6 Dichte-Schätzung

Problem: Gegeben: Daten $x_1, \ldots, x_n \in \mathbb{R}$ als Resultate „unabhängiger“ Wiederholungen eines Zufallsexperimentes.

Annahme: Einzelexperiment genügt absolutstetiger Verteilung.

Gesucht: Dichte der Verteilung für Einzelexperiment.

Hier die einfachste Methode: Approximation der Dichte durch Treppenfunktion.

Formale Beschreibung:

- $(X_i)_{i\in\mathbb{N}}$ iid-ZVen auf $(\Omega, \mathfrak{A}, P)$, X_1 absolutstetig verteilt mit unbekannter Dichte $f := f_{X_1}$.

  Jede dieser ZVen modelliert ein Einzelexperiment.

- Annahme: Daten $x_1, \ldots, x_n$ sind eine Realisierung der ZVen $X_1, \ldots, X_n$, d.h. für ein $\omega \in \Omega$

$$x_1 = X_1(\omega), \ldots, x_n = X_n(\omega)$$

Fixiere Folge $(I_k)_{k\in\mathbb{Z}}$ von beschränkten Intervallen mit

- $\bigcup_{k\in\mathbb{Z}} I_k = \mathbb{R}$,
- $I_k \cap I_\ell = \emptyset$ für $k \neq \ell$.

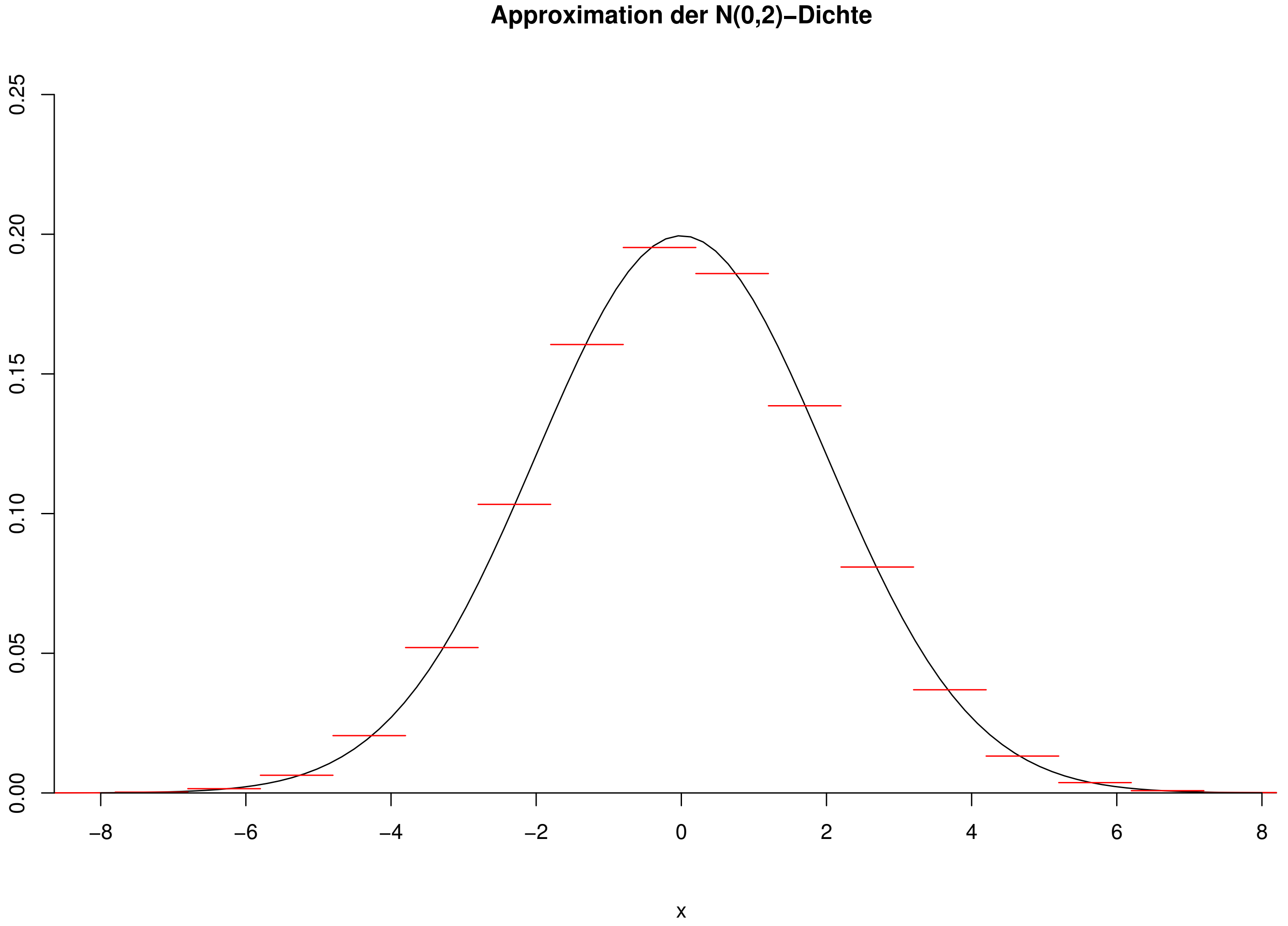
Definiere stückweise konstante Dichte $\widetilde{f} : \mathbb{R} \to [0, \infty[$ durch

