3,120 Methoden: Statistik

Übungsblatt 3: Mehrdimensionale Wahrscheinlichkeitsverteilungen

Mathis Mörke
Michael Schürle
und Alois Weigand

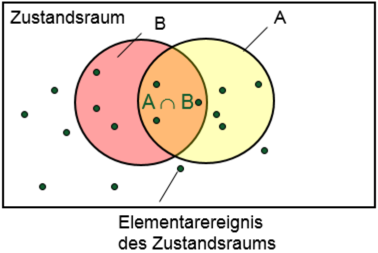
Universität St.Gallen (HSG)

Herbstsemester 2024

- **Mehrdimensionale Verteilungen**
- **Literatur**
  - Schira, Statistische Methoden der VWL und BWL
  - 4. Auflage
  - Kapitel 3 und 10

Refresher

- **Mehrdimensionale Verteilungen**
  - **aus Übung 2:** Ereignisse A und B treten gemeinsam ein $\Rightarrow P(A \cap B)$



  - **Konzept angewandt auf zwei Zufallsvariablen**

Refresher

- **Mehrdimensionale Verteilungen**
  - **aus Übung 2:**
  - für stochastisch unabhängige Ereignisse
    $P(A \cap B) = P(A) \cdot P(B)$
  - für stochastisch abhängige Ereignisse
    $P(A \cap B) = P(A) \cdot P(B|A) = P(B) \cdot P(A|B)$
  - Bedingte Wahrscheinlichkeit (B gegeben A)
    $P(B|A) = \frac{P(A \cap B)}{P(A)}$ für $P(A) > 0$
  - **Konzepte angewandt auf zwei Zufallsvariablen**

Refresher

- **Mehrdimensionale Verteilungen**
  - gemeinsame Verteilungen
  - marginale Verteilung (Randverteilung)
  - bedingte Verteilung
  - ***später: Kovarianz, Regressionsanalyse***
    $\Rightarrow$ ***Übung 6***

Refresher

- **Mehrdimensionale Verteilungen**
  - Zweidimensionale Zufallsvariable (X, Y)
  - **Gemeinsame Verteilung**
    $\rightarrow$ analog zur Verteilungsfunktion
    $f(x_i, y_j) = P(X = x_i \cap Y = y_j)$
  - **Marginale Verteilung (Randverteilung)**
    $\rightarrow$ Verteilungen von X bzw. Y, wenn die jeweils andere Zufallsvariable unberücksichtigt bleibt
    $\Rightarrow$ Zeilen- und Spaltensummen

	y_1	y_2	y_3	y_4	$f(x_i)$
x_1	0.02	0.06	0.02	0.05	0.15
x_2	0.1	0.1	0.2	0.1	0.5
x_3	0.05	0.1	0.15	0.05	0.35
$f(y_j)$	0.17	0.26	0.37	0.2	1

Refresher

- **Mehrdimensionale Verteilungen**

	y_1	y_2	y_3	y_4	$f(x_i)$
x_1	0.02	0.06	0.02	0.05	0.15
x_2	0.1	0.1	0.2	0.1	0.5
x_3	0.05	0.1	0.15	0.05	0.35
$f(y_j)$	0.17	0.26	0.37	0.2	1

  - **Erwartungswert und Varianz**
  - $E(X) = \mu_x = \sum_i x_i f(x_i)$
    $E(Y) = \mu_Y = \sum_j y_j f(y_j)$
  - $Var(X) = \sum_i (x_i - \mu_X)^2 f(x_i)$
    $Var(Y) = \sum_j (y_j - \mu_Y)^2 f(y_j)$

Refresher

- **Mehrdimensionale Verteilungen**
  - ein kleines Beispiel

	$x_1 = 1$	$x_2 = 2$	
$y_1 = 2$	f(x_1,y_1)=0.4	f(x_2,y_1)=0.3	$f(y_1) = 0.7$
$y_2 = 4$	f(x_2,y_1)=0.2	f(x_2,y_2)=0.1	$f(y_2) = 0.3$
	$f(x_1) = 0.6$	$f(x_2) = 0.4$	1

  - $E(X) = \mu_x = 0.6 \cdot 1 + 0.4 \cdot 2 = 1.4$
  - $E(Y) = \mu_y = 0.7 \cdot 2 + 0.3 \cdot 4 = 2.6$

  - $Var(X) = (1 - 1.4)^2 \cdot 0.6 + (2 - 1.4)^2 \cdot 0.4 = 0.24$
  - $Var(Y) = (2 - 2.6)^2 \cdot 0.7 + (4 - 2.6)^2 \cdot 0.3 = 0.84$

