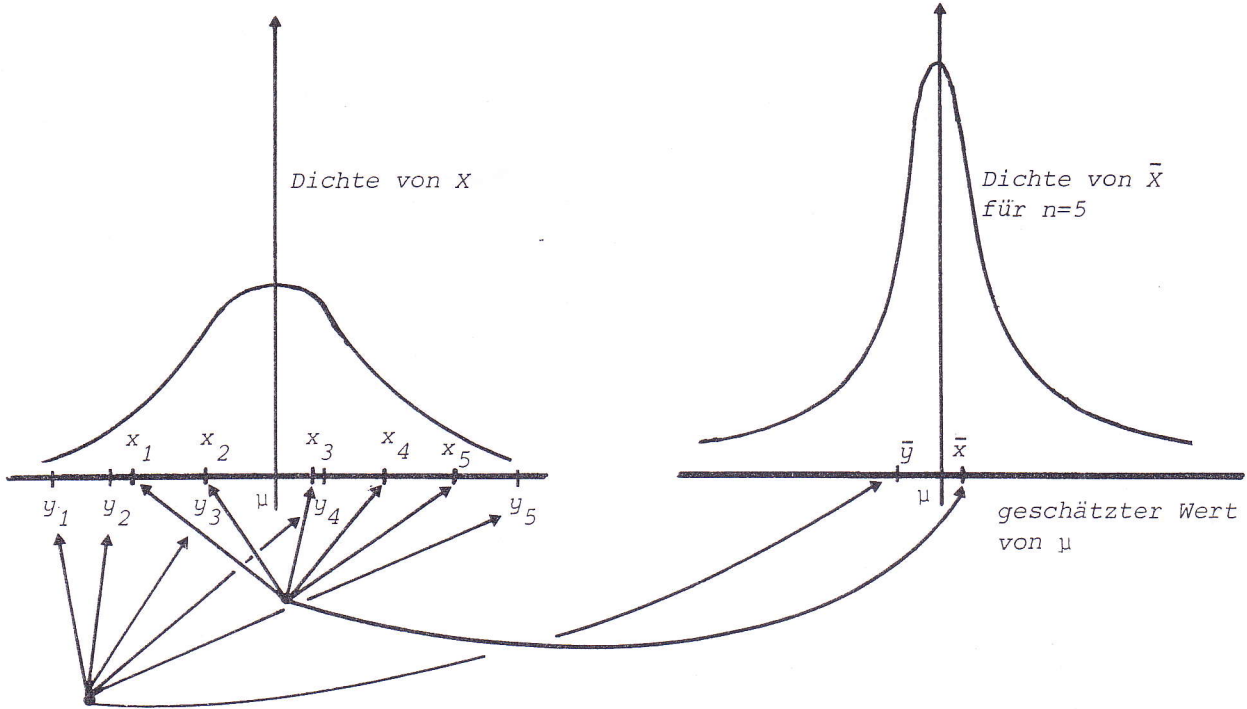


schiedlichen Schätzwerten führen.

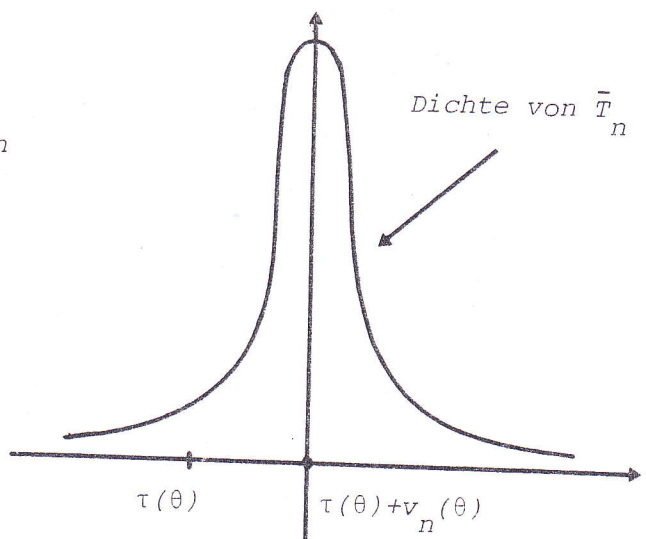
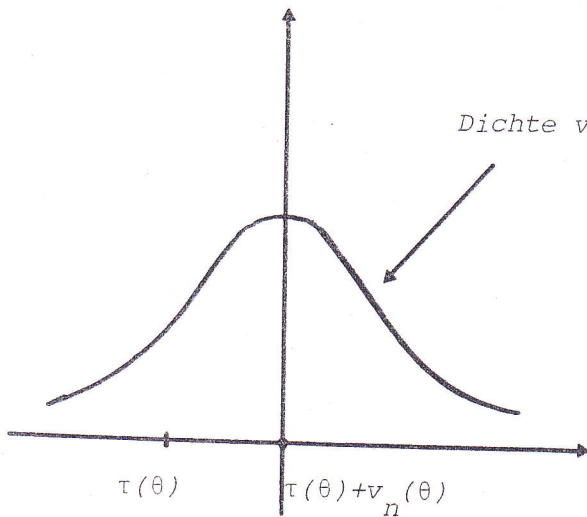


