KAPITEL 3

Empirische Verteilungsfunktion

3.1. Empirische Verteilungsfunktion

Seien $X_1, \ldots, X_n$ unabhängige und identisch verteilte Zufallsvariablen mit theoretischer Verteilungsfunktion

$$F(t) = \mathbb{P}[X_i \leq t].$$

Es sei $(x_1, \ldots, x_n)$ eine Realisierung dieser Zufallsvariablen. Wie können wir die theoretische Verteilungsfunktion F anhand der Stichprobe $(x_1, \ldots, x_n)$ schätzen? Dafür benötigen wir die empirische Verteilungsfunktion.

DEFINITION 3.1.1. Die *empirische Verteilungsfunktion* einer Stichprobe $(x_1, \ldots, x_n) \in \mathbb{R}^n$ ist definiert durch

$$\widehat{F}_n(t) := \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{x_i \leq t} = \frac{1}{n} \# \{i \in \{1, \ldots, n\} : x_i \leq t\}, \quad t \in \mathbb{R}.$$

BEMERKUNG 3.1.2. Die oben definierte empirische Verteilungsfunktion kann wie folgt durch die Ordnungsstatistiken $x_{(1)}, \ldots, x_{(n)}$ ausgedrückt werden

$$\widehat{F}_n(t) = \begin{cases} 0, & \text{falls } t < x_{(1)}, \\ \frac{1}{n}, & \text{falls } x_{(1)} \leq t < x_{(2)}, \\ \frac{2}{n}, & \text{falls } x_{(2)} \leq t < x_{(3)}, \\ \ldots & \ldots \\ \frac{n-1}{n}, & \text{falls } x_{(n-1)} \leq t < x_{(n)}, \\ 1, & \text{falls } x_{(n)} \leq t. \end{cases}$$

BEMERKUNG 3.1.3. Die empirische Verteilungsfunktion $\widehat{F}_n$ hat alle Eigenschaften einer Verteilungsfunktion, denn es gilt

(1) $\lim_{t \to -\infty} \widehat{F}_n(t) = 0$ und $\lim_{t \to +\infty} \widehat{F}_n(t) = 1$.
(2) $\widehat{F}_n$ ist monoton nichtfallend.
(3) $\widehat{F}_n$ ist rechtsstetig.

Parallel werden wir auch die folgende Definition benutzen.

DEFINITION 3.1.4. Seien $X_1, \ldots, X_n$ unabhängige und identisch verteilte Zufallsvariablen. Dann ist die empirische Verteilungsfunktion gegeben durch

$$\widehat{F}_n(t) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{X_i \leq t}, \quad t \in \mathbb{R}.$$

Es sei bemerkt, dass $\widehat{F}_n(t)$ für jedes $t \in \mathbb{R}$ eine Zufallsvariable ist. Somit ist $\widehat{F}_n$ eine zufällige Funktion. Auf die Eigenschaften von $\widehat{F}_n(t)$ gehen wir im folgenden Satz ein.

SATZ 3.1.5. *Seien $X_1, \ldots, X_n$ unabhängige und identisch verteilte Zufallsvariablen mit Verteilungsfunktion F. Dann gilt*

(1) *Die Zufallsvariable $n\widehat{F}_n(t)$ ist binomialverteilt:*

$$n\widehat{F}_n(t) \sim Bin(n, F(t)).$$

Das heißt:

$$\mathbb{P}\left[\widehat{F}_n(t) = \frac{k}{n}\right] = \binom{n}{k} F(t)^k (1 - F(t))^{n-k}, \quad k = 0, 1, \ldots, n.$$

(2) *Für den Erwartungswert und die Varianz von $\widehat{F}_n(t)$ gilt:*

$$\mathbb{E}[\widehat{F}_n(t)] = F(t), \quad \mathrm{Var}[\widehat{F}_n(t)] = \frac{F(t)(1 - F(t))}{n}.$$

Somit ist $\widehat{F}_n(t)$ ein erwartungstreuer Schätzer für $F(t)$.

