5. Zufallsvariablen und Verteilungsfunktionen

Yichen Han

23. November 2023

Wiederholungsquiz

Abbildung: Quiz 8

Stichwörter

Zufallsvariable, Träger, Verteilungsfunktion, Treppenfunktion
Indikatorfunktion, Dichte
Empirische Häufigkeitsverteilung, empirische Verteilungsfunktion (ECDF)

Anmerkung

Eigenschaften einer Verteilungsfunktion:

1. $\lim_{x \to -\infty} F(x) = 0$
2. $\lim_{x \to +\infty} F(x) = 1$
3. $F(x)$ monoton steigend

Eigenschaften der Dichte:

1. $\int_{\mathbb{R}} f(x) dx = 1$
2. $\forall x \in T_X, \; f(x) \geq 0$

Eigenschaften einer ECDF:

1. $\forall x < a_1, \; F_n(x) = 0$
2. $\forall x \geq a_k, \; F_n(x) = 1$
3. $F_n(x)$ monoton wachsende Treppenfunktion
4. rechtsseitig stetig

Inhalt

r, g

Auf einer Hauptstraße regeln Ampeln an vier Kreuzungen unabhängig voneinander den Verkehr. Jede von ihnen gestattet oder verbietet die Weiterfahrt mit einer Wahrscheinlichkeit von 0.5. Wir wollen die Anzahl X der Verkehrsampeln, an denen eine gegebene Person ohne Halt vorbeifahren kann, beschreiben.

(a) Geben Sie die stochastische Grundgesamtheit Ω des hier zugrundeliegenden Laplace-Wahrscheinlichkeitsraums an.

$\Omega := \underline{\underline{\{r, g\}^4}} = \{(r, g, r, g), (r, r, g, g), \dots\}$

LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Auf einer Hauptstraße regeln Ampeln an vier Kreuzungen unabhängig voneinander den Verkehr. Jede von ihnen gestattet oder verbietet die Weiterfahrt mit einer Wahrscheinlichkeit von 0.5. Wir wollen die Anzahl X der Verkehrsampeln, an denen eine gegebene Person ohne Halt vorbeifahren kann, beschreiben.

(a) Geben Sie die stochastische Grundgesamtheit Ω des hier zugrundeliegenden Laplace-Wahrscheinlichkeitsraums an.

Wir nehmen an, die Ampeln haben zwei Ausprägungen: “r”für rot, “g”für grün.

$$\begin{aligned} \Omega &= \{r, g\}^4 = \{(r, r, r, r), (r, r, r, g), \cdots\} \\ |\Omega| &= |\{r, g\}|^4 = 2^4 = 16 \end{aligned}$$

$= |\{r, g\}^4|$

(b) Definieren Sie X und bestimmen Sie die Wahrscheinlichkeitsverteilung.

$$X : \Omega \to T_X := \{0, 1, 2, 3, 4\}$$

$$X(\omega) = \begin{cases} 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{cases}$$

$\omega \in \{(r, x, y, z) :$
$x, y, z \in \{r, g\}\}$

Aufgabe 1.2

(b) Definieren Sie X und bestimmen Sie die Wahrscheinlichkeitsverteilung.

ZV $X : \Omega \to T_X = \{0, 1, 2, 3, 4\}$

$$X(\omega) = \begin{cases} 0 & \omega \in \{(r, x, y, z) | x, y, z \in \{r, g\}\} \\ 1 & \omega \in \{(g, r, x, y) | x, y \in \{r, g\}\} \\ 2 & \omega \in \{(g, g, r, g), (g, g, r, r)\} \\ 3 & \omega \in \{(g, g, g, r)\} \\ 4 & \omega \in \{(g, g, g, g)\} \end{cases}$$