$$E_{\theta}(T_n) = \tau(\theta) + v_n(\theta) \quad \forall \theta \in \Theta, \quad ,$$

so heißt $v_n(\theta)$ Verzerrung oder Bias von T_n .

Im obigen Beispiel ist $v_n(\theta) = +\sigma^2/(n-1)$ bzw. $v_n(\theta) = -\sigma^2/n$.

Welche Auswirkung auf die Güte einer Schätzung hat die Verzerrung eines Schätzers, wenn wiederholt Schätzungen durchgeführt werden? Schätzen wir etwa k -mal mit T_n , so ist auch $\bar{T}_n := \frac{1}{k}(T_n^1 + \dots + T_n^k)$ (T_n^i ist der Schätzer für $\tau(\theta)$ bei der i -ten Schätzung) um den gleichen Betrag wie T_n verzerrt, denn es gilt ja: $E_{\theta}(\bar{T}_n) = \frac{1}{k}(E_{\theta}(T_n^1) + \dots + E_{\theta}(T_n^k)) = \frac{1}{k} \cdot kE_{\theta}(T_n) = E_{\theta}(T_n) = \tau(\theta) + v_n(\theta)$. Die Varianz von \bar{T}_n beträgt nur den k -ten Teil der Varianz von T_n , d. h. die mittels \bar{T}_n erhaltenen Schätzwerte für $\tau(\theta)$ sind eng um den "fälschen" Wert $\tau(\theta) + v_n(\theta)$ konzentriert.

