Fachhochschule Südwestfalen
University of Applied Sciences

Statistik

CH.7 - Anwendungen Wahrscheinlichkeitsrechnung

2024 || Prof. Dr. Buchwitz, Sommer, Henke

Wir geben Impulse

Lernziele

- Verinnerlichen der Wahrscheinlichkeitsrechnung.
- Verknüpfen der Wahrscheinlichkeitsrechnung mit Kombinatorik.
- Verstehen, dass das Rechnen mit Wahrscheinlichkeiten zu präzisieren Aussagen führt als sich auf die eigenen Intuition zu verlassen.

Anwendungen der Wahrscheinlichkeitsrechnung

Bei der Anwendung der Wahrscheinlichkeitsrechnung gibt es zahlreiche Sonderfälle und Überraschungen. Wir diskutieren im Folgenden diese Themen:

- Urnenmodell
- Entscheidungsbäume
- Simpson Paradoxon
- Geburtstagsproblem

Inhaltsübersicht

Urnenmodell: Überblick

Gegeben sei eine Urne mit N Kugeln, davon W weiße und S schwarze ($W + S = N$). Aus der Urne werden n Kugeln ($1 \leq n \leq N$) nacheinander **ohne Zurücklegen** gezogen. Die Wahrscheinlichkeit für das Ereignis A, dass sich unter den n Kugeln genau w weiße und s schwarze befinden, lässt sich folgendermaßen berechnen:

$$P(A) = \frac{\binom{W}{w}\binom{S}{s}}{\binom{N}{n}}$$

Hypergeometrische Verteilung

ohne Zurücklegen

Urnenmodell: Beispiel

- Gegeben sei eine Urne mit $N = 11$ Kugeln, davon $W = 5$ weiße und $S = 6$ schwarze $(5 + 6 = 11)$.

Aus der Urne werden n Kugeln $(1 \leq n \leq 11)$ nacheinander ohne Zurücklegen gezogen. Die Wahrscheinlichkeit für das Ereignis A, dass sich unter den n Kugeln w weiße und s schwarze befinden, lässt sich folgendermaßen berechnen:

$$P(A) = \frac{\binom{W=5}{w}\binom{S=6}{s}}{\binom{N=11}{n}}$$

Noch konkreter: Aus der Urne werden 5 Kugeln ohne Zurücklegen gezogen. Die Wahrscheinlichkeit für das Ereignis A, dass sich unter den 5 Kugeln 2 weiße und 3 schwarze befinden, lässt sich folgendermaßen berechnen:

$$P(A) = \frac{\binom{W=5}{w=2}\binom{S=6}{s=3}}{\binom{N=11}{n=5}}$$

Urnenmodell: Simulation

```r
balls <- factor(c(rep("w",5), rep("s", 6)))
s <- sample(balls, size=1, replace=F)

balls <- factor(c(rep("w",5), rep("s", 6)))
s <- sample(balls, size=5, replace=F)

is.match <- function(x, s, w){
  tab <- table(x)
  return(tab[1] == s && tab[2] == w)
}

s
```

```
## [1] s s w s w
## Levels: s w
```

```r
is.match(s, w=2, s=3)
```

```
## [1] TRUE
```

Urnenmodell: Simulation

```r
set.seed(1)

# Lösung per Simulation
x <- replicate(n = 10000,
               expr = is.match(sample(balls, size=5, replace=F), s=3, w=2))
mean(x)
```

```
## [1] 0.4334
```

```r
# Lösung per Rechnung
(choose(5,2) * choose(6, 3))/choose(11,5)
```

```
## [1] 0.4329
```

Urnenmodell: Anwendungsbeispiel

Ein Unternehmen erhält wiederholt Lieferungen von 800 Flaschen zur Verpackung von flüssigem Waschmittel. Mit dem Lieferanten ist vereinbart, dass Lieferungen mit mehr als 2% fehlerhaften Flaschen zurückgewiesen werden dürfen. Um zu entscheiden, ob es eine Lieferung zurückweist, verfährt das Unternehmen nach folgender Regel: Der Lieferung werden 50 Flaschen zufällig entnommen und geprüft. Die Lieferung wird zurückgewiesen, wenn mehr als eine Flasche nicht dem vereinbarten Qualitätsstandard entspricht.

- Wie groß ist die Wahrscheinlichkeit, dass eine gerade noch zulässige Lieferung, d.h. mit genau 2% fehlerhaften Flaschen, zurückgewiesen wird?

Urnenmodell: Anwendungsbeispiel

Urnenmodell: Anwendungsbeispiel

```r
# Flasche Fehlerfrei = TRUE, Flasche Fehlerhaft = FALSE
f <- c(rep(TRUE, 784), rep(FALSE, 16))
x <- replicate(n=10000, expr=sum(sample(f, size=50, replace = FALSE)))
p_0 <- mean(x == 50)
p_1 <- mean(x == 49)
1 - p_0 - p_1
```

```
## [1] 0.2644
```

```r
# Lösung per Binomialkoeffizient
p_0 <- (choose(16,0)*choose(784,50))/choose(800,50)
p_1 <- (choose(16,1)*choose(784,49))/choose(800,50)
1 - p_0 - p_1
```

```
## [1] 0.2639
```

Inhaltsübersicht

Entscheidungsbäume: Überblick

- Zahlreiche von Unsicherheit geprägte Sachverhalte können als Entscheidungsbäume dargestellt werden. **Beispiel:** Würfeln mit einem und mit zwei Würfeln.