$$P(X=0) = \frac{2^3}{16} = 0{,}5$$

Aufgabe 1.2

LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

(b) Definieren Sie X und bestimmen Sie die Wahrscheinlichkeitsverteilung

ZV $X : \Omega \to T_X = \{0, 1, 2, 3, 4\}$

$$X(\omega) = \begin{cases} 0 & \omega \in \{(r, x, y, z) | x, y, z \in \{r, g\}\} \\ 1 & \omega \in \{(g, r, x, y) | x, y \in \{r, g\}\} \\ 2 & \omega \in \{(g, g, r, g), (g, g, r, r)\} \\ 3 & \omega \in \{(g, g, g, r)\} \\ 4 & \omega \in \{(g, g, g, g)\} \end{cases}$$

$$\Rightarrow f_X(x) = \begin{cases} 8/16 = 0.5 & x = 0 \\ 4/16 = 0.25 & x = 1 \\ 2/16 = 0.125 & x = 2 \\ 1/16 = 0.0625 & x = 3 \\ 1/16 = 0.0625 & x = 4 \\ 0 & \mathit{sonst} \end{cases}$$

Aufgabe 1.3

(c) Bestimmen Sie die Verteilungsfunktion dieser Zufallsvariable.

(c) Bestimmen Sie die Verteilungsfunktion dieser Zufallsvariable.

$$F(x) = P(X \leq x) \overset{\text{diskret}}{=} \sum_{i: x_i \leq x} f_X(x_i).$$

(c) Bestimmen Sie die Verteilungsfunktion dieser Zufallsvariable.

$$F(x) = P(X \leq x) \overset{\text{diskret}}{=} \sum_{i: x_i \leq x} f_X(x_i).$$

Damit gilt: $F(x) = \begin{cases} 0 & \text{für } x < 0 \\ 0.5 & \text{für } 0 \leq x < 1 \\ 0.75 & \text{für } 1 \leq x < 2 \\ 0.875 & \text{für } 2 \leq x < 3 \\ 0.9375 & \text{für } 3 \leq x < 4 \\ 1 & \text{für } x \geq 4. \end{cases}$

(d) Bestimmen Sie die Wahrscheinlichkeit dafür, dass Verkehrsteilnehmer:innen ohne Anzuhalten mindestens bis zur dritten Ampel fahren dürfen.

$1 - P(X < 3) = 1 - P(X \leq 2)$

Aufgabe 1.3

(c) Bestimmen Sie die Verteilungsfunktion dieser Zufallsvariable.

$$F(x) = P(X \leq x) \overset{\text{diskret}}{=} \sum_{i: x_i \leq x} f_X(x_i).$$

Damit gilt: $F(x) = \begin{cases} 0 & \text{für } x < 0 \\ 0.5 & \text{für } 0 \leq x < 1 \\ 0.75 & \text{für } 1 \leq x < 2 \\ 0.875 & \text{für } 2 \leq x < 3 \\ 0.9375 & \text{für } 3 \leq x < 4 \\ 1 & \text{für } x \geq 4. \end{cases}$

(d) Bestimmen Sie die Wahrscheinlichkeit dafür, dass Verkehrsteilnehmer:innen ohne Anzuhalten mindestens bis zur dritten Ampel fahren dürfen.

$$P(X \geq 3) = 1 - P(X < 3) = 1 - P(X \leq 2) = 1 - F_X(2) = 0.125$$

Welche der nachfolgenden Funktionen sind Verteilungsfunktionen stetiger Zufallsvariablen? Bestimmen Sie zu allen gültigen Verteilungsfunktionen jeweils die zugehörige Dichtefunktion.

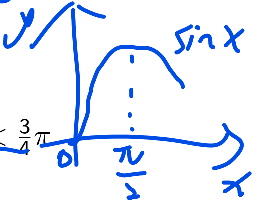
(a) $F'(x) = 2x^{-3}$ 3

$f(x) = (2x^{-3}) I_{[1,\infty[}(x)$

$$F(x) = \begin{cases} 0 & x < 1 \\ 1 - x^{-2} & x \geq 1 \end{cases}$$ 1 2

(b)

$$F(x) = \begin{cases} 0 & x < 0 \\ \log(1 + \sin(x)) & 0 \leq x < \frac{3}{4}\pi \\ 1 & x \geq \frac{3}{4}\pi \end{cases}$$ ~~3~~ 3