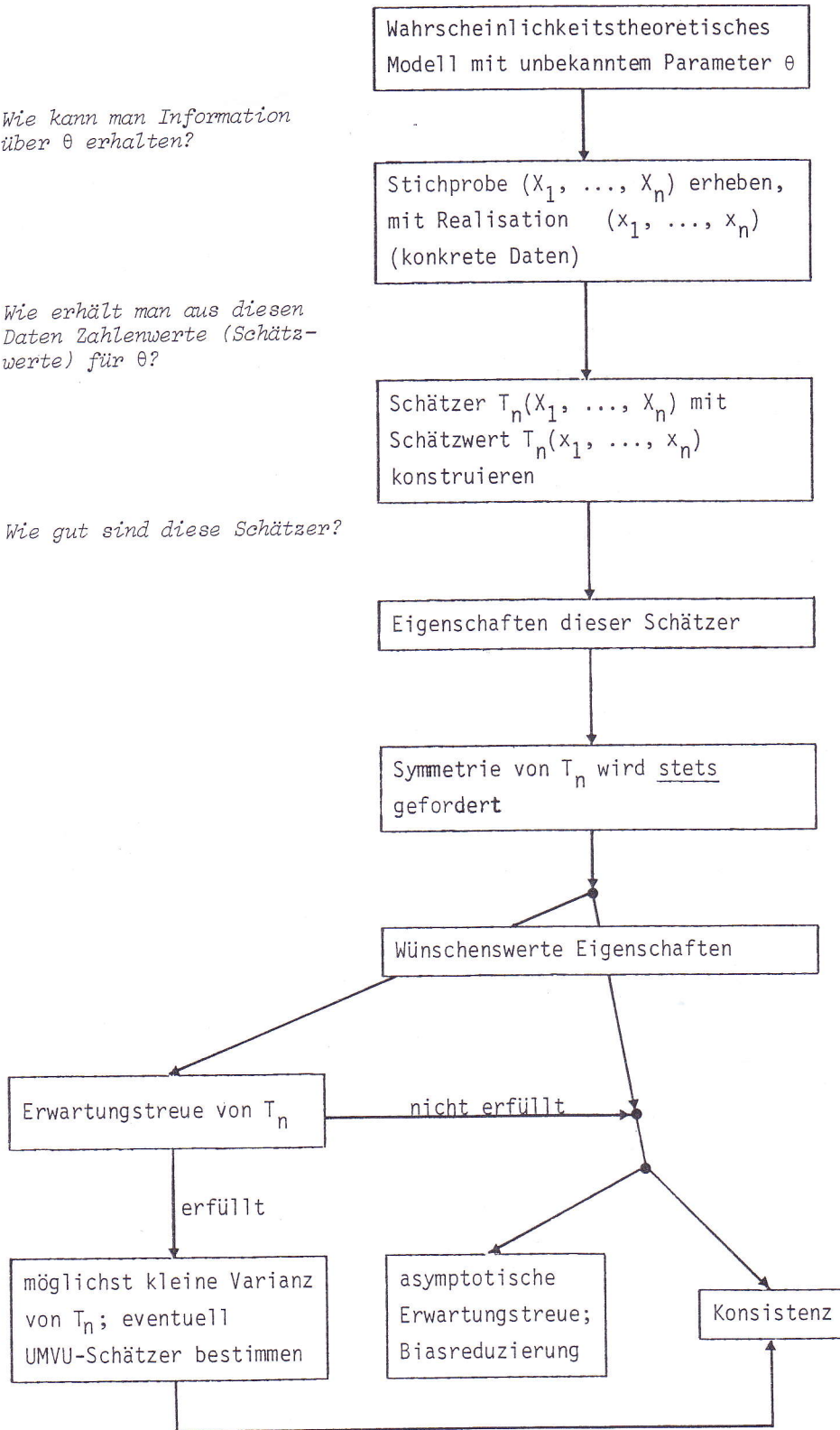


Hätten wir in Beispiel 2.5 das mittlere Gewicht eines Zementsackes nicht mit \bar{X} , sondern mit einem verzerrten Schätzer T_n geschätzt, so würde folgen:

Wie kann man Information über θ erhalten?

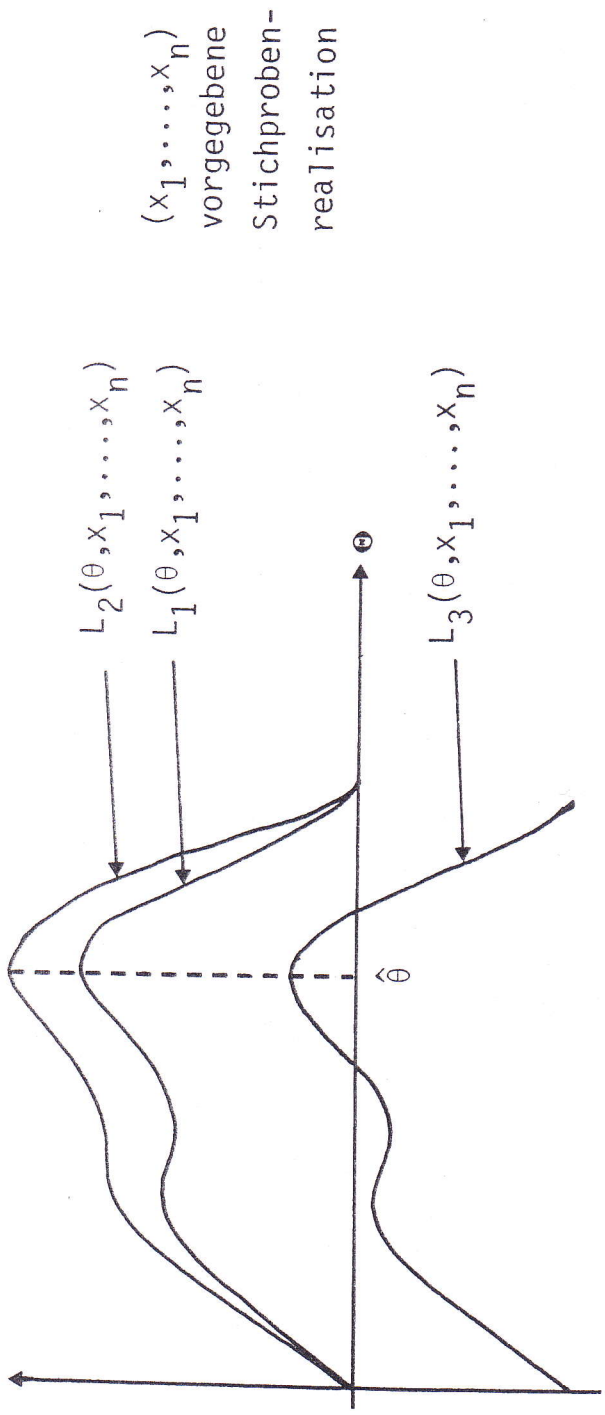
Wie erhält man aus diesen Daten Zahlenwerte (Schätzwerte) für θ ?

