

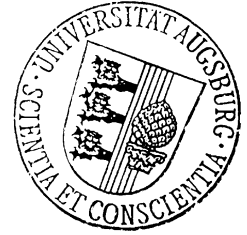
Stochastik für Lehramt

— eine Einführung in
Wahrscheinlichkeitstheorie,
Statistik und Datenanalyse

Antony R. Unwin

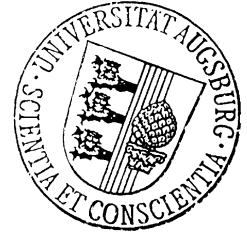
Lehrstuhl für Rechnerorientierte Statistik und Datenanalyse
Institut für Mathematik
Universität Augsburg

tel: 598-2218
email: unwin@math.augsburg.de



Ziele

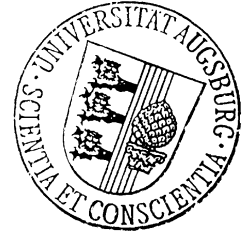
- Wahrscheinlichkeitsmodelle
bauen
modifizieren
interpretieren
- Daten analysieren und darstellen
- Statistische Analysen durchführen



Beispiele von Wahrscheinlichkeiten

Was ist die Wahrscheinlichkeit dafür, daß

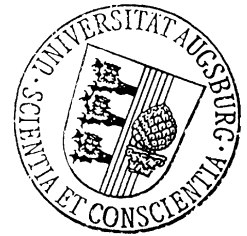
- 1) beim 20-maligen Werfen einer Münze mehr als 16-mal Wappen auftritt?
- 2) jemand vier Karten richtig rät?
- 3) eine Stichprobe von n Personen aus einer Klasse mehr als 25% Frauen enthält?
- 4) Ihr erstes Kind ein Mädchen ist?
- 5) die nächste Bundestagswahl frühzeitig gehalten werden muß?
- 6) morgen ein Flugzeugunglück passiert?
- 7) dieses Jahr eine nukleare Katastrophe passiert?



Wahrscheinlichkeitstheorie

Inhalt

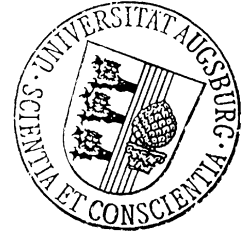
- Was ist Wahrscheinlichkeit?
- Kombinatorik
- Diskrete Modelle (Binomial, Poisson...)
- Stetige Modelle (Normal, Exponential..)
- Grenzwertsätze
- Simulation
- Anwendungen und Interpretation



Datenanalyse und Statistik

Inhalt

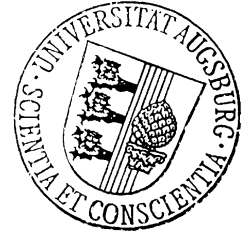
- Daten
erheben, organisieren, darstellen
zusammenfassen, analysieren
interpretieren
- Beschreibende Statistik
- Statistische Graphik
- Schätzen
- Testen



Anwendungen der Wahrscheinlichkeitstheorie

Modellierung von Variabilität und Unsicherheit

- Physik
 - Risikoanalysen
 - Genetische Modelle
 - Kartenspiele (Bridge, Poker ...)
 - und, und, und
- + Statistische Analysen



Anwendungen der Statistik und Datenanalyse

- Lotto Gewinnzahlen
- Studiendauer
- Proportion Zwillinge
- Autoversicherungspraemien
- Schrittmesser
- Verbraucherpreise
- Hit-Paraden
- Renten (Problem für wen?)
- Arbeitslosenzahlen



Bücher

N. Henze, *Stochastik für Einsteiger*
(8. Auflage, Vieweg, Braunschweig, 2009)

Fahrmeir, L., Künstler, R., Pigeot, I., Tutz, G.
Statistik
(7. Auflage) Springer 2009

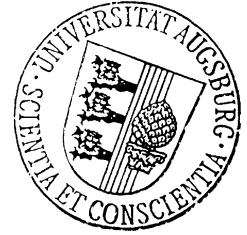
und viele, viele mehr....

Tufte, E.R.
The Visual Display of Quantitative Information
(2. Auflage) Graphic Press 2001

U. Ligges, *Programmieren mit R* (2. Auflage, Springer,
2006)

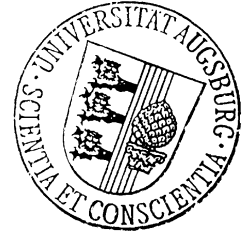
Webseite der Vorlesung:

rosuda.org/lehre/SS10-f/StochLAs10.shtml



Übungen und

- Jede Woche, ein Übungsblatt
Theorie
Modelle und Anwendungen
Datenanalysen, Graphiken
Rechnerbenutzung
- Klausuren für beide Gruppen im Juli
(Zweit-Klausuren im Oktober)
Note = Besser{1. Klausur, 2. Klausur}
- Schein
> 40% in der Klausur
- Zeitungen, Zeitschriften, Fernsehen
— kritisch denken



Pausenvorträge

- (Jede) Vorlesung, eine Präsentation
Beispiele aus den Medien
Theorie
Rechnerbenutzung
Geschichte
??
- Zwei StudentInnen
per Zufall ausgesucht
berichten für 5 Minuten
die Woche darauf
zu Halbzeit



K1 Was ist Wahrscheinlichkeit?

1.1 Ereignisse

Der einzelne Ausgang ist unbekannt, aber die Menge aller möglichen Ausgänge ist bekannt.

- Elementarereignisse $\{\omega\}$
- Ereignisraum Ω
(Stichprobenraum)
- Ereignis $A \subseteq \Omega$
(Ausgang)
- Wahrscheinlichkeitsmaß P



Interpretationen

$A \cup B$ Ereignis A oder B tritt ein

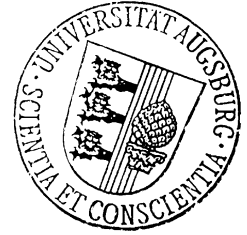
$A \cap B$ beide Ereignisse A und B treten ein

$\bigcup_{i=1} A_i$ mindestens eines von A_i tritt ein

$\bigcap_{i=1} A_i$ alle A_i treten ein

Zwei Ereignisse A und B sind disjunkt, wenn sie keine Elementarereignisse gemeinsam haben:

$$A \cap B = \emptyset$$



1.2 Beispiele

a) Münzen 20-mal werfen

b) Geschlecht des ersten Kinds

c) Fahrzeit von Uni zum Königsplatz

d) Sekunde, wenn ein Tor geschossen wird



1.3 Bestimmung von Wahrscheinlichkeiten

Sind die Situationen identisch?

- Kombinatorische Bestimmung
(Abzählen von Fällen)
- Statistische Schätzungen
(Wie häufig ist es vorher aufgetreten?)
- Logische Überlegungen
(Abgeleitet von Struktur)



1.4

Interpretation von Wahrscheinlichkeiten

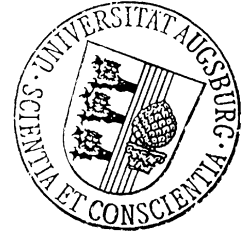
- Relative Häufigkeit

Es gibt eine unendliche Zahl von identischen Situationen, bei denen ein Ereignis A auftreten kann. Die Anzahl der n Versuche, bei denen A eintritt, heißt die absolute Häufigkeit des Ereignisses A , $h_n(A)$, und die relative Häufigkeit von A ist gleich

$$h_n(A)/n$$

Da relative Häufigkeiten, selbst bei noch so großem n , von der tatsächlichen Wahrscheinlichkeit beliebig weit entfernt sein können, gibt es Probleme.

- Subjektive Wahrscheinlichkeit



1.5 Axiome von Kolmogoroff (1933)

Eine Abbildung P von $\mathcal{P}(\Omega)$ in $[0,1]$ heißt Wahrscheinlichkeitsmaß wenn sie die folgenden Eigenschaften hat:

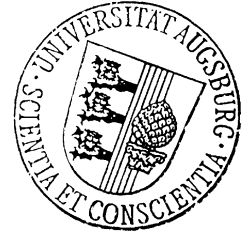
- $P(\Omega) = 1$
- $P(A) \geq 0$ für alle A
- $P(A \cup B) = P(A) + P(B)$
wenn A, B disjunkt sind

Ω ist eine nicht leere Menge der möglichen Versuchsergebnisse

\mathcal{F} ist die Menge der Ereignisse, A_1, A_2, A_3, \dots die uns interessieren.

$P(A)$ ist die Wahrscheinlichkeit von A .

Das Tupel (Ω, \mathcal{F}, P) heißt der dem Experiment zugeordnete Wahrscheinlichkeitsraum.

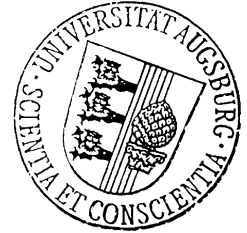


Beispiel für Kolmogoroff

Werfen eines Würfels

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

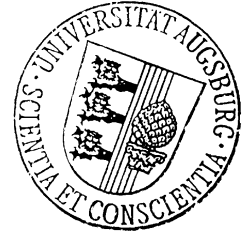
- $P(\text{eine Zahl aus } \{1, \dots, 6\}) = 1$
- $P(\text{Zahl} > 4) = 1/3$
 $P(\text{ungerade Zahl}) = 1/2$
 $P(\text{nicht } 2) = 5/6$
- $P(2 \text{ oder ungerade}) = P(2) + P(\text{ungerade})$
 weil 2 und ungerade disjunkt sind



1.5.1 Folgerungen aus den Axiomen

Für $A, B, A_i \in \mathcal{F}$ gilt

- $P(\bar{A}) = 1 - P(A)$
- $P(\emptyset) = 0$
- Aus $A \subset B$ folgt $P(A) \leq P(B)$
- $P(A \setminus B) = P(A) - P(A \cap B)$
wobei $(A \setminus B = A \cap \bar{B})$
- Wenn A_i disjunkt sind gilt
$$P(\cup A_i) = \sum P(A_i)$$
- Sonst gilt für beliebige A_1, A_2, A_3, \dots
$$P(\cup A_i) \leq \sum P(A_i)$$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$



Folgerungen für das Beispiel

- $P(\leq 4) = 1 - P(> 4)$
 $2/3 = 1 - 1/3$
- $P(\text{keine Zahl}) = 0$
- $\{2,4\} \subset \{\text{gerade Zahl}\}$
 $P(2 \text{ oder } 4) \leq P(\text{gerade Zahl})$
 $1/3 < 1/2$
- $P(\text{gerade Zahl aber nicht } < 3) =$
 $P(\text{gerade Zahl}) - P(\text{gerade Zahl } < 3)$
 $1/3 = 1/2 - 1/6$
- Disjunkte Ereignisse
 $P(3 \text{ oder gerade Zahl}) = 1/6 + 1/2$
- Beliebige Ereignisse
 $P(2 \text{ oder } 5 \text{ oder } \leq 4) \leq 1/6 + 1/6 + 2/3$
- $P(\text{gerade Zahl oder } \leq 3) =$
 $P(\text{gerade Zahl}) + P(\leq 3) - P(\text{gerade Zahl } \leq 3)$
 $5/6 = 1/2 + 1/2 - 1/6$



1.7 Unabhängigkeit

Ereignisse A und B heißen unabhängig wenn

$$P(A \cap B) = P(A) \cdot P(B)$$

$$\Rightarrow P(A) = P(A | B) \text{ und } P(B) = P(B | A)$$

$P(A | B)$ ist die Wahrscheinlichkeit von Ereignis A gegeben Ereignis B

Dieses Konzept ist ein wahrscheinlichkeitstheoretisches Konzept.

z.B.