(3) *Für alle $t \in \mathbb{R}$ gilt*

$$\widehat{F}_n(t) \xrightarrow[n\to\infty]{f.s.} F(t).$$

In diesem Zusammenhang sagt man, dass $\widehat{F}_n(t)$ ein "stark konsistenter" Schätzer für $F(t)$ ist.

(4) *Für alle $t \in \mathbb{R}$ mit $F(t) \neq 0, 1$ gilt:*

$$\sqrt{n}\frac{\widehat{F}_n(t) - F(t)}{\sqrt{F(t)(1 - F(t))}} \xrightarrow[n\to\infty]{d} N(0, 1).$$

In diesem Zusammenhang sagt man, dass $\widehat{F}_n(t)$ ein "asymptotisch normalverteilter Schätzer" für $F(t)$ ist.

BEMERKUNG 3.1.6. Die Aussage von Teil 4 kann man folgendermaßen verstehen: Die Verteilung des Schätzfehlers $\widehat{F}_n(t) - F(t)$ ist für große Werte von n approximativ

$$\mathrm{N}\left(0, \frac{F(t)(1 - F(t))}{n}\right).$$

BEWEIS VON (1). Wir betrachten n Experimente. Beim i-ten Experiment überprüfen wir, ob $X_i \leq t$. Falls $X_i \leq t$, sagen wir, dass das i-te Experiment ein Erfolg ist. Die Experimente sind unabhängig voneinander, denn die Zufallsvariablen $X_1, \ldots, X_n$ sind unabhängig. Die Erfolgswahrscheinlichkeit in jedem Experiment ist $\mathbb{P}[X_i \leq t] = F(t)$. Die Anzahl der Erfolge in den n Experimenten, also die Zufallsvariable

$$n\widehat{F}_n(t) = \sum_{i=1}^{n} \mathbb{1}_{X_i \leq t}$$

muss somit binomialverteilt mit Parametern n (Anzahl der Experimente) und $F(t)$ (Erfolgswahrscheinlichkeit) sein.

Beweis von (2). Wir haben in (1) gezeigt, dass $n\widehat{F}_n(t) \sim \text{Bin}(n, F(t))$. Der Erwartungswert einer binomialverteilten Zufallsvariable ist die Anzahl der Experimente multipliziert mit der Erfolgswahrscheinlichkeit. Also gilt

$$\mathbb{E}[n\widehat{F}_n(t)] = nF(t).$$

Teilen wir beide Seiten durch n, so erhalten wir $\mathbb{E}[\widehat{F}_n(t)] = F(t)$.
Die Varianz einer $\text{Bin}(n,p)$-verteilten Zufallsvariable ist $np(1-p)$, also

$$\text{Var}[n\widehat{F}_n(t)] = nF(t)(1-F(t)).$$

Wir können nun das n aus der Varianz herausziehen, allerdings wird daraus (nach den Eigenschaften der Varianz) n^2. Indem wir nun beide Seiten durch n^2 teilen, erhalten wir

$$\text{Var}[\widehat{F}_n(t)] = \frac{F(t)(1-F(t))}{n}.$$

Beweis von (3). Wir führen die Zufallsvariablen $Y_i = \mathbb{1}_{X_i \le t}$ ein. Diese sind unabhängig und identisch verteilt (da $X_1, X_2, \ldots,$ unabhängig und identisch verteilt sind) mit

$$\mathbb{P}[Y_i = 1] = \mathbb{P}[X_i \le t] = F(t), \quad \mathbb{P}[Y_i = 0] = 1 - \mathbb{P}[X_i \le t] = 1 - F(t).$$

Es gilt also $\mathbb{E}Y_i = F(t)$. Wir können nun das starke Gesetz der großen Zahlen auf die Folge $Y_1, Y_2, \ldots$ anwenden:

$$\widehat{F}_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{X_i \le t} = \frac{1}{n}\sum_{i=1}^{n} Y_i \xrightarrow[n\to\infty]{f.s.} \mathbb{E}Y_1 = F(t).$$