Wie gut sind diese Schätzer?



... wird hier im Anschluss an Definition 2.18 diskutiert.

F



$$L_1 = g_1(\theta, x_1, \dots, x_n) = f_{\theta}(x_1) \cdot \dots \cdot f_{\theta}(x_n) ; L_2 = g_2(\theta, x_1, \dots, x_n) ; L_3 = g_3(\theta, x_1, \dots, x_n)$$

Verteilung	Parameter	Schätzer nach der	
		Likelihoodmethode	Momentenmethode
Binomial- verteilung ¹	p	\bar{X}	\bar{X}
Poisson- verteilung	λ	\bar{X}	\bar{X}
geometrische Verteilung	p	$\frac{1}{\bar{X}}$	$\frac{1}{\bar{X}}$
negative Binomial- verteilung	p	$\frac{r}{r+\bar{X}}$	$\frac{r}{r+\bar{X}}$
diskrete Gleichver- teilung auf {1,...,N}	N	$\max \{X_1, \dots, X_n\}$	$2\bar{X}-1$
Gleichver- teilung auf [0,0]	θ	$\max \{X_1, \dots, X_n\}$	$2\bar{X}$
negative Ex- ponentialver- teilung	λ	$\frac{1}{\bar{X}}$	$\frac{1}{\bar{X}}$
Gamma-Ver- teilung	r λ	Formelmäßig nicht angebbbar. Muß iterativ berechnet wer- den. Als Startwerte können die Momentenschätzwerte ver- wendet werden.	$\bar{X}^2 \cdot \frac{1}{M_2}$ $\bar{X} \cdot \frac{1}{M_2}$
Normal- verteilung ²	μ σ^2	\bar{X} $\frac{1}{n} \sum (X_i - \bar{X})^2$	\bar{X} $\frac{1}{n} \sum (X_i - \bar{X})^2$

¹ Man beachte, daß p aufgrund der Realisationen von n unabhängigen zweipunkt-ver-
teilten Zufallsvariablen geschätzt wird.

² Man beachte, daß beide Parameter unbekannt sind.

- a) Fig. 17-7 oben: Das VI gebildet mit \bar{x} als Punktschätzung überdeckt in 95% der Fälle (allgemein: in $(1-\alpha) \cdot 100\%$ der Fälle) sein "eigenes" μ , d.h. jenes μ , das in der GG vorlag, aus der dieses \bar{x} tatsächlich stammte.
- b) Fig. 17-7 unten: Das VI schneidet aus der Menge aller μ jene heraus, die in extremis noch als "eigenes" μ zu \bar{x} in Frage kommen können..