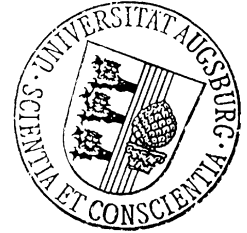
Beim Werfen eines Würfels

sei A: gerade Zahl und B: > 3

dann gelten $P(A) = 1/2$ und $P(A | B) = 2/3$

Sei A: gerade Zahl und B: > 2

dann gelten $P(A) = 1/2$ und $P(A | B) = 1/2$



Stochastische Unabhängigkeit

Wir haben schon die Unabhängigkeit von zwei Ergebnissen besprochen:

A, B unabhängig $\Leftrightarrow P(A \cap B) = P(A) \cdot P(B)$

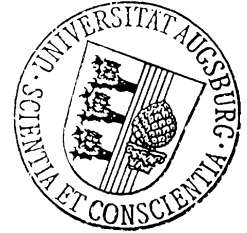
Die Verallgemeinerung ist komplizierter:

$A_1, A_2, A_3, \dots, A_m$ sind unabhängig

\Leftrightarrow

$P(\cap A_j) = \prod P(A_j)$ für jede Teilfamilie

Falls $P(A_i \cap A_j) = P(A_i) \cdot P(A_j)$ für jedes Paar A_i, A_j gelten sollte, sind $A_1, A_2, A_3, \dots, A_m$ paarweise unabhängig aber nicht unbedingt unabhängig.



Beispiel:

Beim zweimaligen Werfen einer Münze:

$$A_1 = \{1. \text{ Wurf Kopf}\}$$

$$A_2 = \{2. \text{ Wurf Zahl}\}$$

$$A_3 = \{\text{Beide verschieden}\}$$

$$P(A_1) = 1/2, P(A_2) = 1/2, P(A_3) = 1/2$$

$$P(A_1 \cap A_2) = 1/4 = P(A_1) \cdot P(A_2)$$

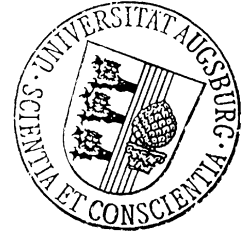
$$P(A_1 \cap A_3) = 1/4 = P(A_1) \cdot P(A_3)$$

$$P(A_2 \cap A_3) = 1/4 = P(A_2) \cdot P(A_3)$$

\Rightarrow jedes Paar A_i, A_j ist unabhängig

$$P(A_1 \cap A_2 \cap A_3) = 1/4 \neq P(A_1) \cdot P(A_2) \cdot P(A_3)$$

$\Rightarrow A_1, A_2, A_3$ sind nicht unabhängig



K2 Kombinatorik

(Abzählen der Fälle)

Annahme —

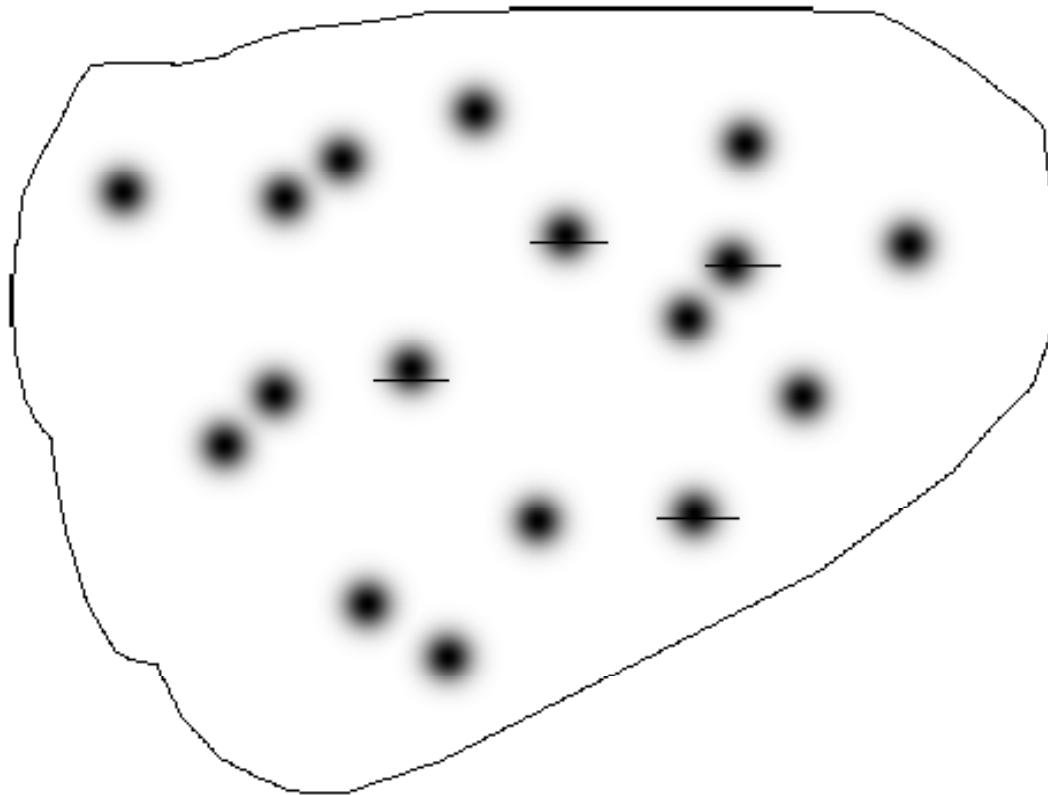
alle Ereignisse gleichwahrscheinlich

Dann können wir die Wahrscheinlichkeit eines Ausgangs als

$$\frac{\#(\text{„günstige“ Fälle})}{\#(\text{alle Fälle})}$$

bestimmen.

Dafür müssen wir wissen, wie man abzählt.



- Ereignisse
- günstige Ereignisse

Wenn alle Ereignisse gleichwahrscheinlich sind, ist die Wahrscheinlichkeit eines günstigen Resultats:



2.1 Permutationen

Aus n verschiedenen Elementen kann man unter Berücksichtigung der Reihenfolge und ohne Rücklegung k Stück ($1 \leq k \leq n$) auf

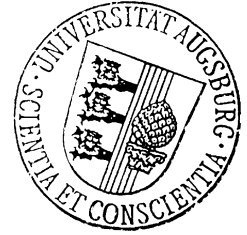
$$n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot (n-k+1) = n! / (n-k)! \quad (2.1.1)$$

verschiedene Arten auswählen.

n Objekte, von denen $n_1, n_2, n_3, \dots, n_r$ gleich sind, lassen sich auf

$$n! / (n_1! n_2! n_3! \dots n_r!) \quad (2.1.2)$$

verschiedene Arten anordnen.



Tennis Beispiel

Welche Reihenfolgen von Resultaten sind in einem Spiel aus fünf Sätzen möglich?

Das letzte Spiel muß gewonnen werden, die andern vier dürfen eine beliebige Reihenfolge haben. Es gibt

$4! / (2! 2!)$ Anordnungen von 2 G und 2 V:

GGVVG

GVGVG

GVVGG

VGGVG

VGVGG

VVGGG

Wären alle in der Tat gleich wahrscheinlich?



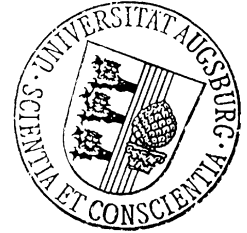
2.2 Kombinationen

Aus n verschiedenen Elementen kann man ohne Berücksichtigung der Reihenfolge und ohne Rücklegung k Stück ($1 \leq k \leq n$) auf

$$n! / [(n-k)! k!] \quad (2.2.1)$$

verschiedene Arten auswählen., d.h.

$$\frac{\# \text{Permutationen}}{k!}$$



2.3 Kombinatorische Resultate

Stichproben

(n Kugeln aus einer Urne mit N Kugeln)

Verteilungen

(n Murmeln verteilt auf N Zellen)

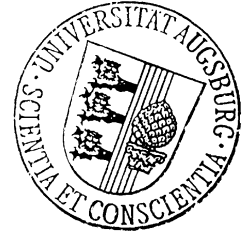


Herzogin

(n Ringe auf N Finger angesteckt)

Eisenbahn

(n Wagen verteilt auf N Rangiergleisen)

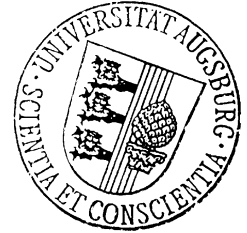


Stichproben vom Umfang n aus N unterschiedlichen Kugeln

	mit Rücklegen	ohne
in Reihenfolge	(2.3.1)	(2.3.2)
ohne Reihenfolge	(2.3.3)	(2.3.4)

Verteilungen von n Murmeln auf N Zellen

	mit Mehrfachbesetzung	ohne
unterscheidbare Murmeln		
ununterscheidbare Murmeln		



Erklärung für (2.3.3)

Eine Stichprobe der Größe n wird aus N unterschiedlichen Kugeln mit Zurücklegen gezogen (die Reihenfolge ist unwichtig).

oder

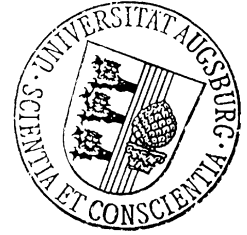
n identische Kugeln sollen auf N unterschiedliche Zellen verteilt werden.