Beweis von (4). Mit der Notation von Teil (3) gilt

$$\mathbb{E}Y_i = F(t) \quad \text{Var}\, Y_i = F(t)(1-F(t)).$$

Wir wenden den zentralen Grenzwertsatz auf die Folge $Y_1, Y_2, \ldots$ an:

$$\sqrt{n}\frac{\widehat{F}_n(t) - F(t)}{\sqrt{F(t)(1-F(t))}} = \sqrt{n}\frac{\frac{1}{n}\sum_{i=1}^{n} Y_i - \mathbb{E}Y_1}{\sqrt{\text{Var}\, Y_1}} = \frac{\sum_{i=1}^{n} Y_i - n\mathbb{E}Y_1}{\sqrt{n\,\text{Var}\, Y_1}} \xrightarrow[n\to\infty]{d} \text{N}(0,1).$$

□

3.2. Empirische Verteilung

Mit Hilfe der empirischen Verteilungsfunktion können wir also die theoretische Verteilungsfunktion schätzen. Nun führen wir auch die empirische Verteilung ein, mit der wir die theoretische Verteilung schätzen können. Zuerst definieren wir, was die theoretische Verteilung ist.

Definition 3.2.1. Sei X eine Zufallsvariable. Die *theoretische Verteilung* von X ist ein Wahrscheinlichkeitsmaß μ auf $(\mathbb{R}, \mathcal{B})$ mit

$$\mu(A) = \mathbb{P}[X \in A] \text{ für jede Borel-Menge } A \subset \mathbb{R}.$$

Der Zusammenhang zwischen der theoretischen Verteilung μ und der theoretischen Verteilungsfunktion F einer Zufallsvariable ist dieses:

$$F(t) = \mu((-\infty, t]), \quad t \in \mathbb{R}.$$

Wie können wir die theoretische Verteilung anhand einer Stichprobe $(x_1, \ldots, x_n)$ schätzen?

DEFINITION 3.2.2. Die *empirische Verteilung* einer Stichprobe $(x_1, \ldots, x_n) \in \mathbb{R}^n$ ist ein Wahrscheinlichkeitsmaß $\widehat{\mu}_n$ auf $(\mathbb{R}, \mathcal{B})$ mit

$$\widehat{\mu}_n(A) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{x_i \in A} = \frac{1}{n} \# \{i \in \{1, \ldots, n\} : x_i \in A\}.$$

Die theoretische Verteilung $\widehat{\mu}_n$ ordnet jeder Menge A die Wahrscheinlichkeit, dass X einen Wert in A annimmt, zu. Die empirische Verteilung ordnet jeder Menge A den Anteil der Stichprobe, der in A liegt, zu.
Die empirische Verteilung $\widehat{\mu}_n$ kann man sich folgendermaßen vorstellen: Sie ordnet jedem der Punkte x_i aus der Stichprobe das gleiche Gewicht $1/n$ zu. Falls ein Wert mehrmals in der Stichprobe vorkommt, wird sein Gewicht entsprechend erhöht. Dem Rest der reellen Geraden, also der Menge $\mathbb{R} \backslash \{x_1, \ldots, x_n\}$, ordnet $\widehat{\mu}_n$ Gewicht 0 zu. Am Besten kann man das mit dem Begriff des Dirac-δ-Maßes beschreiben.

DEFINITION 3.2.3. Sei $x \in \mathbb{R}$ eine Zahl. Das *Dirac-δ-Maß* δ_x ist ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B})$ mit

$$\delta_x(A) = \begin{cases} 1, & \text{falls } x \in A, \\ 0, & \text{falls } x \notin A \end{cases} \quad \text{für alle Borel-Mengen } A \subset \mathbb{R}.$$