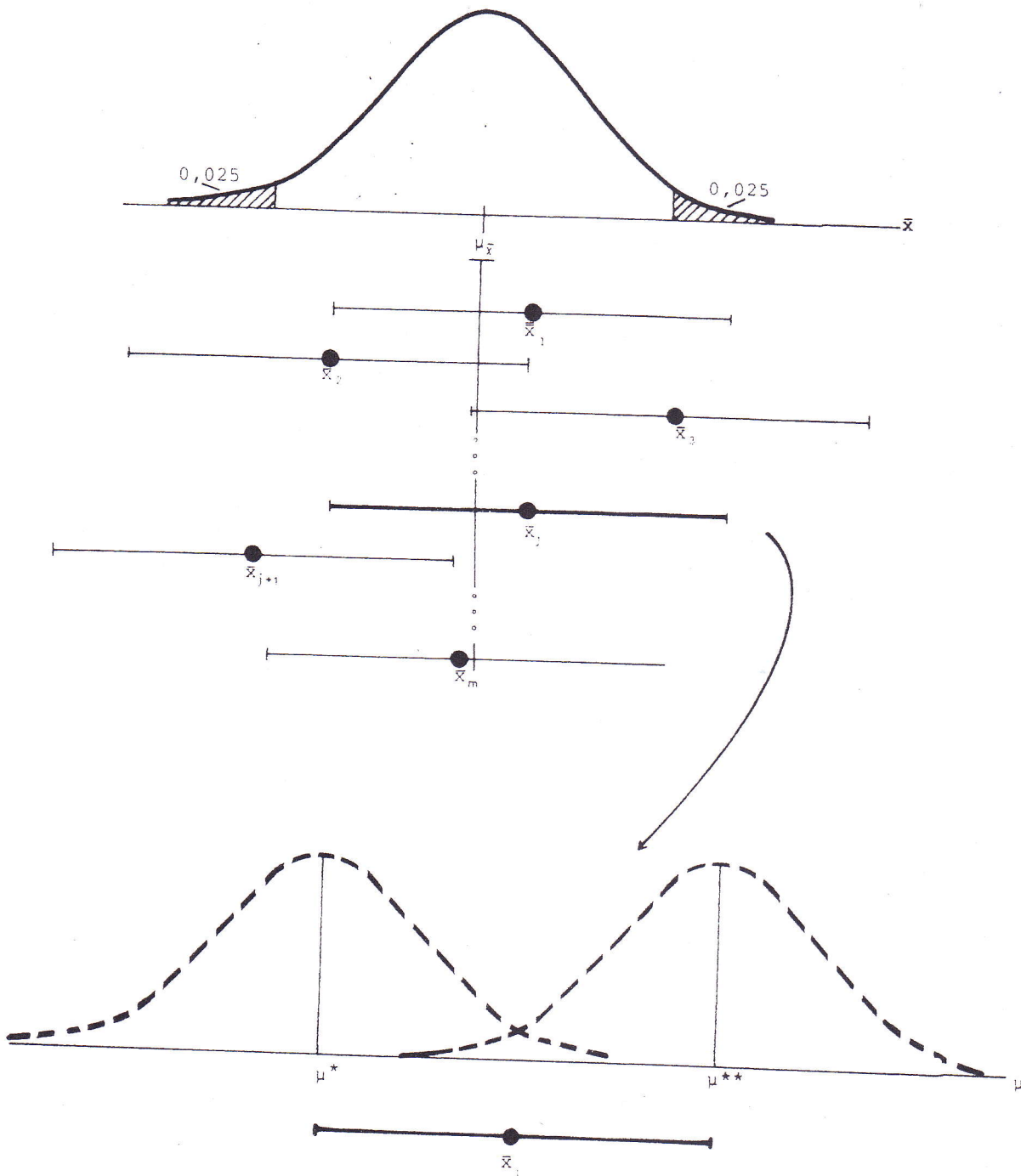