Wieviele Möglichkeiten gibt es?

|000| |0| |00| |.....|0|

Die Anzahl der Anordnungen von n Kugeln (0) und $(N-1)$ Striche (|), die internen Grenzen der N Zellen:

$$(N-1+n)! / [(N-1)!n!]$$



2.4 Kombinatorische Beispiele

2.4.1 Pokerspiel

5 Karten aus 52

Insgesamt gibt es

mögliche Blattkombinationen.

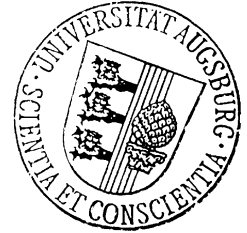
(a)

$P(\text{P-König, H-König, H3, Kr7, Ka2}) =$

(b)

$P(\text{P-König, H-König, ...}) =$

(einschließlich ein Blatt mit allen Königen)



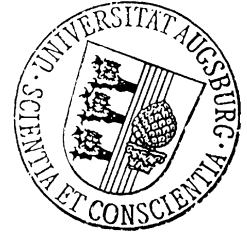
(c)

$P(\text{mindestens 1 schwarzer K und 1 roter K})$

$\{ \neq P(\text{genau 1 schwarzer K und 1 roter K}) \}$

$= P(\text{genau 1 sK, 1 rK}) + P(1 \text{ sK, 2 rK})$
 $+ P(2 \text{ sK, 1 rK}) + P(2 \text{ sK, 2 rK})$

=



1) Anzahl Blätter

Gegeben 52 Karten gibt es $52!$ Reihenfolgen

Für die ersten 5 Karten (ein Blatt in der Reihenfolge des Austeilens) $52! / 47!$

Da die Reihenfolge der 5 unwichtig ist

$$52! / 47! / 5!$$

2) Blatt mit (genau) einem König

$$4 \text{ Könige} + 48 \text{ andere} \Rightarrow 4C1 * 48C4$$

3) Blatt mit (genau) einem schwarzen und einem roten König

$$2 \text{ schwarze K, } 2 \text{ rote K} \Rightarrow 2C1 * 2C1 * 48C3$$



2.4.2 Genetik

In einer genetischen Sequenz wird jede Stelle entweder A, C, G oder T sein. Eine Sequenz der Länge n hat

4^n Möglichkeiten

$$P(\text{Sequenz ohne G}) = 3^n / 4^n$$

2.4.3 Lotto (6 aus 49)

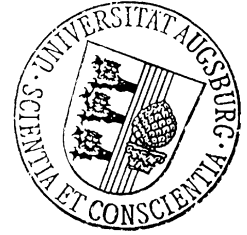
Anzahl Möglichkeiten =

$P(\text{ohne gerade Zahl}) =$

2.4.4 Stichproben

Aus einer Gruppe von 10 Männern und 10 Frauen wird eine Stichprobe der Größe 10 gezogen.

$P(\leq 2 \text{ Frauen}) =$



2.4.5 Ranglisten

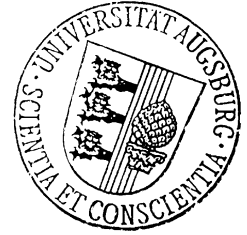
Auf wieviele Weisen können die ersten drei Plätze in der Bundesliga besetzt werden?

$18! / 15!$

2.4.6 Wahlfächern

30 Studentinnen wählen eins aus fünf verschiedenen Fächern. Wieviele mögliche Verteilungen gibt es?

$34! / 30! / 4!$



2.5 W-Theorie und Kombinatorik

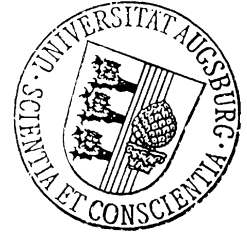
2.5.1 Prozedur

Raum der möglichen Ereignisse bestimmen:

- mit/ohne Rücklegen
- Unterscheidbarkeit
(von Objekten, Zellen....)
- Reihenfolgen

Sind alle Ereignisse gleichwahrscheinlich, dürfen wir kombinatorische Methoden anwenden.

Abzählen der Ereignisse insgesamt
Abzählen der „günstigen“ Ereignisse



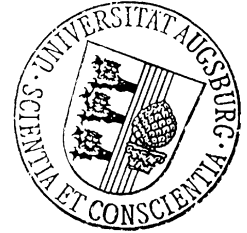
2.5.2 Anwendung und Interpretation

In der Kombinatorik gibt es viele interessante (und manchmal trickreiche) Fragestellungen.

z.B.

- (1) Auf wieviele Weisen können m Ehepaare um einen Tisch sitzen, so dass jede Frau zwischen zwei Männern sitzt und nicht neben Ihrem Mann?
- (2) Zwei Kartenspiele werden getrennt gemischt. Was ist die Wahrscheinlichkeit, dass eine Karte dieselbe Position in beiden Spielen besetzt?

Seitens der Wahrscheinlichkeitstheorie sind die Annahmen und die Interpretationen am wichtigsten.



K3 (Diskrete) Zufallsvariablen

3.1 Basis

$\Omega = \{\omega\}$, $X(\omega)$ ist eine Größe die durch ω bestimmt ist. Bei der zufälligen Auswahl von ω bekommen wir den Wert, $X(\omega)$.

Definition: Ist (Ω, \mathcal{F}, P) ein Wahrscheinlichkeitsraum und X eine beliebige Menge, so nennen wir eine Abbildung
 $X: \Omega \rightarrow X$ eine X -wertige Zufallsvariable

Falls Ω abzählbar ist, wird der Wertebereich von X , $\{X(\omega): \omega \in \Omega\}$ auch abzählbar sein. Die Verteilung von X ist das Wahrscheinlichkeitsmaß

$$P_X(x) = P(\{\omega \in \Omega: X(\omega) = x\})$$



3.2 Erwartungswerte

Die Erwartung einer reellwertigen Zufallsvariablen X existiert wenn $\sum |X(\omega)| P(\omega)$ konvergiert. Wir definieren dann

$$E[X] = \sum X(\omega)P(\omega)$$

als den Erwartungswert von X

Ist x_1, x_2, x_3, \dots eine Abzählung des Wertebereichs von X , so ist

$$E[X] = \sum x_i P(X = x_i)$$

Eigenschaften:

(1) $E[X+a] = E[X] + a$

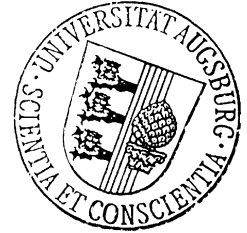
(2) $E[X+Y] = E[X] + E[Y]$

(3) $E[aX] = aE[X]$

(4) sind X, Y unabhängig:

$$E[X*Y] = E[X]*E[Y]$$

(5) $E[f(X)] \neq f(E[X])$



3.2.1 Warum brauchen wir Konvergenz?

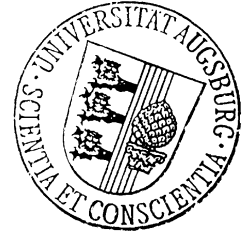
(Das Paradox von St. Petersburg)

Man wirft eine Münze bis Kopf auftritt.
Wenn das zum ersten Mal mit dem r .ten
Wurf geschieht, gewinnt man

$$€ 2^r$$

Nehmen wir an, daß $P(\text{Kopf}) = p$

$$\begin{aligned} E[\text{Gewinn}] &= \sum_{r=1}^{\infty} 2^r p (1-p)^{r-1} \\ &= 2p \sum_{r=1}^{\infty} (2q)^{r-1} \\ &= 2p \sum_{r=0}^{\infty} (2q)^r \\ &= 2p / (1 - 2q) \text{ für } 2q < 1 \\ &\text{sonst } \infty \end{aligned}$$



3.3 Verteilungen auf den Ganzzahlen

$$\Omega = \{0, 1, 2, \dots\}$$

$$P(X = k) = p_k$$

Man interessiert sich insbesondere für

$$P(X = 0) = p_0$$

$$P(X \geq k) = \sum_{i \geq k} p_i$$

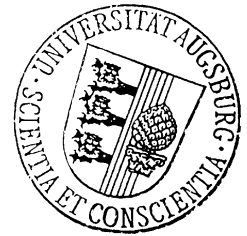
$$P(X = r \mid X \geq k) = p_r / \sum_{i \geq k} p_i$$

$$P(j \leq X \leq k) = \sum_{i=j}^k p_i$$

$$E[X] = \sum i p_i \text{ (arithmetischer Mittel)}$$

$$V[X] = \sum (i - E[X])^2 p_i$$

$$\text{Standardabweichung von } X = \sqrt{V[X]}$$



z.B.

k	p_k
0	0.2
1	0.4
2	0.3
3	0.1
≥4	0.0

$$P(X = 0) =$$

$$P(X \geq 2) =$$

$$P(X = 3 | X \geq 2) =$$

$$P(1 \leq X \leq 2) =$$

$$E[X] =$$

$$V[X] =$$



3.4 Hypergeometrische Verteilung

Eine Urne enthält
 S schwarze und W weiße Kugeln.

Es werden $n \leq S + W$ Kugeln ohne
Rücklegen gezogen.

Die Wahrscheinlichkeit, daß die Stichprobe
genau s schwarze Kugel enthält ist

$$h(s; n, N, S) =$$

wo

$$N = S + W, n = s + w \text{ und } 0 \leq s \leq \min(n, S)$$

Für eine Urne mit mehreren Farben, z. B.
 S schwarze, W weiße, R rote, B blaue Kugeln
gilt:

$$P(s, w, r, b) =$$



3.5 Produktexperimente

Wir haben eine Reihe von unabhängigen Experimenten $\{1, 2, 3, \dots, i, \dots\}$ mit Modellen (Ω_i, F_i, P_i) . Das Gesamtergebnis hat einen Wahrscheinlichkeitsraum (Ω, F, P) mit

$$\Omega = \prod \Omega_i, F = \prod F_i, P = \prod P_i$$

das Produkt der Wahrscheinlichkeitsräume

Man zieht z.B. 4 Kugeln hintereinander aus einer Urne mit 10 schwarze Kugeln und 20 weiße Kugeln.