(c) $f(x) = F'(x) = \begin{cases} 0 & x < 0 \\ e^{-x} & x \geq 0 \end{cases}$

$= e^{-x} I_{[0,\infty[}(x)$

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - \exp(-x) & x \geq 0 \end{cases}$$ 1 2

Aufgabe 3.1

Es sei eine Zufallsvariable mit Dichte

$$f(x) = (cx - 6x^2)\ I(x \in [0, 1]).$$

(a) Bestimmen Sie die Konstante c so, dass $f(x)$ eine gültige Wahrscheinlichkeitsdichte darstellt.

$$\int_{\mathbb{R}} f(x)\,dx = 1$$

$$\forall x \in I_x \; f(x) \geq 0$$

Aufgabe 3.1

Es sei eine Zufallsvariable mit Dichte

$$f(x) = (cx - 6x^2)\ I(x \in [0,1]).$$

(a) Bestimmen Sie die Konstante c so, dass $f(x)$ eine gültige Wahrscheinlichkeitsdichte darstellt.

Zu zeigen: $\int_{\mathbb{R}} f(x)dx = 1$, und $f(x) \geq 0$

Aufgabe 3.1

Es sei eine Zufallsvariable mit Dichte

$$f(x) = (cx - 6x^2)\ I(x \in [0,1]).$$

(a) Bestimmen Sie die Konstante c so, dass $f(x)$ eine gültige Wahrscheinlichkeitsdichte darstellt.

Zu zeigen: $\int_{\mathbb{R}} f(x)dx = 1$, und $f(x) \geq 0$

$$\int_R (cx - 6x^2)\ I_{[0,1]}(x)\, dx$$

$$\begin{aligned}
\int_0^1 cx - 6x^2 dx &= c\int_0^1 x dx - 6\int_0^1 x^2 dx \\
&= \frac{1}{2}c\left[x^2\right]_0^1 - 6\left[\frac{1}{3}x^3\right]_0^1 \\
&= \frac{1}{2}c - 2 \overset{!}{=} 1 \\
\Rightarrow \quad & c = 6
\end{aligned}$$

$$f(x) = 6x - 6x^2 = 6x(1-x) \geq 0 \text{ für } x \in [0,1] \qquad \surd$$

(b) Ermitteln Sie die zugehörige Verteilungsfunktion $F(x)$.

(b) Ermitteln Sie die zugehörige Verteilungsfunktion $F(x)$.

$$
\begin{aligned}
F(x) &= \int_{-\infty}^{x} f(t)dt \\
&= \int_{-\infty}^{x} (6t - 6t^2) I_{[0,1]}(x)\ dt \\
&= \int_{0}^{x} (6t - 6t^2) dt I_{[0,1]}(x) \\
&= \left(\left[\frac{6}{2} t^2 - \frac{6}{3} t^3 \right]_0^x \right) I_{[0,1]}(x) \\
&= (3x^2 - 2x^3)\ I_{[0,1]}(x) + I_{]1,\infty[}(x)
\end{aligned}
$$

äq.
$$
\begin{cases} 0, & x < 0 \\ 3x^2 - 2x^3, & x \in [0,1] \\ 1, & x > 1 \end{cases}
$$

(c) Mit welcher W'keit werden in einer Woche

i) mehr als 0.8 Hektoliter verbraucht?
ii) genau 0.5 Hektoliter verbraucht?
iii) zwischen 0.5 und 0.8 Hektoliter verbraucht?

$1 - F(0.8)$

$P(X > 0.8) = 1 - P(X \leq 0.8)$

$P(X = 0.5) \hat{=} 0$

$P(0.5 \leq X \leq 0.8)$

$= F(0.8) - F(0.5)$

(c) Mit welcher W'keit werden in einer Woche

i) mehr als 0.8 Hektoliter verbraucht?

ii) genau 0.5 Hektoliter verbraucht?

iii) zwischen 0.5 und 0.8 Hektoliter verbraucht?

i)

$$\begin{aligned} P(X > 0.8) &= 1 - F(0.8) \\ &= 1 - 0.896 = 0.104 \end{aligned}$$

ii)

$$P(X = 0.5) = P(0.5 \leq x \leq 0.5) = F(0.5) - F(0.5) = 0$$

iii)

$$\begin{aligned} P(0.5 \leq X \leq 0.8) &= F(0.8) - F(0.5) \\ &= 0.896 - 0.5 = 0.396 \end{aligned}$$

(d) Wie hoch ist der Bierverbrauch pro Woche, der mit einer Wahrscheinlichkeit von 0.5 überschritten wird?