$$\widetilde{f}(x) := \sum_{k\in\mathbb{Z}} c_k \cdot 1_{I_k}(x)$$

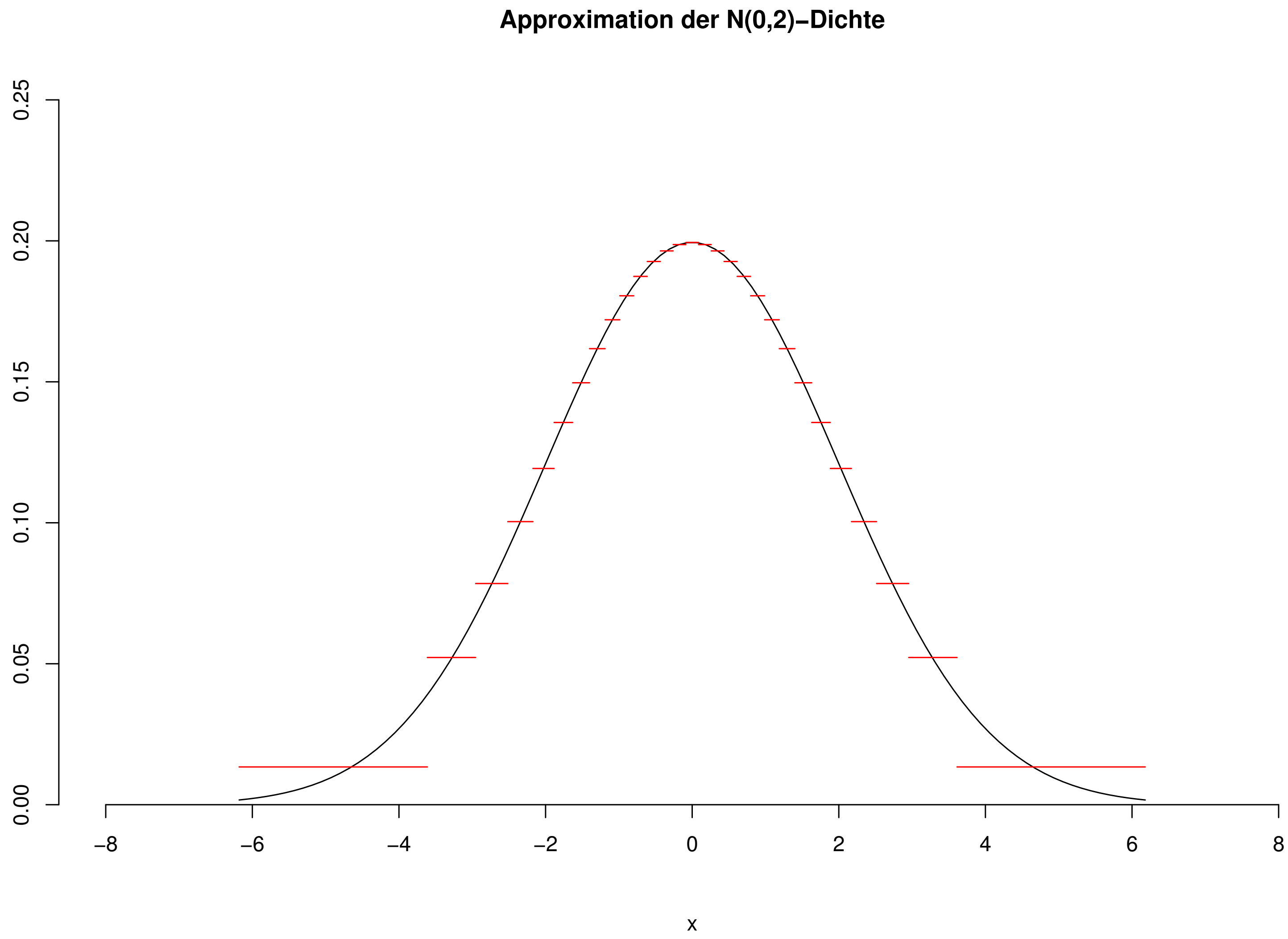
mit

$$c_k := \frac{\int_{I_k} f(u)\, du}{\lambda_1(I_k)} = \frac{P(\{X_1 \in I_k\})}{\lambda_1(I_k)}.$$

56. Beispiel $\mathrm{N}(0,2)$, Intervalle der Länge eins.



57. Beispiel $\mathrm{N}(0,2)$, Intervalle nicht-konstanter Länge.



58. Definition Histogramm $f_n(\cdot; x_1, \dots, x_n) : \mathbb{R} \to [0, \infty[$ zu $x_1, \dots, x_n \in \mathbb{R}$ definiert durch

$$f_n(x; x_1, \dots, x_n) := \sum_{k \in \mathbb{Z}} c_{n,k}(x_1, \dots, x_n) \cdot 1_{I_k}(x)$$