(a) mit Rücklegen

alle 4 Experimente haben denselben Wahrscheinlichkeitsraum (Ω_i, F_i, P_i)

(b) ohne Rücklegen

das Resultat der zweiten Ziehung ist vom Resultat der ersten abhängig und somit ist (Ω_2, F_2, P_2) im voraus nicht bestimmbar.



Produktexperimente — Beispiele

Geometrische Verteilung

In jedem Experiment gibt es zwei mögliche Ausgänge, $\{1, 0\}$ mit Wahrscheinlichkeiten p bzw. $q (= 1-p)$.

Wie groß ist die Wahrscheinlichkeit, daß k Nullen vor dem ersten Einser auftreten?

$$p_k = pq^k \quad k = 0, 1, 2, \dots$$

Binomiale Verteilung

Es gibt genau N der obigen Experimente. Wie groß ist die Wahrscheinlichkeit, daß man k Einsen bekommt?

$$\left[\frac{N!}{k! (N - k)!} \right] p^k (1 - p)^{N-k}$$

Negative Binomiale Verteilung

Es gibt keine Begrenzung der Zahl der Experimente. Wie groß ist die Wahrscheinlichkeit, daß man k Nullen vor dem r .ten Einser bekommt?

$$\left[\frac{(k + r - 1)!}{k! (r - 1)!} \right] p^r (1 - p)^k$$



3.6 Geometrische Verteilung

3.6.1 Formeln

$$p_k = pq^k \quad (q = 1 - p) \quad (3.6.1)$$

$$p_0 = p$$

$$P(X \geq k) = \sum_{i \geq k} p_i = pq^k \sum_{i \geq 0} q^i = q^k$$

$$P(X = r \mid X \geq k) = p_r / \sum_{i \geq k} p_i = pq^{r-k}$$

$$P(j \leq X \leq k) = \sum_{i=j}^k p_i = q^j (1 - q^{k+1-j})$$

$$E[X] = \sum_{i \geq 0} ip_i = pq \sum_{i \geq 1} iq^{i-1} = q/p$$

$$V[X] = \sum_{i \geq 0} (i - E[X])^2 p_i = q/p^2$$



3.6.2 Anwendungen — Geometrische Verteilung

1) Zuverlässigkeitsmodelle

Wieviele Backups?

$$P(\text{Absturz}) = p$$

$$P(k \text{ stürzen ab}) = p^k$$

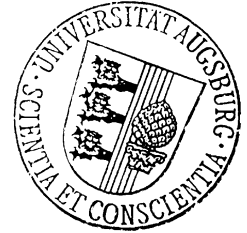
Unabhängigkeit ist eine wichtige Annahme

2) „Coupons“ Problem

$$P(\text{man bekommt den seltenen Fall}) = p$$

$$P(k + 1 \text{ kaufen müssen}) = q^k p$$

3) Zeit zum ersten Erfolg



3.7 Binomialverteilung $B(N, p)$

3.7.1 Formeln

$$p_k = N! / (k! (N-k)!) p^k q^{N-k} \quad (3.7.1)$$

$$p_0 = q^N$$

$$P(X \geq k) = \sum_{i \geq k} p_i$$

$$P(X = r \mid X \geq k) = p_r / \sum_{i \geq k} p_i$$

$$P(j \leq X \leq k) = \sum_{i=j}^k p_i$$

$$E[X] = \sum_{i \geq 0} i p_i = Np$$

$$V[X] = \sum_{i \geq 0} (i - E[X])^2 p_i = Npq$$

(Bernoulli Experimente)

Für N unabhängige Zufallsvariablen $\{X_i\}$,
wo $P(X_i = 1) = p$ und $P(X_i = 0) = 1 - p$,
hat $X = \sum X_i$ eine Binomialverteilung.



3.7.2 Anwendungen — Binomialverteilung

1) Qualitätskontrolle

Stichproben der Größe N . Jedes Produkt hat die gleiche Wahrscheinlichkeit, p , fehlerhaft zu sein.

2) „Multiple Choice“

N Fragen

r Antworten pro Frage $P(\text{richtig}) = 1/r$

Die Verteilung der #(richtige Antworten)

$$B(N, 1/r)$$

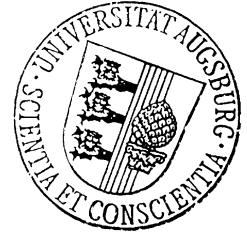
3) „ESP“

N Versuche $p = P(\text{richtig})$

$$P(X \geq k) = \sum_{i=k}^N \frac{N!}{[i!(N-i)!]} p^i q^{N-i} = g_k$$

M Personen $p = g_k$

$$P(\text{mehr als } K \text{ Personen } \geq k \text{ richtig haben}) = \sum_{j>K}^M \frac{M!}{[j!(M-j)!]} g_k^j (1 - g_k)^{M-j}$$



3.8 Poisson Verteilung

3.8.1 Formeln

$$p_k = e^{-\lambda} \lambda^k / k! \quad k = 0, 1, 2, \dots \quad (3.8.1)$$

$$p_0 = e^{-\lambda}$$

$$P(X \geq k) = e^{-\lambda} \sum_{i \geq k} \lambda^i / i! = \Gamma_k(\lambda) \quad (*)$$

$$P(X = r \mid X \geq k) = p_r / \sum_{i \geq k} p_i$$

$$P(j \leq X \leq k) = \Gamma_j(\lambda) - \Gamma_{k+1}(\lambda)$$

$$E[X] = \lambda$$

$$V[X] = \lambda$$

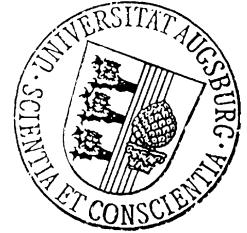
$$(*) \quad \Gamma_k(\lambda) = \int_0^\lambda e^{-t} t^{k-1} dt / (k-1)!$$

$\gamma(k, \lambda) = \int_0^\lambda e^{-t} t^{k-1} dt$ ist die (untere) unvollständige Gammafunktion



3.8.2 Anwendungen — Poisson Verteilung

- 1) # Unfälle —
z.B. Todesfälle nach Huftritten von
Pferden in der preußischen Armee
14 Korps 1875-1894 196 Todesfälle
Das Gesetz der kleinen Zahlen
L. von Bortkewitsch (Teubner 1898)
- 2) # Fehler — z.B. Tippfehler
- 3) Tore im Sport (?)
- 4) Bombenverteilung in London
(Feller s. 160)
- 5) # Anrufe



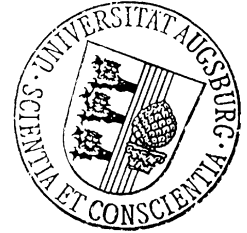
Beispiel 3.8.2 (4) Bomben in London (V1 und V2)

Das Gebiet von Süd-London wird in 576 kleinere Gebiete von je 0.25 km^2 aufgeteilt. Die Tabelle zeigt die Anzahl Gebiete N_k , in denen genau k Bomben gefallen sind.

k	N_k	Poisson($k; 0.9323$)
0	229	226.74
1	211	211.39
2	93	98.54
3	35	30.62
4	7	7.14
≥ 5	1	1.57

λ wird durch $537/576$ geschätzt.

Wie soll man dieses Resultat deuten?



Eine wichtige Begründung in vielen Anwendungen ist die folgende Eigenschaft der Poisson Verteilung:

Satz 3.8.2

X, Y Poissonverteilte ZV mit Parametern λ_1, λ_2

$$X \sim P(\lambda_1) \text{ und } Y \sim P(\lambda_2)$$

X, Y unabhängig $\Rightarrow X+Y \sim P(\lambda_1 + \lambda_2)$

Beweis

Entweder über

(a) Faltung

oder

(b) Wahrscheinlichkeitserzeugende Funktionen

Diskrete Verteilungen — Rekapitulation

H: Hypergeometrische — ziehen ohne Rücklegen

$$X \sim H(k; n, N, K)$$

B: Binomial — Wieviele Treffer aus N?

$$X \sim B(N, p)$$

G: Geometrische — Wieviele Versuche bis Erfolg?

$$X \sim G(p)$$

NB: Negative Binomial – Wieviele Versuche zum r. Erfolg?

$$X \sim NB(r, p)$$

P: Poisson — Wieviele Ereignisse?

$$X \sim P(\lambda)$$

B, G, NB, P — alle nehmen Unabhängigkeit an.

und Spezialfälle wie Benford's Gesetz:

Die Verteilung der ersten Ziffer in Zahlen ist

$$P(\text{erste Ziffer} = n) = \log_{10} \left(1 + \frac{1}{n} \right) \quad n = 1, \dots, 9$$

Mathematische Resultate für diskrete Verteilungen

Für eine diskrete Zufallsvariable X mit Verteilung

$$P(X = x) = p(x) \quad x = 0, 1 \dots$$

gilt

$$\sum_{x=0}^{\infty} p(x) = 1$$

$$E[X] = \sum_{x=0}^{\infty} xp(x)$$

$$V[X] = \sum_{x=0}^{\infty} (x - E[X])^2 p(x)$$

z.B. $X \sim B(N, p)$ (Binomial Verteilung)

$$\sum_{x=0}^N \binom{N}{x} p^x (1-p)^{N-x} = 1$$

$$E[X] = \sum_{x=0}^N x \binom{N}{x} p^x (1-p)^{N-x}$$

$$= Np \sum_{y=0}^{N-1} \binom{N-1}{y} p^y (1-p)^{N-1-y} = Np$$

Beispiele von Anwendungen diskreter Verteilungen

1. 10% der Bevölkerung bringen Leistungen besser als der Wert L . Wieviele Leute muss man testen, bis man 20 hat?
2. Die Krankheit ALS hat eine Prävalenzrate von 3 bis 8 pro 100.000. Wäre es überraschend, 20 Patienten in Augsburg zu finden?
3. Nehmen wir an, dass 25% der Bevölkerung SPD wählen wollen. Was ist die Verteilung von SPD Wählern in einer Stichprobe der Größe 10? 100? 1000?
4. Für ein Fernsehprogramm soll ein Wähler von jeder Partei beim Einkaufen in der Stadt befragt werden. Wenn die Linken 5,5% Unterstützung haben, was ist die Verteilung der Anzahl Leute, die angesprochen werden muss, bis ein Wähler der Linken gefunden wird?
5. Im Durchschnitt gibt es ein Erdbeben der Stärke 8 oder höher jedes Jahr. Wie könnte man die Verteilung der Anzahl Erdbeben im nächsten Jahr modellieren?