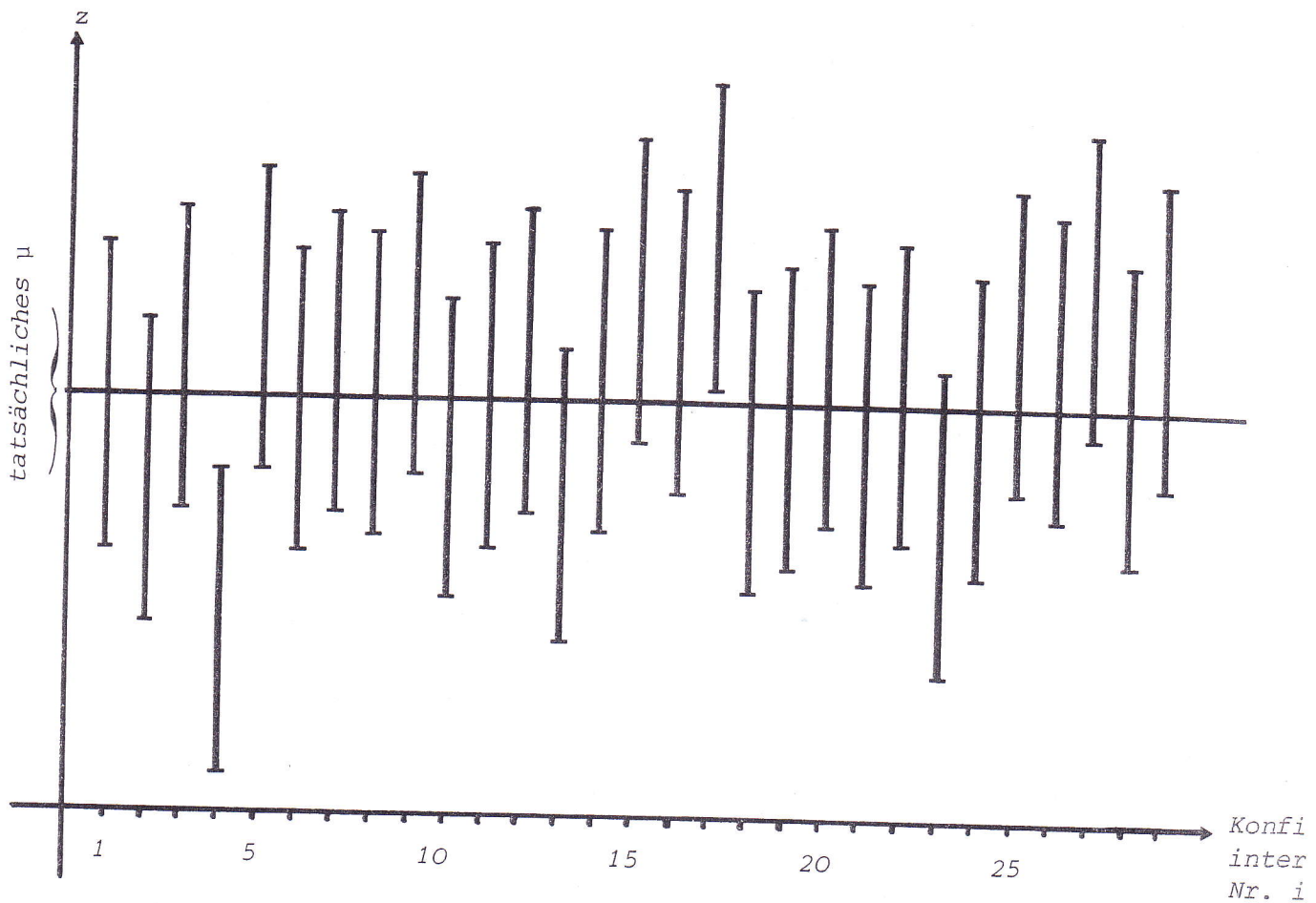