mit

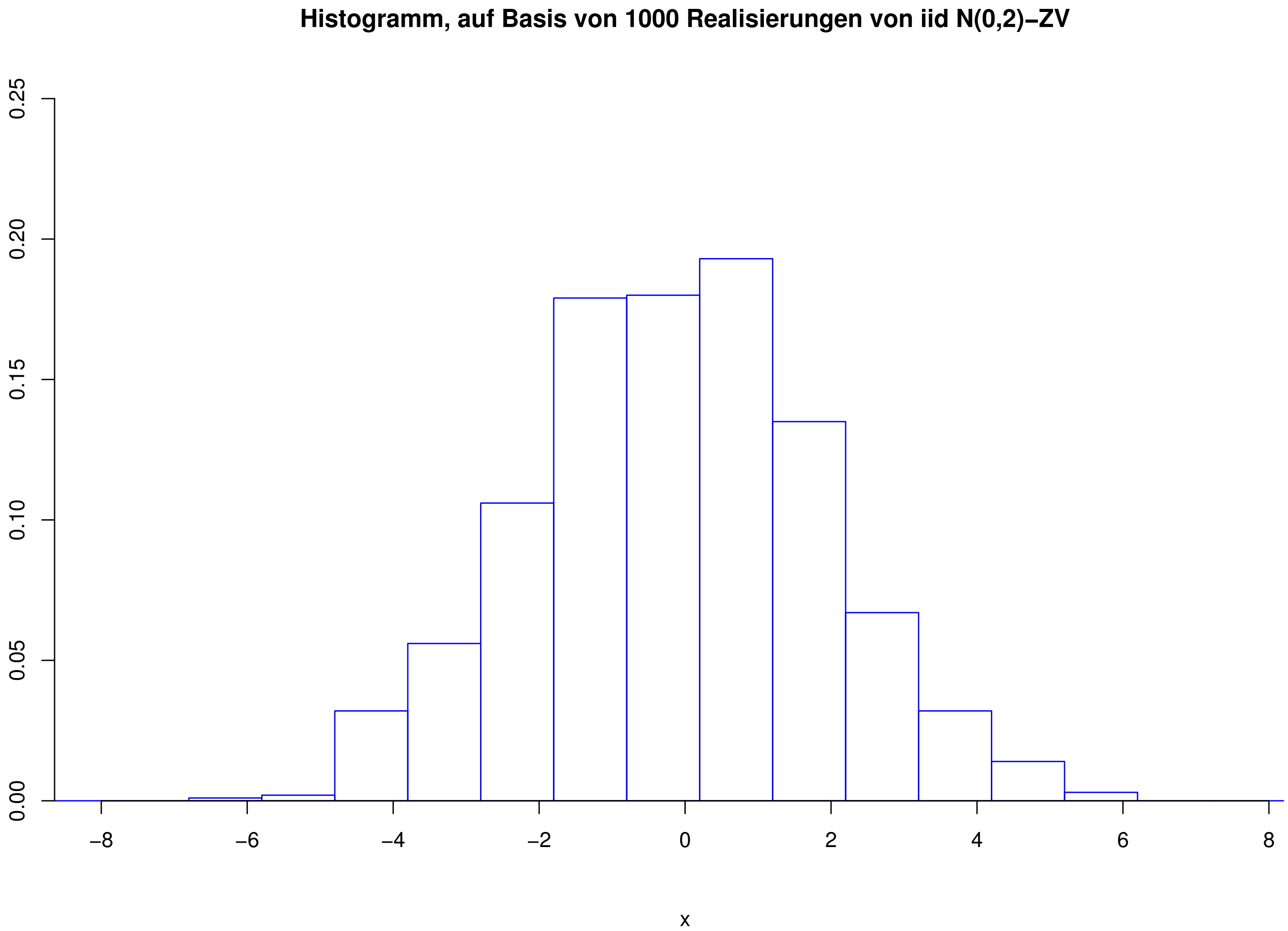
$$c_{n,k}(x_1, \dots, x_n) := \frac{|\{i \in \{1, \dots, n\} : x_i \in I_k\}|}{n \cdot \lambda_1(I_k)}.$$

59. Beispiel PROJEKTOR

60. Bemerkung Histogramm $f_n(\cdot; x_1, \dots, x_n)$ ist stückweise konstante Dichte, die nur auf endlich-vielen Intervallen ungleich null ist.

61. Beispiel

Histogramm, $n = 1000$, Intervalle der Länge eins.



62. Satz Für alle $x \in \mathbb{R}$ gilt für fast alle ω

$$\lim_{n\to\infty} f_n(x; X_1(\omega), \ldots, X_n(\omega)) = \widetilde{f}(x).$$

Verschärfung: fast sicher gleichmäßige Konvergenz.

