Stochastik

FH Aachen – Studienstandort Aachen
28. November 2025

Prof. Dr. Florian Heinrichs

Empirische Verteilungsfunktion

Stichproben und Schätzer

FH AACHEN
UNIVERSITY OF APPLIED SCIENCES

- Seien $X_1, \ldots, X_n$ unabhängige Zufallsvariablen mit Verteilungsfunktion F, $\mu = \mathbb{E}[X_1]$ und $\sigma^2 = \mathrm{var}(X_1)$
- Wir wissen bereits:
  - Starkes Gesetz der großen Zahlen: $\bar{X}_n \xrightarrow{f.s.} \mu$
  - Folgerung aus dem Gesetz der großen Zahlen:

$$\frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2 \xrightarrow{f.s.} \sigma^2$$

- Sei $x_1, \ldots, x_n$ eine Stichprobe von F, d.h. Realisierungen von $X_1, \ldots, X_n$, dann ist $\bar{x}_n \approx \mu$ und $\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x}_n)^2 \approx \sigma^2$.
- Können wir mit Hilfe der Stichprobe auch F approximieren?

Empirische Verteilungsfunktion

Definition 59

Seien $X_1, \ldots, X_n$ Zufallsvariablen. Die Funktion

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq x\}}$$

heißt *empirische Verteilungsfunktion*.

> Die empirische Verteilungsfunktion hängt von Zufallsvariablen ab (ist also selbst zufällig)
  > Genauer: Für jedes $x \in \mathbb{R}$ ist $F_n(x)(\omega) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{X_i(\omega) \leq x\}}$ eine Zufallsvariable

> Für eine Stichprobe $x_1, \ldots, x_n$, heißt $F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{x_i \leq x\}}$ ebenfalls empirische Verteilungsfunktion

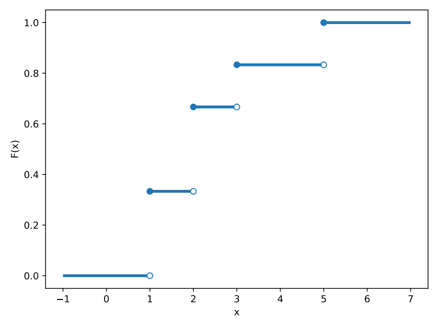
Empirische Verteilungsfunktion

Beispiel 135

Angenommen wir kennen die Analysis I Noten von sechs Studierenden:

i	1	2	3	4	5	6
x_i	2	3	2	1	5	1

Wie sieht die empirische Verteilungsfunktion der Stichprobe aus?



Empirische Verteilungsfunktion

Übung 83

Eine Person hat in den letzten Tagen $2, 3, 1, 2, 0, 1, 2$ Tassen. Kaffee getrunken. Zeichnen Sie die zugehörige empirische Verteilungsfunktion.

Empirische Verteilungsfunktion und relative Häufigkeit

- Sei $x_1, \ldots, x_n$ eine Stichprobe einer diskreten Verteilung mit Ausprägungen $(a_i)_{i \in I}$ für eine diskrete Indexmenge I.
- Die absolute Häufigkeit ist $h_i = \sum_{j=1}^n \mathbb{1}_{\{x_j = a_i\}}$.
- Die relative Häufigkeit ist $r_i = \frac{1}{n} h_i = \frac{1}{n} \sum_{j=1}^n \mathbb{1}_{\{x_j = a_i\}}$.
- Für die empirische Verteilungsfunktion F_n gilt

$$F_n(a_i) = \frac{1}{n} \sum_{j=1}^n \mathbb{1}_{\{x_j \leq a_i\}}.$$

- Falls die Ausprägungen sortiert sind, d.h. $a_1 < a_2 < \ldots$, gilt

$$F_n(a_i) = \frac{1}{n} \sum_{j=1}^n \mathbb{1}_{\{x_j \leq a_i\}} = \sum_{k=1}^i \frac{1}{n} \sum_{j=1}^n \mathbb{1}_{\{x_j = a_k\}} = \sum_{k=1}^i r_k.$$

- Die emp. VF heißt deshalb auch *relative Summenhäufigkeit*.
- Die *absolute Summenhäufigkeit* ist gegeben durch

$$nF_n(a_i) = n \sum_{k=1}^i r_k = \sum_{k=1}^i h_k.$$

Empirische Verteilungsfunktion

Beispiel 135: Fortsetzung

Angenommen wir kennen die Analysis I Noten von 6 Studierenden:

i	1	2	3	4	5	6
x_i	2	3	2	1	5	1

Was sind jeweils relative und absolute (Summen-)häufigkeit?

Ausprägung (a_i)	h_i	r_i	$F_n(a_i)$	$nF_n(a_i)$
1	2	$\frac{2}{6}$	$\frac{2}{6}$	2
2	2	$\frac{2}{6}$	$\frac{4}{6}$	4
3	1	$\frac{1}{6}$	$\frac{5}{6}$	5
4	0	$\frac{0}{6}$	$\frac{5}{6}$	5
5	1	$\frac{1}{6}$	$\frac{6}{6}$	6

Empirische Verteilungsfunktion

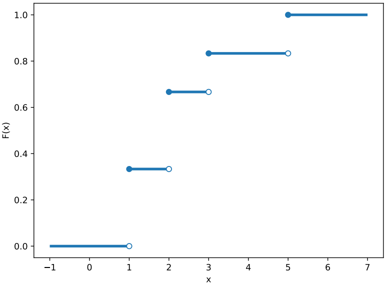
Übung 84

Eine Person hat in den letzten Tagen $2, 3, 1, 2, 0, 1, 2$ Tassen. Kaffee getrunken. Bestimmen Sie jeweils die relative und absolute (Summen-)häufigkeit.