Fig. 17 - 7: Die zwei "Eigenschaften" eines VI

Stichprobenrealisation abhängt.

Graphisch lassen sich die erhaltenen Konfidenzintervalle sehr schön veranschaulichen.



Tragen wir eine große Anzahl von Konfidenzintervallen in unser "Koordinatensystem" ein, so müßten in etwa $(1-\alpha)100\%$ dieser Intervalle (angedeutet durch die senkrechten Strecken) die Gerade $z=\mu$ schneiden.

Wir haben uns anfangs auf die Sprechweise "Das Konfidenzintervall enthält den wahren Parameter!" geeinigt; manchmal sagen wir auch: "Das Konfidenzintervall überdeckt den wahren Parameter". Es wäre allerdings irreführend zu sagen, der unbekannte Parameter liege mit der Wahrscheinlichkeit $1-\alpha$ im Konfidenzintervall. In diesem Fall würden wir dem Parameter stochastischen Charakter zubilligen, den er ja nicht besitzt.

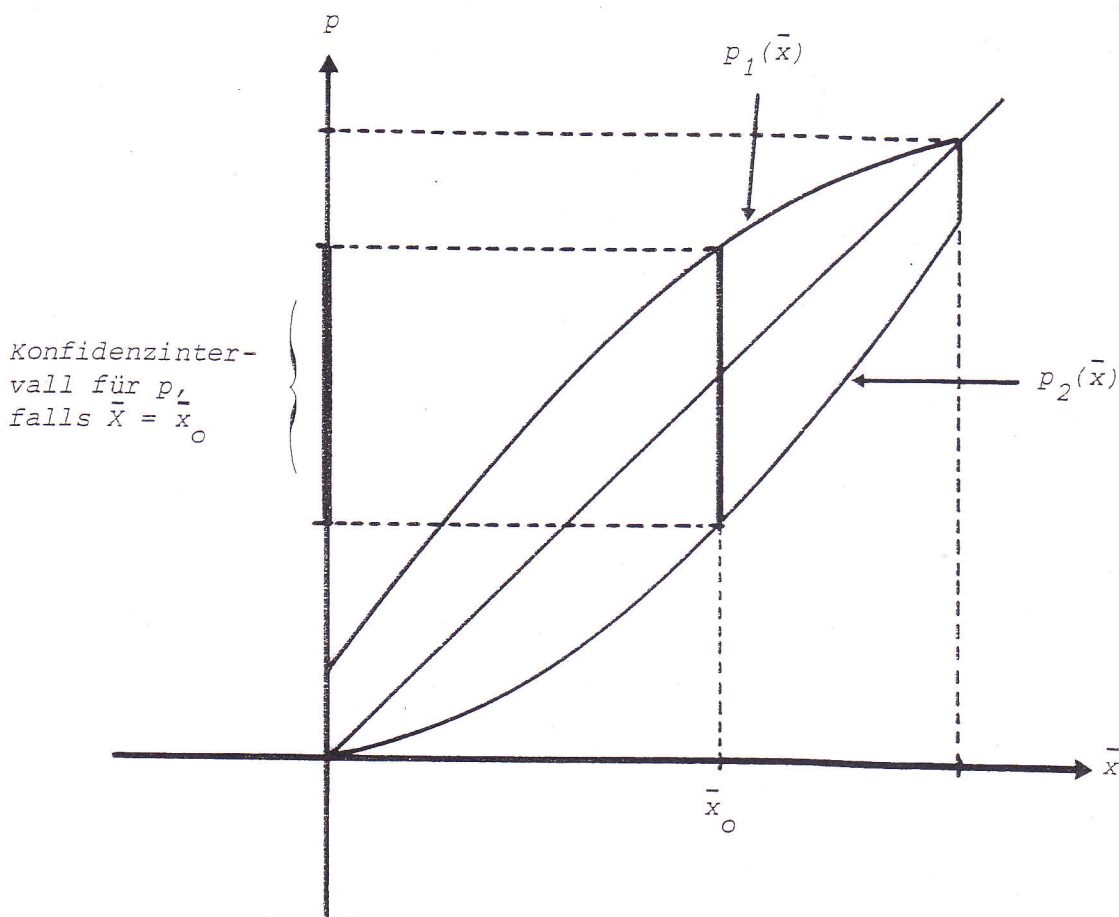
An unserem einführenden Beispiel haben wir die Vorgehensweise...

$$p_{1,2}(\bar{x}) = \frac{\bar{x} + \frac{c^2}{2n} \pm \sqrt{\bar{x}(1-\bar{x}) \frac{c^2}{n} + \frac{c^4}{4n^2}}}{1 + \frac{c^2}{n}}$$