Beweis. Sei $x \in I_k$. Für

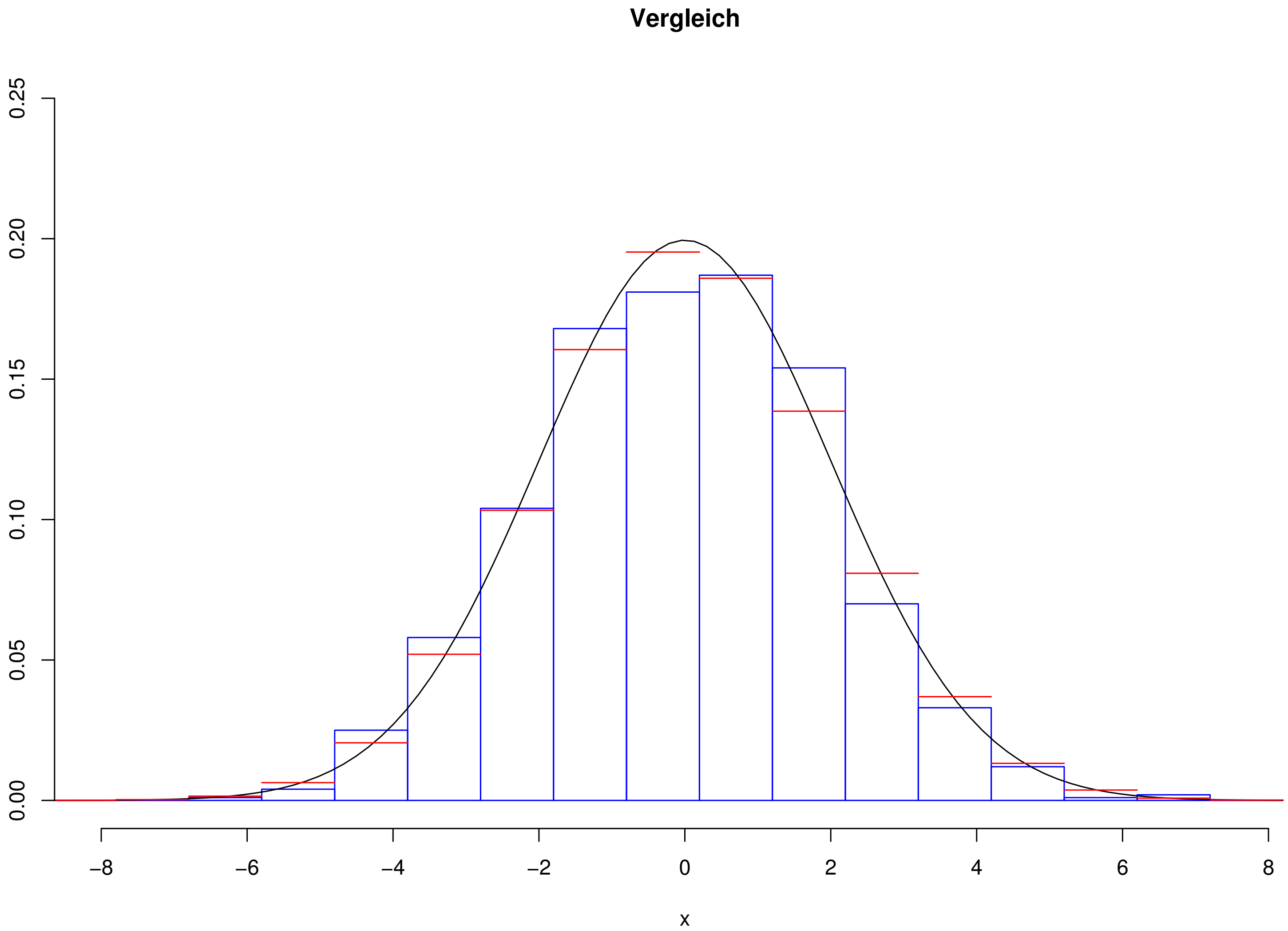
$$Z_i := 1_{I_k}(X_i)$$

gilt $(Z_i)_{i\in\mathbb{N}}$ iid mit $Z_1 \sim \mathbf{B}(1, p)$ für

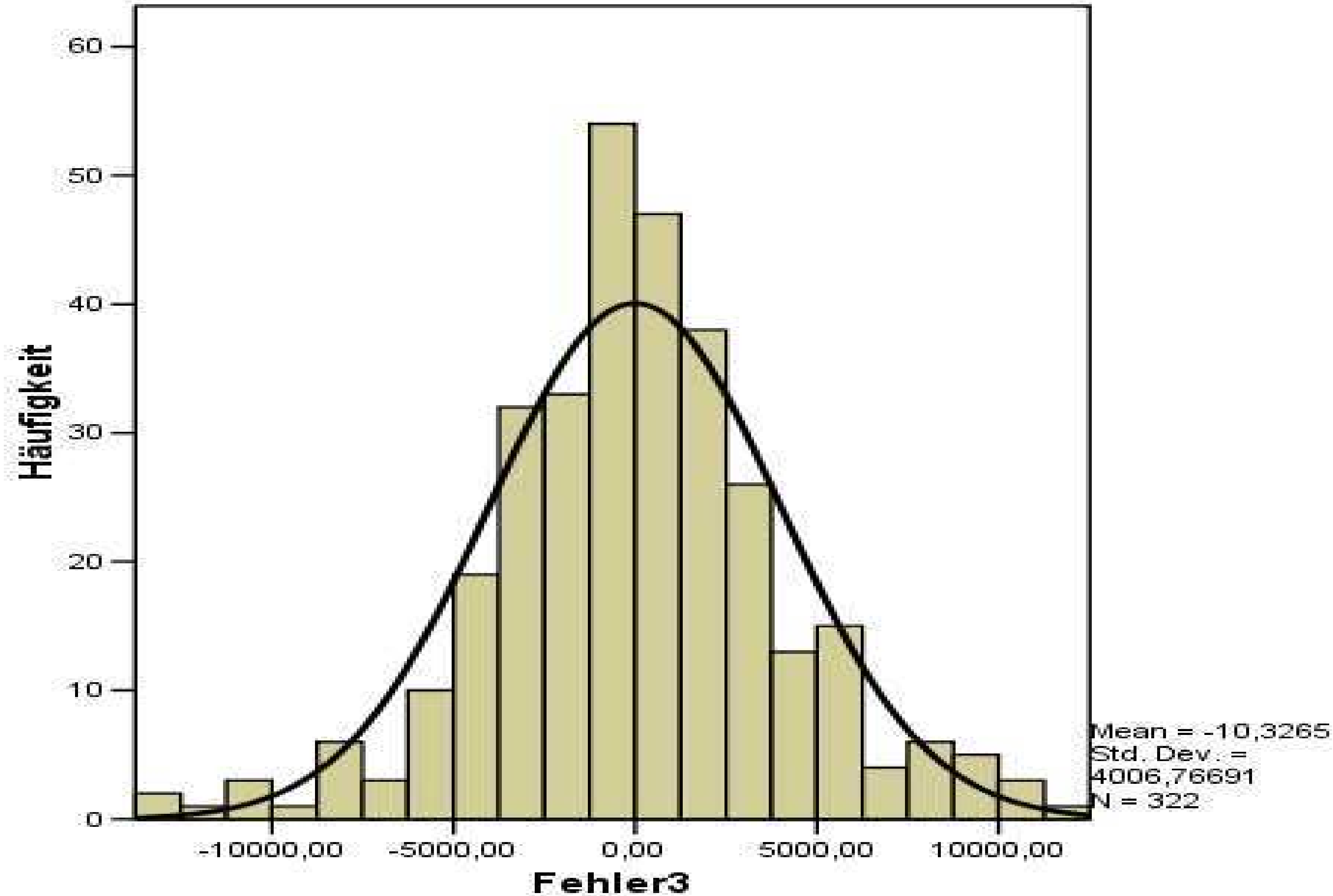
$$p = P(\{X_1 \in I_k\}) = \int_{I_k} f(u)\, du.$$

Wende Satz IV.7 an. $\square$

63. Beispiel Vergleich der Schätzung aus Bsp. 61 mit der Dichte und ihrer Approximation aus Bsp. 56.



64. Beispiel Prognosefehler für Tagesumsätze in Geldautomat, siehe Kapitel I.



65. Bemerkung Fehler der Approximation von f durch $\widetilde{f}$?

Setze $\delta := \sup_{k\in\mathbb{Z}} \lambda_1(I_k)$. Dann

$$\sup_{x\in\mathbb{R}} |f(x) - \widetilde{f}(x)|$$
$$\leq \sup\{|f(s) - f(t)| : s, t \in \mathbb{R},\ |s-t| \leq \delta\}.$$

Anwendbar, falls f gleichmäßig stetig, also speziell falls f stetig und $\lim_{x\to\infty} f(x) = \lim_{x\to-\infty} f(x) = 0$.