3.8.3 Die Poisson Verteilung als eine Approximation für die Binomialverteilung

$$X \sim B(N, p)$$

$$P(X = k) = \binom{N}{k} p^k (1 - p)^{N-k}$$

Sei $Np = \lambda$, dann ist $X \sim B(N, \lambda/N)$ und

$$\begin{aligned} P(X = k) &= \binom{N}{k} \left(\frac{\lambda}{N}\right)^k \left(1 - \frac{\lambda}{N}\right)^N \left(1 - \frac{\lambda}{N}\right)^{-k} \\ &= \left(1 * \frac{N-1}{N} \dots \frac{N-k+1}{N}\right) \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{N}\right)^N \left(1 - \frac{\lambda}{N}\right)^{-k} \end{aligned}$$

Für $N \rightarrow \infty$ und λ konstant gilt

$$P(X = k) \rightarrow \frac{\lambda^k}{k!} e^{-\lambda}$$

eine Poisson Verteilung.

Man findet auch

”Für $N \rightarrow \infty, p \rightarrow 0$ unter der Bedingung, dass $Np \rightarrow \lambda$, ist $P(Np)$ eine gute Approximation für $B(N, p)$.”

3.8.4 Die Poisson Verteilung ist die Zufallsverteilung

Seien

$$p_1(i) = P(i \text{ Ereignisse im Zeitintervall } t_1)$$

$$p_2(j) = P(j \text{ Ereignisse im Zeitintervall } t_2)$$

wo $t_1 \cap t_2 = \emptyset$ und

$$p_0(n) = P(n \text{ Ereignisse im Zeitintervall } t_0 = t_1 \cup t_2)$$

Dann

$$p_0(n) = \sum_{i=0}^n p_1(i)p_2(n-i)$$

Wir möchten, dass die Verteilungen $\{p_k(i)\}$ alle dieselbe Form haben. Wir nehmen an, dass die Ereignisse rein zufällig und unabhängig von einander sind.

\Rightarrow Die Wahrscheinlichkeit ein Ereignis in t_1 fällt, gegeben, dass es in $t_0 = t_1 \cup t_2$ fällt, ist

$$\alpha = \frac{t_1}{t_0}$$

Dann gilt

$$p_1(i)p_2(n-i) = \binom{n}{i} \alpha^i (1-\alpha)^{n-i} p_0(n)$$

Insbesondere haben wir

$$p_1(n)p_2(0) = \alpha^n p_0(n) \quad (\Rightarrow p_2(0) > 0)$$

und

$$p_1(n-1)p_2(1) = n\alpha^{n-1}(1-\alpha)p_0(n)$$

So dass

$$\frac{p_1(n)}{p_1(n-1)} = \frac{p_2(1)}{p_2(0)} \left(\frac{\alpha}{1-\alpha} \right) \frac{1}{n}$$

Für $n = 1$

$$\frac{p_1(1)}{p_1(0)} \frac{1}{\alpha} = \frac{p_2(1)}{p_2(0)} \frac{1}{1-\alpha} = c$$

$$\Rightarrow \frac{p_1(n)}{p_1(n-1)} = \frac{\alpha c}{n}$$

und so

$$p_1(n) = \frac{(\alpha c)^n}{n!} p_1(0)$$

eine Poisson Verteilung mit $\lambda = \alpha c$

3.8.5 Noch zwei Poisson Beispiele

1. Geburtstage

Gegeben 100 Studenten, wieviele haben heute Geburtstag? Als Approximation nehmen wir:

$$X \sim P\left(\frac{100}{365}\right)$$

$$P(\text{mindestens 1 Geburtstag}) \approx 1 - 0.76$$

2. Lotto

Wieviel Gewinner im Lotto?

$$P(\text{Größter Gewinn}) = \frac{1}{\binom{49}{6}} * 0.1$$

Gegeben 100.000.000 Tipps, die alle zufällig ausgewählt werden, modellieren wir die Anzahl Gewinner mit $X \sim P(0.715112)$

X	0	1	2	≥ 3
$P(X)$	0.489	0.350	0.125	0.036

3.9 R und diskrete Verteilungen

Für jede Verteilung gibt es vier Kommandos. z.B. für die Binomial Verteilung $X \sim B(N, p)$ haben wir

dbinom(x, N, p) $P(X = x)$

pbinom(y, N, p) $P(X \leq y)$

qbinom(s, N, p) $P(X \leq z) = s$

rbinom(n, N, p) eine Zufallsstichprobe der Größe n

Bei einer Poisson Verteilung hat man

dpois(x, λ), **ppois(y, λ)**, **qpois(z, λ)**, **rpois(n, λ)**

Für die Negativ Binomial Verteilung hat man

dnbinom(x, N, p) usw

Für die Hypergeometrische hat man

dhyper(k, K, N - K, n) usw

K4 Bedingte Wahrscheinlichkeiten

4.1 Definition

Die bedingte Wahrscheinlichkeit von A bei gegebenem B

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (1)$$

Meistens benutzen wir diese Form

$$P(A \cap B) = P(A|B)P(B)$$

weil $P(A \cap B)$ schwer zu berechnen ist.

N.B. $P(A|B) = P(A)$ bedeutet A, B unabhängig.

Beispiel (1) Lebensdauer

Sei $p_k = P(\text{jemand im } k.\text{ten Lebensjahr stirbt})$
unabhängig von seinem Geburtsjahr.

Dann ist $P(\text{jemand das } k.\text{te Jahr erreicht})$

$$s_k = p_k + p_{k+1} + \dots$$

$P(\text{im Alter } r \text{ zu sterben, wenn man schon } k \text{ Jahre alt ist})$

$$= 0 \quad r < k$$

$$= \frac{p_r}{s_k} \quad r \geq k$$

Beispiel (2) Lotto

Nehmen wir an, dass die Anzahl Gewinner des Jackpots X Poissonverteilt sei.

$$X \sim P(\lambda)$$

1. Gegeben, dass der Jackpot gewonnen wird, was ist die Wahrscheinlichkeit, dass er geteilt wird?

$$P(X > 1 | X > 0) = \frac{1 - e^{-\lambda} - \lambda e^{-\lambda}}{1 - e^{-\lambda}}$$

2. Gegeben, dass ich den Jackpot gewonnen habe, was ist die Wahrscheinlichkeit, dass ich ihn teilen muss?
Sei $p = P(\text{ich gewinne den Jackpot})$

$$P(\text{ich und } x \text{ andere gewinnen}) = p \frac{\lambda^x}{x!} e^{-\lambda}$$

$$\Rightarrow P(x \text{ andere gewinnen} | \text{ich}) = \frac{\lambda^x}{x!} e^{-\lambda}$$

$$\Rightarrow P(x \text{ ich den Jackpot teilen muss}) = 1 - e^{-\lambda}$$

4.2 Formel von Bayes

(a) Formel von der totalen Wahrscheinlichkeit

$\{B_1, B_2, B_3, \dots\}$ heisst Zerlegung von Ω wenn die B_i disjunkt sind, und ihre Vereinigung Ω ist.

Für jede Zerlegung und jedes Ereignis A gilt

$$P(A) = \sum P(B_k)P(A|B_k) \quad (2)$$

(Ist $P(B_k) = 0$, so ist $P(A|B_k)$ nicht definiert, aber das Produkt kann $= 0$ gesetzt werden.)

(b) Bayes

Gilt $P(A) > 0$ und gelten alle Voraussetzungen von (a) , so ist $\forall i$

$$P(B_i|A) = \frac{P(B_i)P(A|B_i)}{\sum P(B_k)P(A|B_k)} \quad (3)$$

Beweis von §4.2 (3) (Bayes Formel)

$$P(A \cap B_i) = P(A|B_i)P(B_i) = P(B_i|A)P(A)$$

\Rightarrow

$$P(B_i|A) = \frac{P(B_i)P(A|B_i)}{P(A)}$$

$$= \frac{P(B_i)P(A|B_i)}{\sum P(B_k)P(A|B_k)}$$

$\{P(B_i)\}$ sind die apriori Wahrscheinlichkeiten

$\{P(B_i|A)\}$ sind die aposteriori Wahrscheinlichkeiten

Mit Bayes werden Wahrscheinlichkeiten berechnet. Das erlaubt aber keinen Rückschluss auf etwaige Kausalzusammenhänge zwischen den Ereignissen.

4.3.1 Beispiel mit Bayes (Autoversicherung)

Drei Gruppen von KFZ Versicherten:

Gut	G
Mittelmässig	M
Schlecht	S

Nehmen wir an, dass

$$P(\text{Unfall im Jahr} | G) = 0.01$$

$$P(\text{Unfall im Jahr} | M) = 0.03$$

$$P(\text{Unfall im Jahr} | S) = 0.10$$

und $P(G) = 0.3$, $P(M) = 0.5$, $P(S) = 0.2$

Jemand hat im Jahr einen Unfall, wie sehen die aposteriori Wahrscheinlichkeiten aus?

$$P(S|\text{Unfall}) = \frac{0.2*0.1}{0.2*0.1 + 0.5*0.03 + 0.3*0.01} = 0.5263$$

4.3.2 Beispiel mit Bayes (Sport-Kommentatoren)

”Bayern gewinnt, wenn Robben ein Tor schießt.”

Nehmen wir an, dass

$$P(\text{Bayern gewinnt}) = 0.8$$

$$P(\text{Bayern gewinnt} | \text{Robben Tor}) = 0.9$$

$$P(\text{Robben Tor}) = 0.3$$

Dann

$$P(\text{Robben Tor} | \text{Bayern gewinnt}) =$$

4.3.3 Beispiel mit Bayes (Medizinische Tests)

Die Genauigkeit des ELISA Tests für HIV wird von Wikipedia (8.5.2010) so angegeben:

Sensitivität [$P(+ \text{ Test} | \text{ HIV})$] 99.9%

Spezifität [$P(- \text{ Test} | \text{ kein HIV})$] 99.8%

Es gibt ungefähr 40,000 HIV-infizierte in Deutschland
⇒ eine Prävalenzrate von 0.0005.