Refresher

- **Mehrdimensionale Verteilungen**

	y_1	y_2	y_3	y_4	$f(x_i)$
x_1	0.02	0.06	0.02	0.05	0.15
x_2	0.1	0.1	0.2	0.1	0.5
x_3	0.05	0.1	0.15	0.05	0.35
$f(y_j)$	0.17	0.26	0.37	0.2	1

  - Analog zur bedingten Wahrscheinlichkeit
    $\Rightarrow P(X = x_i | Y = y_j) = \frac{P(X=x_i \cap Y=y_j)}{P(Y=y_j)}$

  - **Bedingte Verteilungen**
    $\Rightarrow$ gemeinsame Verteilung geteilt durch Randverteilung
    $f(x_i | Y = y_j) = \frac{f(x_i, y_j)}{f(y_j)}$

    $f(y_j | X = x_i) = \frac{f(x_i, y_j)}{f(x_i)}$

Refresher

- **Mehrdimensionale Verteilungen**

	y_1	y_2	y_3	y_4	$f(x_i)$
x_1	0.02	0.06	0.02	0.05	0.15
x_2	0.1	0.1	0.2	0.1	0.5
x_3	0.05	0.1	0.15	0.05	0.35
$f(y_j)$	0.17	0.26	0.37	0.2	1

  - Bedingte Verteilungen

    $\Rightarrow$ Spalte von $Y = y_4$ als Population

  - für y_4: $f(x_1|y_4) = \frac{0.05}{0.2} = 0.25$, $f(x_2|y_4) = \frac{0.1}{0.2} = 0.5$, $f(x_3|y_4) = \frac{0.05}{0.2} = 0.25$

Refresher

- **Mehrdimensionale Verteilungen**

	y_1	y_2	y_3	y_4	f(x_i)
x_1	0.02	0.06	0.02	0.05	0.15
x_2	0.1	0.1	0.2	0.1	0.5
x_3	0.05	0.1	0.15	0.05	0.35
f(y_j)	0.17	0.26	0.37	0.2	1

  - Bedingte Verteilungen

    $\Rightarrow$ Zeile von $X = x_2$ als Population

  - für x_2: $f(y_1|x_2) = \frac{0.1}{0.5} = 0.2$, $f(y_2|x_2) = \frac{0.1}{0.5} = 0.2$, $f(y_3|x_2) = \frac{0.2}{0.5} = 0.4$, $f(y_4|x_2) = \frac{0.1}{0.5} = 0.2$

Refresher

- **Mehrdimensionale Verteilungen**
  - **Stochastische Abhängigkeit**
  - analog zu Wahrscheinlichkeiten:
    Ereignisse A und B sind stochastisch abhängig, wenn
    $P(A|B) \neq P(A|\overline{B})$

Refresher

- **Mehrdimensionale Verteilungen**
  - **Stochastische Abhängigkeit**

    (i) Test über bedingte Verteilungen
  - X und Y sind **stochastisch abhängig**, wenn die bedingten Verteilungen für verschiedene Bedingungen y_1 und y_2 verschieden sind $\Rightarrow f(x|y_1) \neq f(x|y_2)$
    $\Rightarrow$ Verteilung von X hängt davon ab, welchen Wert Y annimmt

Refresher

- **Mehrdimensionale Verteilungen**

  (ii) Test über Randverteilungen

  - Zufallsvariablen X und Y sind **stochastisch unabhängig**, wenn die gemeinsame Verteilungsfunktion gleich dem Produkt der beiden Randverteilungen ist
    $f(x, y) = f(x)f(y)$

  - **Warum?**
    für unabhängige Variablen gilt:
    $f(x|y) = \frac{f(x,y)}{f(y)} = \frac{f(x)f(y)}{f(y)} = f(x)$

    $f(y|x) = \frac{f(x,y)}{f(x)} = \frac{f(x)f(y)}{f(x)} = f(y)$

    $\Rightarrow$ alle bedingte Verteilungen einer Zufallsvariablen sind gleich

Aufgabe 10

In einer Bank kommen 80% der männlichen Kreditkunden ihren Kreditverpflichtungen pünktlich nach, 15% schleppend nach, und bei 5% muss die Bank den Kredit abschreiben. Bei den weiblichen Kreditkunden sind die entsprechenden Zahlen 85%, 10% und 5%. Von den Kreditkunden der Bank sind 70% männlich.

(a) Wie gross ist die Wahrscheinlichkeit, dass eine zufällig aus dem Kreis aller Kreditkunden ausgewählte Person ihren Kreditverpflichtungen pünktlich nachkommt?