Entscheidungsbäume: Überblick

Entscheidungsbäume

- Unternehmen müssen täglich Entscheidungen treffen, z.B.:
  - über den Standort eines neuen Werkes
  - zwischen mehreren unterschiedlichen Anlageformen
  - über Investitionen in neue Maschinen etc.
- Was ist dabei zu beachten:
  - Nicht alle Informationen, die der Entscheider gerne zur Verfügung hätte, sind bekannt
  - Das Unternehmen ist darauf angewiesen, Wahrscheinlichkeiten für das Eintreffen der verschiedenen Ereignisse abzuschätzen
  - Basierend auf diesen Wahrscheinlichkeiten werden Entscheidungen getroffen
  - Während des Entscheidungsprozesses kann es möglich sein, dass das Unternehmen an Zusatzinformationen gelangt
  - Durch diese Zusatzinformationen verändern sich die Wahrscheinlichkeiten für das Eintreffen der verschiedenen Ereignisse.

Entscheidungsbäume: Beispiel

Entscheidungsbäume: Beispiel

Entscheidungsbäume: Beispiel

Inhaltsübersicht

Simpsons Paradoxon

```r
UCBAdmissions[ , ,1] # Department A
```

```
##           Gender
## Admit      Male Female
##   Admitted  512     89
##   Rejected  313     19
```

Simpsons Paradoxon

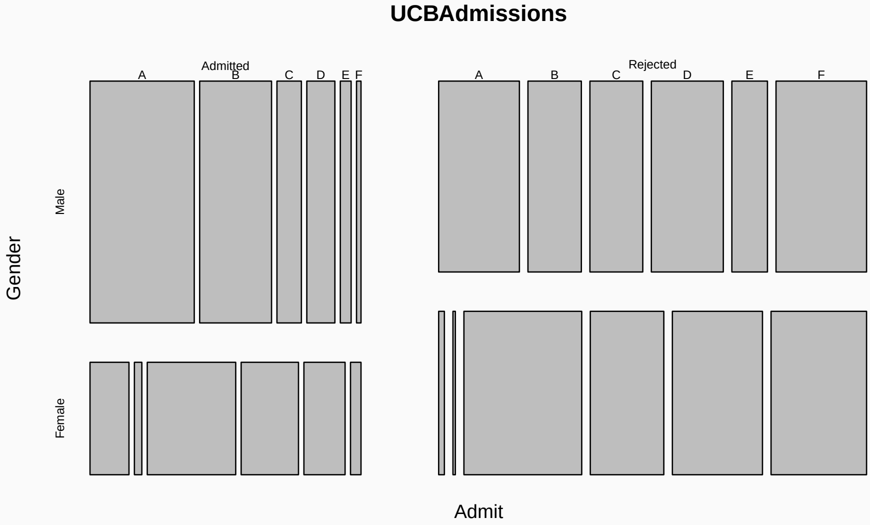
- Frauen haben eine niedrigere Zulassunsgsquote

```
apply(UCBAdmissions,c(1,2),sum)
```

```
##           Gender
## Admit      Male Female
##   Admitted 1198    557
##   Rejected 1493   1278
```

Simpsons Paradoxon

```r
plot(UCBAdmissions)
```



Simpsons Paradoxon

Bewertung verschiedener Gruppen fällt scheinbar unterschiedlich aus, je nachdem, ob man die Ergebnisse der Gruppen kombiniert oder nicht.

Inhaltsübersicht

Geburtstagsproblem: Überblick

- **Frage:** Wie groß ist die Wahrscheinlichkeit, dass in einer Statistikvorlesung mit k = 100 Studierenden (mindestens) zwei Personen am gleichen Tag Geburtstag haben (Ereignis A)?

Geburtstagsproblem: Lösung

$$P(A) = 1 - P(\bar{A}) = 1 - \frac{\text{Anzahl der für } \bar{A} \text{ günstigen Geburtstagsanordenungen}}{\text{Anzahl der möglichen Geburtstagsanordnungen}}$$

Hierbei handelt es sich um Kombinationen k-ter Ordnung von 365 Tage mit Berücksichtigung der Reihenfolge und ohne Wiederholung, sodass die Anzahl der möglichen Anordnungen für $\bar{A}$ folgt

$$\text{Anzahl der für } \bar{A} \text{ günstigen Geburtstagsanordnungen} = \frac{365!}{(365-k)!}$$

Bei den möglichen Anordnungen handelt es sich um Kombinationen k-ter Ordnung von 365 Tage mit Berücksichtigung der Reihenfolge und mit Wiederholung, sodass

$$\text{Anzahl der möglichen Geburtstagsanordnungen} = 365^k$$

Geburtstagsproblem: Lösung

$$P(A) = 1 - P(\bar{A}) = 1 - \frac{365 \cdot 364 \cdot \ldots \cdot (365 - k + 1)}{365^k}$$

Zwei Personen in einer Gruppe mit $k = 100$ Personen haben am gleichen Tag Geburtstag:

```r
k <- 100
1- prod(365:(365-k+1))/365^k
```

```
## [1] 1
```

Intuitiv werden bestimmte Wahrscheinlichkeiten häufig falsch eingeschätzt.