Wenn eine zufällig ausgewählte Person einen positiven Test hat, was ist die Wahrscheinlichkeit, dass sie HIV-infiziert ist?

$$P(\text{HIV} | +) = \frac{0.999 * 0.0005}{0.999 * 0.0005 + 0.002 * 0.9995}$$

4.3.4 Beispiel mit Bayes (Das Auto und Ziegen Problem)

Es gibt drei Türe: A, B, und C. Hinter einer Tür steht ein Auto und hinter den anderen zwei steht je eine Ziege. Sie wählen eine Tür aus. Bevor der Moderator die Tür öffnet, öffnet er eine der anderen, um eine Ziege zu zeigen. "Gut, dass Sie diese Tür nicht gewählt haben", sagt er. Jetzt bietet er Ihnen die Möglichkeit, Ihre Tür zu wechseln, sollten Sie?

o.B.d.A. wählen Sie zuerst Tür A.

$$P(\text{Auto hinter Tür A}) = \frac{1}{3}$$

Der Moderator öffnet Tür B. Was ist

$P(\text{Auto hinter Tür A} | \text{A gewählt, B geöffnet})$?

$$= \frac{\frac{1}{3} * \frac{1}{2}}{\frac{1}{3} * \frac{1}{2} + \frac{1}{3} * 1}$$

K5 Wahrscheinlichkeitsmaße mit Dichten

5.1 Ereignisräume

Ein endlicher Raum ohne Reihenfolge:

(rot, schwarz, blau)

Ein endlicher geordneter Raum:

$(1, 2, 3, 4, 5, 6)$

Ein abzählbarer Raum:

$(0, 1, \dots)$

Ein nichtabzählbarer Raum:

$[0, \infty)$ oder $[0, 1]$ oder $(-\infty, \infty)$

Bis jetzt haben wir höchstens abzählbare Ereignisräume betrachtet. Diskrete Zufallsvariablen können wir als Abbildungen auf \mathbb{N}_0 betrachten. Bei unendlichen nicht-abzählbaren Räumen Ω werden wir darauf verzichten, $P(A)$ für alle $A \in \Omega$ zu definieren. Wir wollen mit kontinuierlichen Zufallsvariablen arbeiten können, die auf $(-\infty, \infty)$ definiert sind.

z.B. Länge einer Krankheit, Betriebsergebnisse, Meßfehler, Wartezeiten, Schadenssummen, Wasserverbrauch,

5.2 Wahrscheinlichkeitsräume

Def 5.2.1

Ein meßbarer Raum ist ein Paar (Ω, \mathcal{A}) bestehend aus einer nichtleeren Menge Ω und einer σ -Algebra \mathcal{A} von Teilmengen von Ω . Ein Wahrscheinlichkeitsmaß P ist eine auf \mathcal{A} definierte Funktion mit Werten in $[0, 1]$ welche den folgenden Bedingungen genügt:

$$P(A) \geq 0 \quad \forall A \in \mathcal{A}$$

$$P(\Omega) = 1$$

P ist σ -additiv, d.h. für disjunkte $A_1, A_2, \dots \in \mathcal{A}$ gilt

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

(Ω, \mathcal{A}, P) heißt Wahrscheinlichkeitsraum.

5.3 Verteilungsfunktionen und Dichten

5.3.1 Definitionen

Eine Funktion F auf \mathbb{R} mit Werten in $[0, 1]$ heißt Verteilungsfunktion, wenn sie rechtsstetig und monoton wachsend ist (nicht notwendig strikt monoton) und wenn für

$$x \rightarrow -\infty \quad F(x) \rightarrow 0$$

$$x \rightarrow \infty \quad F(x) \rightarrow 1$$

gilt.

Ist P ein Wahrscheinlichkeitsmaß auf \mathbb{R} und setzt man $F(x) = P((-\infty, x])$, so ist F eine Verteilungsfunktion. Umgekehrt

$$P((a, b]) = F(b) - F(a)$$

Eine Dichte auf \mathbb{R} ist eine nichtnegative Funktion mit

$$\int_{-\infty}^{\infty} f(x) dx = 1$$

Das Integral muß wohldefiniert sein: f muß stetig bis auf höchstens endlich viele Sprungstellen sein.

Ist f eine Dichte, dann ist

$$F(x) = \int_{-\infty}^x f(t) dt$$

eine Verteilungsfunktion. Die Wahrscheinlichkeit von Intervallen ist durch

$$P((a, b]) = \int_a^b f(t) dt$$

gegeben.

Ist F eine beliebige stetige Verteilungsfunktion auf \mathbb{R} , die auf dem Komplement einer endlichen oder leeren Menge C stetig differenzierbar ist, so wird durch

$$f(x) = \frac{d}{dx} F(x) \quad x \in \mathbb{R} \setminus C$$

eine Dichte f zu F definiert.

Ist (a_n) eine wachsende Folge mit $a_n < b$, die gegen b strebt, so ist $\{b\}$ der Durchschnitt der $(a_n, b]$ und die über diese Intervalle erstreckten Integrale streben gegen 0.

$$\Rightarrow P(\{b\}) = 0 \quad \forall b$$

Kein Ergebnis $b \in \mathbb{R}$ hat positive Wahrscheinlichkeit.

5.3.2 Beispiele von Dichten

- Die Gleichverteilung (z.B. Zufallszahlen)

$$f(x) = \frac{1}{b-a} \quad a \leq x \leq b$$

Insbesondere

$$f(x) = 1 \quad 0 \leq x \leq 1$$

- Die Exponentialverteilung (z.B. Zeit zum Ereignis)

$$f(x) = \lambda e^{-\lambda x} \quad x \geq 0$$

- Die Normalverteilung (z.B. Meßfehler)

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad -\infty < x < \infty$$

5.3.3 Zwei Eigenschaften der Exponentialverteilung

Sei $X \sim P(\lambda t)$ d.h. λ ist die erwartete Anzahl Ereignisse pro Zeiteinheit. Sei T eine Zufallsvariable, die Zeit zum nächsten Ereignis.

$$P(T > t) = P(X = 0 \text{ in } t)$$

$$1 - F(t) = e^{-\lambda t}$$

$F(t)$ ist die Verteilungsfunktion von t .

$$F(t) = 1 - e^{-\lambda t}$$

T hat eine Exponentialverteilung, $T \sim E(\lambda)$

Für eine Exponentialverteilung gilt, $b > a$,

$$P(T > b | T > a) = P(T > b - a)$$

weil

$$\frac{\int_b^{\infty} \lambda e^{-\lambda t} dt}{\int_a^{\infty} \lambda e^{-\lambda t} dt} = e^{-\lambda(b-a)}$$

”Gedächtnislosigkeit”

5.3.4 Eine wichtige Eigenschaft der Normalverteilung

Sei X_1 und X_2 unabhängig normalverteilt mit Erwartungswerten und Varianzen (μ_i, σ_i^2)

$$X_1 \sim N(\mu_1, \sigma_1^2), \quad X_2 \sim N(\mu_2, \sigma_2^2)$$

dann gilt

$$W = X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

und

$$V = X_1 - X_2 \sim N(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$$

Beweis durch Faltung oder Charakteristische Funktionen.

Die wichtigste Eigenschaft der Normalverteilung ergibt sich aus dem Zentralen Grenzwertsatz.

5.4 Erwartungswerte

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx$$

5.4.1 Eigenschaften von Erwartungswerten

(wie schon für den diskreten Fall im §3.2)

- für alle $a, b \in \mathbb{R}$

$$E[a + bX] = a + bE[X]$$

- ob X und Y unabhängig sind oder nicht

$$E[X + Y] = E[X] + E[Y]$$

- im allgemeinen gilt

$$E[g(X)] \neq g(E[X])$$

- für X und Y unabhängig gilt

$$E[XY] = E[X] * E[Y]$$

aber sonst nicht.

5.5 Und in der Praxis? Modellierung

- Was ist eine Zufallsvariable?
z.B. Länge einer Dienstreise
Niederschlag pro Tag
Unsicherheit bei Geschwindigkeitsmessungen
- Eine Verteilungsfunktion? $F(x) = P(X \leq x)$
z.B. Was ist die Wahrscheinlichkeit, dass die Fahrt weniger als eine Stunde dauert?
Dass wir weniger als 2 cm Regen haben?
Dass der Meßfehler kleiner 5 km/Stunde ist?
- Eine Dichte? $P(X \leq x) = \int_{-\infty}^x f(x)dx$
Keine Wahrscheinlichkeit!
- Ein Erwartungswert? $E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx$
z.B. Wie lange dauert die Fahrt im Durchschnitt?
Wieviel Regen erwarten wir?
Sind die Messungen verzerrt?

5.5.1 Einige Anwendungsfragen

- (1) Wieviele Leute sollen in einem Aufzug erlaubt werden?
- (2) Wieviel Gepäck darf man im Flugzeug mitnehmen?
- (3) Wenn die Standards verlangen, dass 95% aller Schokoriegel mindestens soviel wiegen, wie auf der Hülle versprochen wird, wie hoch muß das Durchschnittsgewicht sein?
- (4) Teil A muß mit Teil B zusammenpassen. Wie groß dürfen die Variabilitäten sein?
- (5) Blutdruck sollte nicht zu hoch sein. Wieviele haben einen zu hohen Blutdruck?
- (6) Es wird angenommen, dass Aktienpreise lognormalverteilt sind. d.h. für einen Aktien Preis X_t

$$Y_t = \log X_t \sim N(\mu_t, \sigma_t^2)$$

wobei $\mu_t = \log X_0$. Was ist $P(a < X_t \leq b)$?