Das Dirac-δ-Maß δ_x ordnet dem Punkt x das Gewicht 1 zu. Der Menge $\mathbb{R} \backslash \{x\}$ ordnet es das Gewicht 0 zu. Die empirische Verteilung $\widehat{\mu}_n$ lässt sich nun wie folgt darstellen:

$$\widehat{\mu}_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i}.$$

Zwischen der empirischen Verteilung $\widehat{\mu}_n$ und der empirischen Verteilungsfunktion $\widehat{F}_n$ besteht der folgende Zusammenhang:

$$\widehat{F}_n(t) = \widehat{\mu}_n((-\infty, t]).$$

3.3. Satz von Gliwenko–Cantelli

Wir haben in Teil 3 von Satz 3.1.5 gezeigt, dass für jedes $t \in \mathbb{R}$ die Zufallsvariable $\widehat{F}_n(t)$ gegen die Konstante $F(t)$ fast sicher konvergiert. Man kann auch sagen, dass die empirische Verteilungsfunktion $\widehat{F}_n$ punktweise fast sicher gegen die theoretische Verteilungsfunktion $F(t)$ konvergiert. Im nächsten Satz beweisen wir eine viel stärkere Aussage. Wir zeigen nämlich, dass die Konvergenz mit Wahrscheinlichkeit 1 sogar *gleichmäßig* ist.

DEFINITION 3.3.1. Der *Kolmogorov-Abstand* zwischen der empirischen Verteilungsfunktion $\widehat{F}_n$ und der theoretischen Verteilungsfunktion F wird folgendermaßen definiert:

$$D_n := \sup_{t \in \mathbb{R}} |\widehat{F}_n(t) - F(t)|.$$

Satz 3.3.2 (von Gliwenko–Cantelli). *Für den Kolmogorov-Abstand D_n gilt*

$$D_n \xrightarrow[n\to\infty]{f.s.} 0.$$

Mit anderen Worten, es gilt

$$\mathbb{P}\left[\lim_{n\to\infty} D_n = 0\right] = 1.$$

Beispiel 3.3.3. Da aus der fast sicheren Konvergenz die Konvergenz in Wahrscheinlichkeit folgt, gilt auch

$$D_n \xrightarrow[n\to\infty]{P} 0.$$

Somit gilt für alle $\varepsilon > 0$:

$$\lim_{n\to\infty} \mathbb{P}\left[\sup_{t\in\mathbb{R}} |\widehat{F}_n(t) - F(t)| > \varepsilon\right] = 0.$$

Also geht die Wahrscheinlichkeit, dass bei der Schätzung von F durch $\widehat{F}_n$ ein Fehler von mehr als ε entsteht, für $n \to \infty$ gegen 0.

Bemerkung 3.3.4. Für jedes $t \in \mathbb{R}$ gilt offenbar

$$0 \le |\widehat{F}_n(t) - F(t)| \le D_n.$$

Aus dem Satz von Gliwenko–Cantelli und dem Sandwich–Lemma folgt nun, dass für alle $t \in \mathbb{R}$

$$|\widehat{F}_n(t) - F(t)| \xrightarrow[n\to\infty]{f.s.} 0,$$

was exakt der Aussage von Satz 3.1.5, Teil 3 entspricht. Somit ist der Satz von Gliwenko–Cantelli stärker als Satz 3.1.5, Teil 3.

Beweis von Satz 3.3.2. Wir werden den Beweis nur unter der vereinfachenden Annahme führen, dass die Verteilungsfunktion F stetig ist. Sei also F stetig. Sei $m \in \mathbb{N}$ beliebig.

Schritt 1. Da F stetig ist und von 0 bis 1 monoton ansteigt, können wir Zahlen

$$z_1 < z_2 < \ldots < z_{m-1}$$

mit der Eigenschaft

$$F(z_1) = \frac{1}{m}, \ldots, F(z_k) = \frac{k}{m}, \ldots, F(z_{m-1}) = \frac{m-1}{m}$$

finden. Um die Notation zu vereinheitlichen, definiern wir noch $z_0 = -\infty$ und $z_m = +\infty$, so dass $F(z_0) = 0$ und $F(z_m) = 1$.