Empirische Verteilungsfunktion

Was können wir aus der empirischen Verteilungsfunktion ablesen?

- Wie viel Prozent der Studierenden haben bestanden?
- Wie viel Prozent der Studierenden haben eine Note besser als 3?
- Allgemein: Verteilungsfunktion $F_n(x)$
- Mit welcher Note gehören Studierenden zu den besten 25%?
- Mit welcher Note sind 50% der Studierenden jeweils schlechter und besser?
- Allgemein: $q(p)$ (p-Quantil von F_n)
  $q(p) = \inf\{x \in \mathbb{R} : F_n(x) \geq p\}$



Achtung: Emp. Quantil (vgl. Def. 53) $\neq$ Quantil der emp. VF F_n:

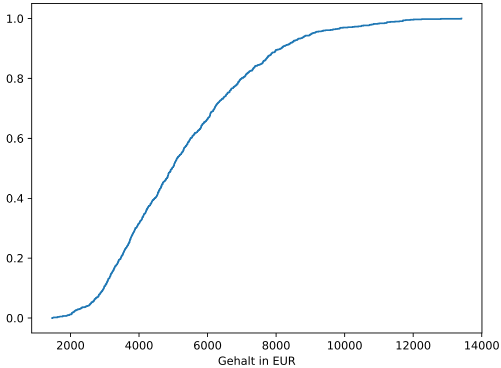
$$\inf\{x \in \mathbb{R} : F_n(x) \geq p\} \neq \begin{cases} X_{(\lfloor n \cdot p + 1 \rfloor)} & \text{falls } n \cdot p \notin \mathbb{N} \\ \frac{1}{2}\left(X_{(n \cdot p)} + X_{(n \cdot p + 1)}\right) & \text{falls } n \cdot p \in \mathbb{N} \end{cases}$$

Empirische Verteilungsfunktion

Übung 85

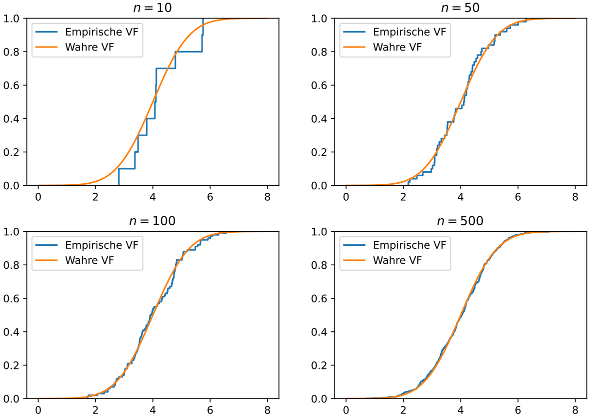
Es wurden 1000 Personen nach ihrem Gehalt befragt. Beantworten Sie mit Hilfe der empirischen VF die folgenden Fragen (approximativ):

1. Wie viel Prozent der Befragten verdienen mehr als 6000 EUR?
2. Wie viel Prozent der Befragten verdienen weniger als 3000 EUR?
3. Mit welchem Gehalt gehört eine Person zu den Top-10%-Spitzenverdienern?
4. Mit welchem Gehalt verdient eine Person jeweils mehr und weniger als 50% der Befragten?



Empirische Verteilungsfunktion

- Was passiert, wenn die Stichprobengröße n wächst?
- Sei $x_1, \ldots, x_n$ eine Stichprobe der Verteilung $\mathcal{N}(4, 1)$.



Empirische Verteilungsfunktion

- Seien $X_1, \dots, X_n$ unabhängige Zufallsvariablen mit VF F
- Für festes $x \in \mathbb{R}$ definiere $Y_i := \mathbb{1}_{\{X_i \le x\}}$, für $i = 1, \dots, n$
  - Dann gilt $Y_i \sim Ber\big(F(x)\big)$, denn $\mathbb{P}(Y_i = 1) = F(x)$
- Nach dem starken Gesetz der großen Zahlen gilt für jedes $x \in \mathbb{R}$

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{X_i \le x\}} = \frac{1}{n} \sum_{i=1}^{n} Y_i \xrightarrow{f.s.} \mathbb{E}[Y_1] = F(x)$$

  - Punktweise Konvergenz!
- Es gilt sogar $\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \xrightarrow{f.s.} 0$
  - Gleichmäßige Konvergenz!
- **Ausgangsfrage:** Wie können wir F mit Hilfe einer Stichprobe $x_1, \dots, x_n$ approximieren?

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{x_i \le x\}} \approx F(x)$$

Literatur I

Bundesministerium für Gesundheit (2024).
Infektionsradar.
https://infektionsradar.gesund.bund.de/de/covid/inzidenz.
Abgerufen: 2024-10-15.

Dehling, H. and Haupt, B. (2006).
Einführung in die Wahrscheinlichkeitstheorie und Statistik.
Springer-Verlag.

Literatur II

Dinnes, J., Sharma, P., Berhane, S., van Wyk, S., Nyaaba, N., Domen, J., Taylor, M., Cunningham, J., Davenport, C., Dittrich, S., Emperador, D., Hooft, L., Leeflang, M., McInnes, M., Spijker, R., Verbakel, J., Takwoingi, Y., Taylor-Phillips, S., Van den Bruel, A., and Deeks, J. (2022).
Rapid, point-of-care antigen tests for diagnosis of sars-cov-2 infection.
Cochrane Database of Systematic Reviews, (7).

Henze, N. et al. (1997).
Stochastik für Einsteiger, volume 4.
Springer.