5.6 Summen von Zufallsvariablen: Faltung

5.6.1 Diskrete Zufallsvariablen — Poisson

Seien X, Y unabhängige poissonverteilte Zufallsvariablen:

$$X \sim P(\lambda_1) \text{ und } Y \sim P(\lambda_2)$$

Für $Z = X + Y$ gilt

$$\begin{aligned} P(Z = z) &= \sum_{x=0}^z P(X = x)P(Y = z - x) \\ &= e^{-\lambda_1}e^{-\lambda_2} \sum_{x=0}^z \frac{\lambda_1^x}{x!} \frac{\lambda_2^{z-x}}{(z-x)!} \\ &= \frac{e^{-(\lambda_1+\lambda_2)}}{z!} \sum_{x=0}^z \frac{z!}{x!(z-x)!} \lambda_1^x \lambda_2^{z-x} \\ &= e^{-(\lambda_1+\lambda_2)} \frac{(\lambda_1 + \lambda_2)^z}{z!} \end{aligned}$$

so dass $Z \sim P(\lambda_1 + \lambda_2)$

5.6.2 Stetige Zufallsvariablen — Normal

Seien X, Y unabhängige normalverteilte Zufallsvariablen:

$$X \sim N(\mu_1, \sigma_1^2) \text{ und } Y \sim N(\mu_2, \sigma_2^2)$$

Für $Z = X + Y$ gilt

$$\begin{aligned} f(z) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(z-x-\mu_2)^2}{2\sigma_2^2}} dx \\ &= \frac{1}{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2)}} e^{-\frac{(z-\mu_1-\mu_2)^2}{2(\sigma_1^2 + \sigma_2^2)}} \end{aligned}$$

so dass $Z \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

Wir brauchen

$$ax^2 + 2bx + c = (\sqrt{a}x + b/\sqrt{a})^2 - b^2/a + c$$

5.7.4 Auswertung von Wahrscheinlichkeiten für $N(\mu, \sigma^2)$

$$P(a \leq X < b) = \int_a^b \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

Numerisch integrieren — besser Experten überlassen.

$$P(a \leq X \leq b) = P\left(\frac{a - \mu}{\sigma} \leq Z \leq \frac{b - \mu}{\sigma}\right)$$

wo $Z \sim N(0, 1)$ die Standardnormalverteilung

Zum Beispiel, gegeben $X \sim N(i, 1)$ was ist $P(X < 0)$?

$P(Z < i)$	Excel	R
0	0.5	0.5
-1	0.1587	0.1587
-5	2.867E-07	2.867E-07
-10	7.620E-24	7.620E-24

Nach Marsaglia "Evaluating the Normal Distribution Function" ist $P(Z < -10) = 7.61985302416053E - 24$

Mit Excel bekommen wir $7.619853E - 24$

Und R gibt uns $7.61985302416053E - 24$

Beispiele von Zufallsvariablen

(1) In einer klinischen Studie wird eine neue Behandlung (n_1 Patienten) mit einer alten (n_2 Patienten) verglichen. Sei p_1 bzw. p_2 die Wahrscheinlichkeit einer Besserung. Was ist die Wahrscheinlichkeit, dass mindestens d mehr Patienten mit der neuen Behandlung besser werden? Z.B. mit $n_1 = n_2 = 50$, $p_1 = 0.4$, $p_2 = 0.6$ und $d = 10$.

(2) Ernste Nebenwirkungen eines Medikaments sind angeblich selten. Gegeben eine Häufigkeit von 10 pro 100000, was ist die Wahrscheinlichkeit, 5 oder mehr Fälle in einer Gruppe von 20000 zu beobachten?

(3) Im Schach werden Spielern Elo Ratings gegeben. In der einfachen Form nimmt man an, dass die Leistungen eines Spielers normalverteilt sind, mit Rating als Erwartungswert und Standardabweichung 200. Was ist die Wahrscheinlichkeit, dass ein Spieler mit Rating 2000 eine Leistung besser als 2200 bringt? Was ist die Wahrscheinlichkeit, dass dieser Spieler gegen einen Spieler mit Rating 2200 gewinnt?

(4) Die Zahl 28 ist seit dem 23. Januar 2010 unter den sechs Lotto Gewinnzahlen nicht mehr gezogen. Wie wahrscheinlich ist das? Jedes Jahr gibt es 104 Ziehungen. Was sind der Erwartungswert und die Standardabweichung für die Anzahl der Ziehungen mit der Zahl 28?

(5) Die Lebensdauer von Geräten wird oft mit Exponentialverteilungen modelliert. Gegeben, dass man zwei Geräte in Parallel laufen läßt, was ist die Wahrscheinlichkeit, dass beide vor Zeitpunkt T nicht mehr am Leben sind? Z.B. für $\lambda = 0.01$ und $T = 200$.

6. Datenanalyse

- Welche Information enthält ein Datensatz?
Wie sind die Variablenwerte verteilt?
Gibt es ungewöhnliche (falsche?) Werte?
Welche Assoziationen gibt es zwischen Variablen?
Welche Strukturen gibt es?
- Wie kann man die Daten zusammenfassen?
Mit Statistiken
- Wie kann man die Daten darstellen?
Mit Graphiken

6.1 Beschreibende Statistiken

Gegeben ein Datensatz

mit n Fällen und m Variablen $\{X_1, X_2, \dots, X_m\}$

X_{ij} ist der Wert von Fall i für Variable j

6.1.1 Wo sind die Daten?

- Mittelwert: Durchschnitt

$$\bar{X}_j = \frac{1}{n} \sum_{i=1}^n X_{ij}$$

z.B. Messfehler, Kosten, Temperatur

- Median: der mittlere Wert
“50% der Beobachtungen sind größer, 50% kleiner”
Die Daten werden aufsteigend sortiert:
 $x_{(1)}$ ist der kleinste Wert, $x_{(n)}$ ist der größte

$$\text{median}(X_j) = x_{(\frac{n+1}{2})} \text{ für } n \text{ ungerade}$$

$$\text{median}(X_j) = \frac{1}{2}(x_{(\frac{n}{2})} + x_{(\frac{n+2}{2})}) \text{ für } n \text{ gerade}$$

z.B. Einkommen

- Modus: der häufigste Wert
z.B. Anzahl Kinder, Krankheitsdauer

6.1.2 Wie variabel sind die Daten?

- Minimum und Maximum

$$\min X = x_{(1)} \quad \max X = x_{(n)}$$

- Spannbreite:

$$= \max X - \min X = x_{(n)} - x_{(1)}$$

- Varianz

$$V[X_j] = \frac{1}{n-1} \sum_{i=1}^n (X_{ij} - \bar{X}_j)^2$$

- Standardabweichung

$$sd[X_j] = \sqrt{V[X_j]}$$

6.2 Graphiken für stetige Variablen

6.2.1 Histogramme

Die Werte der Variable X werden in nicht überlappende Klassen aufgeteilt:

$$a < a + d < a + 2d \dots < a + Kd$$

wo $a \leq x_{(1)}$ und $x_{(n)} \leq a + Kd$

Klassenhäufigkeiten werden als Rechtecke gezeichnet.

Hauptparameter: Ankerpunkt a und Klassenbreite d

Ziele:

Genug Klassen, aber nicht zuviele

Sinnvolle Klassengrenzen

Formatierungsparameter:

Skalierung, Größe, Seitenverhältnis

linksoffene oder rechtsoffene Klassen

Beschriftung, Legende, Farben/Schattierung

Fragen zu Histogramme

- Gibt es Sonderwerte? (Fehler, Ausreißer...)
- Ist die Verteilung symmetrisch?
- Gibt es mehrere Modi?
- Gibt es Gruppen/Cluster?
- Gibt es Lücken/Täler?
- Gibt es Häufungen/bevorzugte Werte?
- Andere Muster oder Struktur?

6.2.2 Streudiagramme

Darstellungen von zwei stetigen Variablen.

Eine Variable wird als Y (vertikale Achse) und eine als X (horizontale Achse) genommen.

Formatierungsparameter:

Punkt-Größe und -Farbe

Skalierung, Größe, Seitenverhältnis

Beschriftung, Legende

Fragen zu Streudiagramme

- Gibt es Sonderwerte? (Fehler, Ausreißer...)
- Sind die Variablen assoziiert?
- Gibt es Gruppen/Cluster?
- Gibt es Lücken?
- Andere Muster oder Struktur?

6.3 Graphiken für kategorielle Daten

6.3.1 Säulendiagramme

Es gibt eine Säule für jede Kategorie und Lücken zwischen den Säulen.

Alle Säulen sind gleich breit. Die Säulenhöhe stellt die Anzahl der Fälle in der Kategorie dar.

Formatierungsparameter:

Größe und Seitenverhältnis

Skalierung und Abstand zwischen den Säulen

Farbe und Schattierung

Beschriftung, Legende

Fragen zu Säulendiagramme

- Gibt es zuviele Kategorien (Fehler, ...)?
- Welche Kategorien sind am häufigsten vertreten?
- Welche Kategorien kommen kaum vor?
- Gibt es eine aufschlussreiche Reihenfolge der Kategorien?
- Andere Muster oder Struktur?

6.4 Datensätze

6.4.1 Noten aus einer irischen Schule

126 Schüler und Schülerinnen haben Klausuren in bis zu 9 Fächern geschrieben. Jede Note liegt zwischen 0 und 100 Punkte. Um nicht durchzufallen, muss man mindestens 40 Punkte bekommen. Die Namen sind geändert worden.

6.4.2 Deutsche Wahlergebnisse

Die Wahlergebnisse (Stimmen und Prozente pro Partei) für 2005 und 2009 für die 299 Wahlkreise.

6.4.3 Zehnkampf 2006

Die Leistungen und Punktzahl für die Bestleistungen aller Zehnkämpfer, die im Jahr 2006 mindestens einmal 6800 Punkte insgesamt erreicht haben.

6.4.4 Körpermessungen

24 Körpermessungen von 247 Männern und 260 Frauen.

K7 Grenzwertsätze

7.1 Tschebychew Ungleichung

Sei X eine stetige ZV mit endlicher Varianz, für $\epsilon > 0$

$$P(|X - E[X]| \geq \epsilon) \leq \frac{V[X]}{\epsilon^2} \quad (4)$$

Beweis:

Setzen wir $Z = (X - E[X])$, dann ist $E[Z] = 0$ und $V[Z] = V[X]$. Definieren wir weiter

$$W = 0 \quad \text{für } |Z| < \epsilon \quad \text{und} \quad W = \epsilon^2 \quad \text{für } |Z| \geq \epsilon$$

$$\Rightarrow W \leq |Z|^2$$

$$V[X] = V[Z] = E[|Z|^2] \geq E[W]$$

$$= 0 * P(|Z| < \epsilon) + \epsilon^2 * P(|Z| \geq \epsilon)$$

$$\Rightarrow P(|Z| \geq \epsilon) = P(|X - E[X]| \geq \epsilon) \leq \frac{V[X]}{\epsilon^2}$$

7.2 Tschebychew: Bemerkungen

- Tschebychew kann keine gute Annäherung sein, da sie für alle ZV gilt. Sie ist aber in der Theorie wichtig.
- Für $\epsilon < \sigma_X$ ist Tschebychew wertlos.
- Für unimodale Dichten gilt das Resultat

$$P(|X - \mu| \geq \epsilon) \leq \frac{4\sigma^2}{9\epsilon^2} \quad \forall \epsilon \geq \sqrt{\frac{8}{3}}\sigma$$
$$\leq \frac{4\sigma^2}{3\epsilon^2} - \frac{1}{3} \quad \forall \epsilon \leq \sqrt{\frac{8}{3}}\sigma$$