Aufgabe 3.4

(d) Wie hoch ist der Bierverbrauch pro Woche, der mit einer Wahrscheinlichkeit von 0.5 überschritten wird?

$$F(x) = 3x^2 - 2x^3 \stackrel{!}{=} 0.5$$

Numerische (mit dem Rechner) Lösung!

(d) Wie hoch ist der Bierverbrauch pro Woche, der mit einer Wahrscheinlichkeit von 0.5 überschritten wird?

$$F(x) = 3x^2 - 2x^3 \stackrel{!}{=} 0.5$$

Numerische (mit dem Rechner) Lösung!

```
library(rootSolve)
# Define the function representing the equation
equation <- function(x) {
  return(3 * x^2 - 2 * x^3 - 0.5)
}
# Find roots of the equation
solutions <- uniroot.all(equation, c(-1, 2))
print(solutions)
```

Output:

```
[1]  0.5000000   -0.3658408   1.3658408
```

Aufgabe 4.1

Der Besitzer des Kinos *Cinemania* macht sich Gedanken über die Wirtschaftlichkeit seines Hauses. An 100 Tagen zählt er daher die Anzahl der Besucher.

a_j	41	42	43	44	45	46	47	48	49	50	51
h_j	1	9	13	13	20	15	10	7	5	4	3

(a) Wie heißt das untersuchte Merkmal und wie ist es skaliert?

(b) Berechnen Sie die relativen und kumulierten relativen Häufigkeiten.

Der Besitzer des Kinos *Cinemania* macht sich Gedanken über die Wirtschaftlichkeit seines Hauses. An 100 Tagen zählt er daher die Anzahl der Besucher.

a_j	41	42	43	44	45	46	47	48	49	50	51
h_j	1	9	13	13	20	15	10	7	5	4	3

(a) Wie heißt das untersuchte Merkmal und wie ist es skaliert?

(b) Berechnen Sie die relativen und kumulierten relativen Häufigkeiten.

X: Besucherzahl, absolutskaliert.

Der Besitzer des Kinos *Cinemania* macht sich Gedanken über die Wirtschaftlichkeit seines Hauses. An 100 Tagen zählt er daher die Anzahl der Besucher.

a_j	41	42	43	44	45	46	47	48	49	50	51
h_j	1	9	13	13	20	15	10	7	5	4	3

(a) Wie heißt das untersuchte Merkmal und wie ist es skaliert?

(b) Berechnen Sie die relativen und kumulierten relativen Häufigkeiten.

X: Besucherzahl, absolutskaliert.

- Relative Häufigkeit für a_j: $f_j = f(a_j) = \frac{h_j}{n}$ mit $n = 100$
- Kumulierte relative Häufigkeit für a_j: $F_j = \sum_{l=1}^{j} f_l$

Aufgabe 4.1

Der Besitzer des Kinos *Cinemania* macht sich Gedanken über die Wirtschaftlichkeit seines Hauses. An 100 Tagen zählt er daher die Anzahl der Besucher.

a_j	41	42	43	44	45	46	47	48	49	50	51
h_j	1	9	13	13	20	15	10	7	5	4	3

(a) Wie heißt das untersuchte Merkmal und wie ist es skaliert?

(b) Berechnen Sie die relativen und kumulierten relativen Häufigkeiten.

X: Besucherzahl, absolutskaliert.