(b) Wie gross ist die Wahrscheinlichkeit, dass eine zufällig aus dem Kreis aller Kreditkunden ausgewählte Person weiblich ist, falls die Person ihren Kreditverpflichtungen nur schleppend nachkommt?

(c) Sind die Ereignisse Kunde ist männlich und Kunde zahlt pünktlich stochastisch unabhängig?

Aufgabe 10

- **Kreuztabelle**
- Zahlungsmoral x_i mit p(ünktlich), s(chleppend), n(icht)
- Geschlecht y_i mit Ausprägungen: *Mann*, *Frau*

	Mann	Frau	f(x)
p	56	25.5	81.5
s	10.5	3	13.5
n	3.5	1.5	5.0
f(y)	70	30	100

  - 70% sind männlich:
    $70 \cdot \frac{80}{100} = 56,\ 70 \cdot \frac{15}{100} = 10.5,\ 70 \cdot \frac{5}{100} = 3.5$
  - 30% sind weiblich:
    $30 \cdot \frac{85}{100} = 25.5,\ 30 \cdot \frac{10}{100} = 3,\ 30 \cdot \frac{5}{100} = 1.5$

Aufgabe 10

(a) Wie gross ist die Wahrscheinlichkeit, dass eine zufällig aus dem Kreis aller Kreditkunden ausgewählte Person ihren Kreditverpflichtungen pünktlich nachkommt?

- Unabhängig vom Geschlecht!
- Zahlungsmoral: p(ünktlich), s(chleppend), n(icht)

	Mann	Frau	
p	56	25.5	81.5
s	10.5	3	13.5
n	3.5	1.5	5.0
	70	30	100

- $f(p) = \frac{56}{100} + \frac{25.5}{100} = 81.5\%$

Aufgabe 10

(b) Wie gross ist die Wahrscheinlichkeit, dass eine zufällig aus dem Kreis aller Kreditkunden ausgewählte Person weiblich ist, falls die Person ihren Kreditverpflichtungen nur schleppend nachkommt?

- Zahlungsmoral: p(ünktlich), s(chleppend), n(icht)

	Mann	Frau	
p	56	25.5	81.5
s	10.5	3	13.5
n	3.5	1.5	5.0
	70	30	100

- $f(w|s) = \frac{f(w,s)}{f(s)} = \frac{3/100}{13.5/100} = \frac{3}{13.5} = 22.22\%$

Aufgabe 10

(c) Sind die Ereignisse Kunde ist männlich und Kunde zahlt pünktlich stochastisch unabhängig?

- Zahlungsmoral: p(ünktlich), s(chleppend), n(icht)

	Mann	Frau	
p	56	25.5	81.5
s	10.5	3	13.5
n	3.5	1.5	5.0
	70	30	

  - stochastisch unabhängig: $f(x, y) = f(x)f(y)$
  - $f(M) = \frac{56}{100} + \frac{10.5}{100} + \frac{3.5}{100} = 70\%$
  - $f(p) = \frac{56}{100} + \frac{25.5}{100} = 81.5\%$

    $\Rightarrow \frac{70}{100} \cdot \frac{81.5}{100} = 0.5705 \neq \frac{56}{100} = 0.56$
    $\Rightarrow$ scheinen stochastisch abhängig

Aufgabe 11

Die folgenden statistischen Daten bezüglich des Wahlverhalten (Anzahl der Wahlstimmen in Mio.) in einem Bundesland werden in einer Stichprobe bereitgestellt:

Partei	weibliche Wähler	männliche Wähler
Die Linke	10.4	12.8
Die Konservativen	11.3	13.9
Die Liberalen	1.6	2.5
Die Grünen	1.4	1.8

(a) Bilden Sie sowohl Zeilen- als auch Spaltensummen und berechnen Sie die relativen Häufigkeiten.

(b) Bestimmen Sie die folgenden bedingten Häufigkeiten: Wie viele männliche Wähler haben für die Linke gestimmt (Bedingung: männlich)? Wie viele konservative Wähler sind weiblich (Bedingung: konservative Wähler)?

(c) Untersuchen Sie, ob die Variablen stochastisch unabhängig sind.