Schritt 2. Wir werden nun die Differenz zwischen $\widehat{F}_n(z)$ und $F(z)$ an einer beliebigen Stelle z durch die Differenzen an den Stellen z_k abschätzen. Für jedes $z \in \mathbb{R}$ können wir ein k mit $z \in [z_k, z_{k+1})$ finden. Dann gilt wegen der Monotonie von $\widehat{F}_n$ und F:

$$\widehat{F}_n(z) - F(z) \le \widehat{F}_n(z_{k+1}) - F(z_k) = \widehat{F}_n(z_{k+1}) - F(z_{k+1}) + \frac{1}{m}.$$

Auf der anderen Seite gilt auch

$$\widehat{F}_n(z) - F(z) \ge \widehat{F}_n(z_k) - F(z_{k+1}) = \widehat{F}_n(z_k) - F(z_k) - \frac{1}{m}.$$

SCHRITT 3. Definiere für $m \in \mathbb{N}$ und $k = 0, 1, \ldots, m$ das Ereignis

$$A_{m,k} := \left\{ \omega \in \Omega : \lim_{n\to\infty} \widehat{F}_n(z_k; \omega) = F(z_k) \right\}.$$

Dabei sei bemerkt, dass $\widehat{F}_n(z_k)$ eine Zufallsvariable ist, weshalb sie auch als Funktion des Ausgangs $\omega \in \Omega$ betrachtet werden kann. Aus Satz 3.1.5, Teil 3 folgt, dass

$$\mathbb{P}[A_{m,k}] = 1 \text{ für alle } m \in \mathbb{N}, k = 0, \ldots, m.$$

SCHRITT 4. Definiere das Ereignis $A_m := \cap_{k=0}^{m} A_{m,k}$. Da ein Schnitt von endlich vielen fast sicheren Ereignis wiederum fast sicher ist, folgt, dass

$$\mathbb{P}[A_m] = 1 \text{ für alle } m \in \mathbb{N}.$$

Da nun auch ein Schnitt von abzählbar vielen fast sicheren Ereignissen wiederum fast sicher ist, gilt auch für das Ereignis $A := \cap_{m=1}^{\infty} A_m$, dass $\mathbb{P}[A] = 1$.

SCHRITT 5. Betrachte nun einen beliebigen Ausgang $\omega \in A_m$. Dann gibt es wegen der Definition von $A_{m,k}$ ein $n(\omega, m) \in \mathbb{N}$ mit der Eigenschaft

$$|\widehat{F}_n(z_k; \omega) - F(z_k)| < \frac{1}{m} \text{ für alle } n > n(\omega, m) \text{ und } k = 0, \ldots, m.$$

Aus Schritt 2 folgt, dass

$$D_n(\omega) = \sup_{z\in\mathbb{R}} |\widehat{F}_n(z; \omega) - F(z)| \leq \frac{2}{m} \text{ für alle } \omega \in A_m \text{ und } n > n(\omega, m).$$

Betrachte nun einen beliebigen Ausgang $\omega \in A$. Somit liegt ω im Ereignis A_m, und das für alle $m \in \mathbb{N}$. Wir können nun das, was oben gezeigt wurde, auch so schreiben: Für alle $m \in \mathbb{N}$ existiert ein $n(\omega, m) \in \mathbb{N}$ so dass für alle $n > n(\omega, m)$ die Ungleichung $0 \leq D_n(\omega) < \frac{2}{m}$ gilt. Das bedeutet aber, dass

$$\lim_{n\to\infty} D_n(\omega) = 0 \text{ für alle } \omega \in A.$$

Da nun die Wahrscheinlichkeit des Ereignisses A laut Schritt 4 gleich 1 ist, erhalten wir

$$\mathbb{P}\left[\left\{ \omega \in \Omega : \lim_{n\to\infty} D_n(\omega) = 0 \right\} \right] \geq \mathbb{P}[A] = 1.$$

Somit gilt $D_n \xrightarrow[n\to\infty]{f.s.} 0$. □