- Relative Häufigkeit für a_j: $f_j = f(a_j) = \frac{h_j}{n}$ mit $n = 100$
- Kumulierte relative Häufigkeit für a_j: $F_j = \sum_{l=1}^{j} f_l$

51

a_j	41	42	43	44	45	46	47	48	49	50	5
h_j	1	9	13	13	20	15	10	7	5	4	3
f_j	0.01	0.09	0.13	0.13	0.20	0.15	0.10	0.07	0.05	0.04	0.03
F_j	0.01	0.10	0.23	0.36	0.56	0.71	0.81	0.88	0.93	0.97	1

Aufgabe 4.2

(c) Zeichnen Sie die empirische Verteilungsfunktion.

42 42 42 ··· 9mal

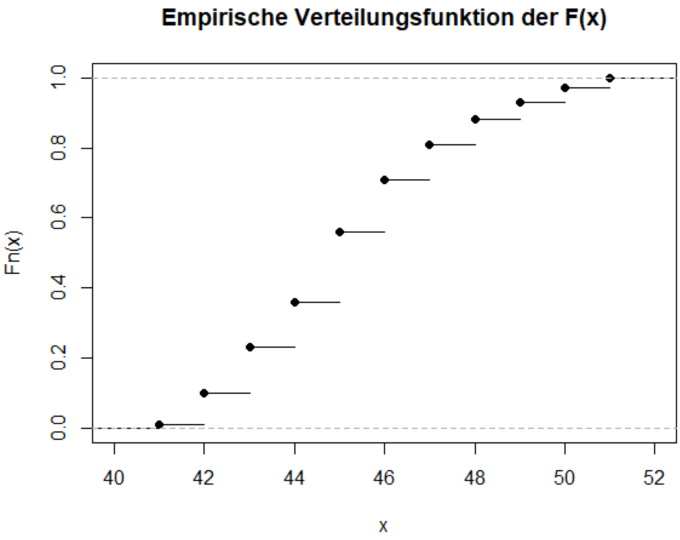
(c) Zeichnen Sie die empirische Verteilungsfunktion.

```
a_j <- c(41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51)
h_j <- c(1, 9, 13, 13, 20, 15, 10, 7, 5, 4, 3)

data <- data.frame(a_j, h_j)

sample <- rep(data$a_j, times = data$h_j)

plot(ecdf(sample),
    main = "Empirische Verteilungsfunktion der F(x)")
```


Empirische Verteilungsfunktion der F(x)
Fn(x)
0.0
0.2
0.4
0.6
0.8
1.0
40
42
44
46
48
50
52
x

Aufgabe 4.3

(d) Es wird ein Einheitspreis von 6 Euro erhoben. Wie groß ist der Anteil der Tage, an denen der Kinobesitzer weniger als 270 Euro einnimmt?

$$< \frac{270}{6} = 45$$

$$F(44)$$

Aufgabe 4.3

(d) Es wird ein Einheitspreis von 6 Euro erhoben. Wie groß ist der Anteil der Tage, an denen der Kinobesitzer weniger als 270 Euro einnimmt?

$$N = \frac{260}{6} = 45$$

$$\Rightarrow \text{max. } 44 \text{ Besucher}$$

$$F(44) = \frac{1+9+13+13}{100} = 36\%$$

(e) Wie groß ist der Anteil der Tage, an denen mindestens 45, aber maximal 50 Besucher kommen?

$P(45 \leq X \leq 50)$

$F(50) - F(45)$

(d) Es wird ein Einheitspreis von 6 Euro erhoben. Wie groß ist der Anteil der Tage, an denen der Kinobesitzer weniger als 270 Euro einnimmt?

$$
\begin{aligned}
N &= \frac{260}{6} = 45 \\
&\Rightarrow \text{max. 44 Besucher} \\
F(44) &= \frac{1+9+13+13}{100} = 36\%
\end{aligned}
$$

(e) Wie groß ist der Anteil der Tage, an denen mindestens 45, aber maximal 50 Besucher kommen?

$$P(45 \leq x \leq 50) = F(50) - F(45) = 0.97 - 0.36 = 61\%$$

Nächste Woche: **Statistische Grafiken**
– ohne Zweifel das wichtigste Thema des Semesters!

Vorbereitung auf das Blatt 6 notwendig! vsl. keine Zeit für Wiederholung.