Aufgabe 11

- Bilden Sie sowohl Zeilen- als auch Spaltensummen und berechnen Sie die relativen Häufigkeiten.
  (relative Häufigkeiten gerundet)

Geschlecht / Partei	weiblich absolut	weiblich (relativ)	männlich absolut	männlich (relativ)	n_i	(f_i)
Die Linke	10.4	(0.19)	12.8	(0.23)	23.2	(0.42)
Die Konservativen	11.3	(0.20)	13.9	(0.25)	25.2	(0.45)
Die Liberalen	1.6	(0.03)	2.5	(0.04)	4.1	(0.07)
Die Grünen	1.4	(0.03)	1.8	(0.03)	3.2	(0.06)
n_j (f_j)	24.7	(0.45)	31	(0.55)	55.7	(1.00)

Aufgabe 11

- Wie viele männliche Wähler haben für die Linke gestimmt (Bedingung: männlich)?

Geschlecht	weiblich		männlich		n_i	(f_i)
Partei	absolut	(relativ)	absolut	(relativ)		
Die Linke	10.4	(0.19)	12.8	(0.23)	23.2	(0.42)
Die Konservativen	11.3	(0.20)	13.9	(0.25)	25.2	(0.45)
Die Liberalen	1.6	(0.03)	2.5	(0.04)	4.1	(0.07)
Die Grünen	1.4	(0.03)	1.8	(0.03)	3.2	(0.06)
$n_j \quad (f_j)$	24.7	(0.45)	31	(0.55)	55.7	(1.00)

  - $f(Linke|männlich) = \frac{f(männlich \cap Linke)}{f(männlich)}$
  - $f(Linke|männlich) = \dfrac{12.8}{31} = 0.413$

Aufgabe 11

- Wie viele konservative Wähler sind weiblich (Bedingung: konservative Wähler)?

Geschlecht Partei	weiblich absolut	 (relativ)	männlich absolut	 (relativ)	n_i	(f_i)
Die Linke	10.4	(0.19)	12.8	(0.23)	23.2	(0.42)
Die Konservativen	11.3	(0.20)	13.9	(0.25)	25.2	(0.45)
Die Liberalen	1.6	(0.03)	2.5	(0.04)	4.1	(0.07)
Die Grünen	1.4	(0.03)	1.8	(0.03)	3.2	(0.06)
n_j (f_j)	24.7	(0.45)	31	(0.55)	55.7	(1.00)

- $f(weiblich|konservativ) = \dfrac{f(weiblich \cap konservativ)}{f(konservativ)}$
- $f(weiblich|konservativ) = \dfrac{11.3}{25.2} = 0.448$

Aufgabe 11

- Untersuchen Sie, ob die Variablen stochastisch unabhängig sind.

Geschlecht / Partei	weiblich absolut	weiblich (relativ)	männlich absolut	männlich (relativ)	n_i	(f_i)
Die Linke	10.4	(0.19)	12.8	(0.23)	23.2	(0.42)
Die Konservativen	11.3	(0.20)	13.9	(0.25)	25.2	(0.45)
Die Liberalen	1.6	(0.03)	2.5	(0.04)	4.1	(0.07)
Die Grünen	1.4	(0.03)	1.8	(0.03)	3.2	(0.06)
n_j (f_j)	24.7	(0.45)	31	(0.55)	55.7	(1.00)

- Stochastische Unabhängigkeit: über bedingte Verteilungen

  Untersuchen Sie, ob

  $$f(DieLinke|maennlich) = \frac{0.23}{0.55} = 0.418 = 0.42 = f(DieLinke)$$

  $$f(DieLinke|weiblich) = \frac{0.19}{0.45} = 0.422 = 0.42 = f(DieLinke)$$

  $\Rightarrow$ Die Werte sind nahe beieinander $\Rightarrow$ nahezu unabhängig.

Aufgabe 11

- Untersuchen Sie, ob die Variablen stochastisch unabhängig sind.

Geschlecht	weiblich		männlich		n_i	(f_i)
Partei	absolut	(relativ)	absolut	(relativ)		
Die Linke	10.4	(0.19)	12.8	(0.23)	23.2	(0.42)
Die Konservativen	11.3	(0.20)	13.9	(0.25)	25.2	(0.45)
Die Liberalen	1.6	(0.03)	2.5	(0.04)	4.1	(0.07)
Die Grünen	1.4	(0.03)	1.8	(0.03)	3.2	(0.06)
n_j (f_j)	24.7	(0.45)	31	(0.55)	55.7	(1.00)

  - Alternativ: Test über Randverteilungen
  - z.B., Die Liberalen (0.07) und weiblich (0.45): $0.07 \cdot 0.45 = 0.0312$
  - z.B., Die Linke (0.42) und männlich (0.55): $0.42 \cdot 0.55 = 0.231$

  $\Rightarrow$ Die Werte sind nahe bei f_{ij} $\Rightarrow$ nahezu unabhängig.