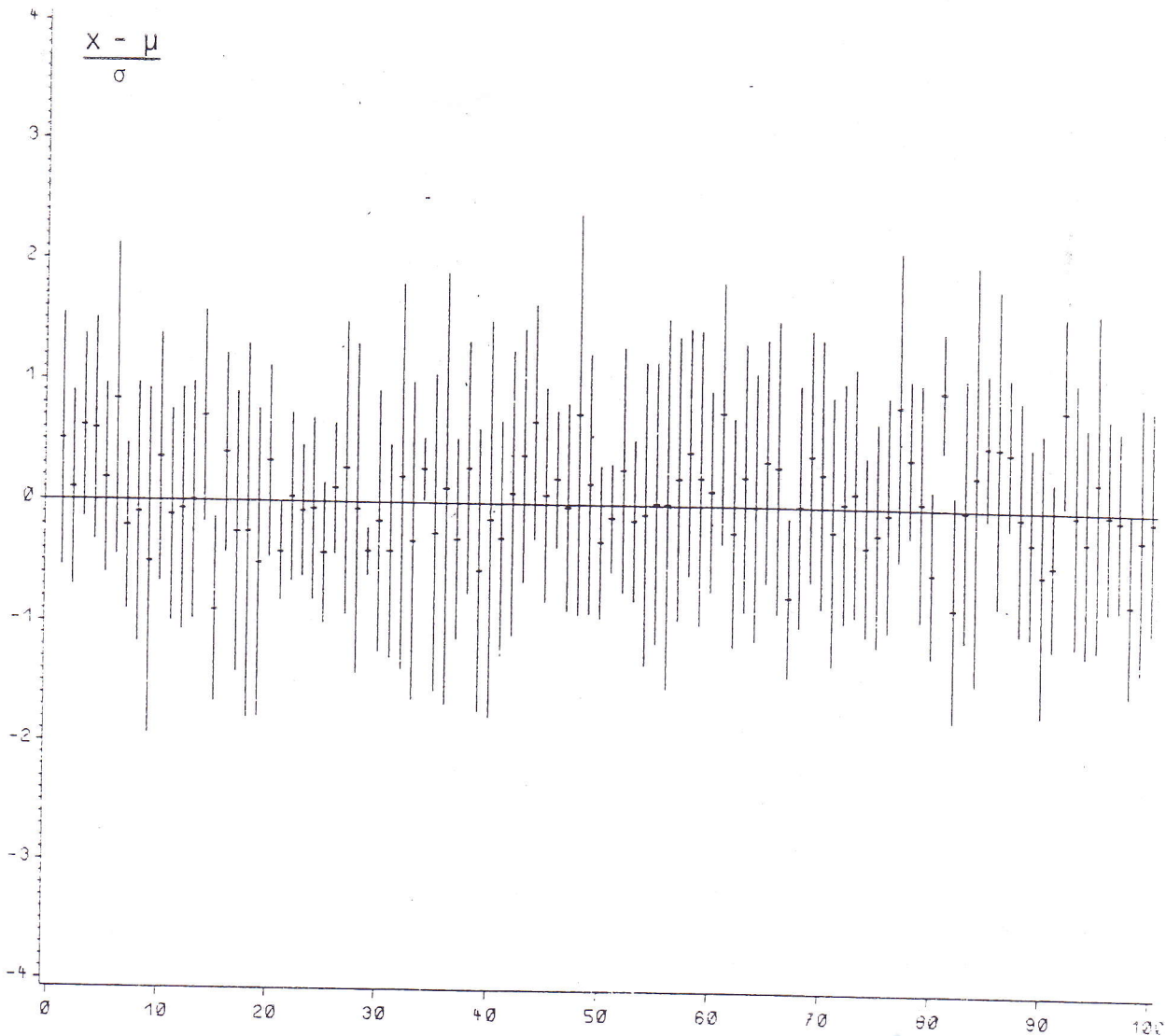
Das approximative $(1-\alpha)100\%$ Konfidenzintervall für p lautet dann:

$$\left[\frac{\bar{x} + \frac{c^2}{2n} - \sqrt{\bar{x}(1-\bar{x}) \frac{c^2}{n} + \frac{c^4}{4n^2}}}{1 + \frac{c^2}{n}}, \frac{\bar{x} + \frac{c^2}{2n} + \sqrt{\bar{x}(1-\bar{x}) \frac{c^2}{n} + \frac{c^4}{4n^2}}}{1 + \frac{c^2}{n}} \right]$$

Die Konfidenzgrenzen $p_{1,2}(\bar{x})$ lassen sich graphisch sehr leicht ermitteln.



Eine etwas übersichtlichere (approximative) Darstellung erhalten wir, wenn in $(\bar{x}-p)/\sqrt{\frac{pq}{n}}$ der Nenner durch $\sqrt{\bar{x}(1-\bar{x})/n}$ geschätzt wird. Es ergibt sich dann:



46: Monte Carlo-Experiment mit Konfidenzintervallen.

bemerken,

1. daß die Länge der Intervalle ebenso wie ihre Lage (Intervallmittelpunkt: das jeweilige \bar{x}) zufällig fluktuiert und
2. daß von den 100 unabhängig voneinander generierten 95%-Konfidenzintervallen 92 den wahren Lageparameter μ einschließen (zwei weitere Intervalle sind "Grenzfälle").

Fig 29: Wir haben als Übungsbeispiel in Abschnitt 5.3 des ersten Teils ein Probenexperiment beschrieben, in dem durch Zufallsziehung sechzehn 25er-Proben aus einer Gesamtheit von 505 Klausurpunktzahlen erzeugt wurden. In der Tabelle 4 (Teil 1, S. 331) wurden alle Resultate $\bar{x}_{(25)}$, $s_{(25)}$ zusammenge-