So dass für $\epsilon = 3\sigma$ hat man

$$P(|X - \mu| \geq 3\sigma) \leq \frac{4}{81} < 0.05$$

(cf. Pukelsheim American Statistician 1994 s88-91)

7.3 Tschebychew: Beispiele

Beispiel (1)

$$X \sim N(\mu, \sigma^2)$$

nach Tscheb $P(|X - \mu| \geq 2\sigma) \leq \frac{\sigma^2}{4\sigma^2} = 0.25$

unimodal $P(|X - \mu| \geq 2\sigma) \leq \frac{1}{9} = 0.11$

In der Tat $P(|X - \mu| \geq 2\sigma) = 0.0456$

Beispiel (2)

$$X \sim E(\lambda)$$

nach Tscheb $P\left(\left|X - \frac{1}{\lambda}\right| \geq 2\frac{1}{\lambda}\right) \leq \frac{\lambda^2}{4\lambda^2} = 0.25$

unimodal $P\left(\left|X - \frac{1}{\lambda}\right| \geq 2\frac{1}{\lambda}\right) \leq \frac{1}{9} = 0.11$

In der Tat $P\left(\left|X - \frac{1}{\lambda}\right| \geq 2\frac{1}{\lambda}\right) = P\left(X \geq \frac{3}{\lambda}\right) = 0.0498$

Beispiel (3)

$$X \sim P(\lambda)$$

nach Tscheb $P(|X - \lambda| \geq 2\sqrt{\lambda}) \leq \frac{\lambda}{4\lambda} = 0.25$

In der Tat $P(|X - \lambda| \geq 2\sqrt{\lambda})$ hängt von λ ab

$$(a) \lambda = 1 \Rightarrow P(|X - 1| \geq 2) = P(X \geq 3) =$$

$$(b) \lambda = 4 \Rightarrow P(|X - 4| \geq 4) = P(X \leq 0) + P(X \geq 8)$$

=

$$(c) \lambda = 9 \Rightarrow P(|X - 9| \geq 6) = P(X \leq 3) + P(X \geq 15)$$

=

7.4 Konvergenz

(1) In Verteilung

$$\text{für } n \rightarrow \infty \quad P(X_n \leq x) \rightarrow P(X \leq x)$$

(2) In Wahrscheinlichkeit

$$\text{für } n \rightarrow \infty \quad P(|X_n - X| > \epsilon) \rightarrow 0 \quad \forall \epsilon > 0$$

(3) Fast sicher

$$\text{für } n \rightarrow \infty \quad P(X_n(\omega) \rightarrow X(\omega) \quad \forall \omega \in \Omega) = 1$$

$$(3) \Rightarrow (2) \Rightarrow (1)$$

z.B.

Sei X eine ZV mit $P(X = 0) = P(X = 1) = 0.5$
und $X_i = X \quad \forall i$. $X_n \rightarrow X$ in Verteilung und für $Y = 1 - X$ gilt auch $X_n \rightarrow Y$ in Verteilung. Aber $|X_n - Y| = 1$ immer, so dass (2) und (3) nicht gelten können.

7.5 Das schwache Gesetz der großen Zahlen

Für das schwache Gesetz der großen Zahlen gilt Konvergenz in Wahrscheinlichkeit (stochastische Konvergenz).

Seien X_1, X_2, \dots, X_n unabhängige ZV mit gleichem Erwartungswert und Varianzen

$$V[X_i] \leq M < \infty \quad \forall i$$

Dann gilt $\forall \epsilon > 0$

$$P \left(\left| \frac{1}{n} \sum_{i=1}^n X_i - E[X] \right| \geq \epsilon \right) \leq \frac{M}{n\epsilon^2} \rightarrow 0$$

Beweis

Sei

$$Y_n = \frac{1}{n} \sum_{i=1}^n X_i$$

$$E[Y_n] = E[X]$$

$$V[Y_n] = \frac{1}{n^2} \sum_{i=1}^n V[X_i] \leq \frac{M}{n}$$

Dann wendet man Tschebychew an.

7.6 Das starke Gesetz der großen Zahlen

Sei $\{X_i\}$ eine Folge von reellwertigen unkorrelierten ZV und

$$V[X_i] \leq M < \infty \quad \forall i$$

dann konvergiert die Folge

$$Z_n = \frac{1}{n} \sum_{i=1}^n (X_i - E[X_i])$$

fast sicher gegen 0.

7.7 Der Zentrale Grenzwertsatz

Sei X_1, X_2, \dots eine Folge von reellwertigen, u.i.v. ZV mit

$$E[X_i] = \mu < \infty \quad \text{und} \quad V[X_i] = \sigma^2 < \infty$$

und sei

$$S_n = \sum_{i=1}^n X_i$$

Dann gilt, dass

$$U_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} \rightarrow N(0, 1)$$

in Verteilung.

7.8 Anwendungen des zentralen Grenzwertsatzes

(1) Binomial (De Moivre)

$$X \sim B(N, p)$$

$$N \rightarrow \infty \quad \Rightarrow \quad X \sim N(Np, Npq)$$

(2) Poisson

$$X \sim P(\lambda)$$

$$\lambda \rightarrow \infty \quad \Rightarrow \quad X \sim N(\lambda, \lambda)$$

(3) In der Statistik

Seien $\{X_i\}$ u.i.v. ZV mit $E[X] = \mu$ und $V[X] = \sigma^2$, dann gilt

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

7.9 Grenzwert-Resultate Beispiele

7.9.1 Gleichverteilung

$$X_i \sim U(0, 1) \quad \text{u.i.v.}$$

$$E[X_i] = \frac{1}{2} \quad V[X_i] = \frac{1}{12}$$

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \quad E[\bar{X}_n] = \frac{1}{2} \quad V[\bar{X}_n] = \frac{1}{12n}$$

Tschebychew:

$$P\left(\left|\bar{X}_n - \frac{1}{2}\right| \geq \epsilon\right) \leq \frac{1}{12n\epsilon^2}$$

Schwaches Gesetz:

$$P\left(\left|\bar{X}_n - \frac{1}{2}\right| \geq \epsilon\right) \rightarrow 0 \text{ für } n \rightarrow \infty$$

Starkes Gesetz:

$$P\left(\left|\bar{X}_n - \frac{1}{2}\right| \geq \epsilon \quad \forall n > N_\epsilon\right) \rightarrow 0 \text{ für } n \rightarrow \infty$$

Zentraler Grenzwert Satz:

$$\begin{aligned} P\left(\left|\bar{X}_n - \frac{1}{2}\right| \geq \epsilon\right) &\rightarrow 2 \int_{\frac{\epsilon}{\sqrt{12n}}}^{\infty} \frac{\sqrt{12n}}{\sqrt{2\pi}} e^{-6nt^2} dt \\ &= 2(1 - \Phi(\sqrt{12n}\epsilon)) \end{aligned}$$

7.9.2 Summen und Mittelwerte

Seien $\{X_i\}$ u.i.v. ZV mit $E[X] = \mu$ und $V[X] = \sigma^2$,
dann gilt für $n \rightarrow \infty$

$$S_n = \sum_{i=1}^n X_i \sim N(n\mu, n\sigma^2)$$
$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

So dass

$$P(-1.96 * \sqrt{n}\sigma < S_n - n\mu < 1.96 * \sqrt{n}\sigma) = 0.95$$

d.h., der Abstand von S_n zu $n\mu$ wächst

$$P\left(-1.96 * \frac{\sigma}{\sqrt{n}} < \bar{X}_n - \mu < 1.96 * \frac{\sigma}{\sqrt{n}}\right) = 0.95$$

d.h., der Abstand von \bar{X}_n zu μ schrumpft.

K 8 Schätzen

Was ist der Durchschnittseinkommen einer Familie?

Welcher Prozentsatz der Bevölkerung lebt unter der Armutsgrenze?

Welchen Marktanteil hat der X-Box?

Welcher Prozentsatz der Wähler unterstützt die SPD?

Was war die Inflationsrate für April?

Wieviele Leute sterben an Lungenkrebs?

Wie zuverlässig sind Volkswagen Autos?

Wieviele Leute wohnen in Deutschland?

Wieviel Bier trinkt ein Bayer?

Wir wissen nicht, wir müssen schätzen

Theorie

Von einer Stichprobe zu einer Verteilung

Die Verteilung hat Parameter θ

θ_j wird durch die Funktion

$$\hat{\theta}_j(X_1, X_2, \dots, X_n)$$

geschätzt. ($\hat{\theta}$ bedeutet Schätzer)

d.h. Aufgrund der Beobachtungen (x_1, x_2, \dots, x_n) von Zufallsgrößen (X_1, X_2, \dots, X_n) soll eine "möglichst gute" Schätzung eines unbekanntes Parameters θ_j der Verteilungen der X_i angegeben werden.

Wir werden meistens o.B.d.A. nur einen Parameter schätzen.

8.1 Eigenschaften von Schätzern: Was heißt “gut”?

(a) Erwartungstreue

$$E[\hat{\theta}] = \theta$$

Beispiel 1

$$E[\bar{X}] = \frac{1}{n} E\left[\sum X_i\right] = \mu$$

Beispiel 2

$$X \sim E(\lambda)$$

$$f(x) = \lambda e^{-\lambda x}$$

$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ ist ein erwartungstreuer Schätzer für $\frac{1}{\lambda}$.

Ist $\frac{n}{\sum X_i}$ ein erwartungstreuer Schätzer für λ ?

$$E\left[\frac{n}{\sum X_i}\right] = \int_{X_1} \cdots \int_{X_n} \frac{n}{\sum X_i} \lambda^n e^{-\lambda \sum x_i} dx_1 \cdots dx_n$$

Sogar für $n = 1$ geht es nicht:

$$E\left[\frac{1}{X}\right] = \int_0^{\infty} \frac{1}{x} \lambda e^{-\lambda x} dx$$

(b) Konsistenz (im statistischen Sinne)

$$n \rightarrow \infty \Rightarrow P(|\hat{\theta}_n - \theta| > \epsilon) \rightarrow 0$$

θ wird asymptotisch richtig geschätzt.

Die Bedingungen (a) und (b) sagen wenig über die Güte eines Schätzers aus, insbesondere bei kleineren Stichproben.

(c) Mittlere quadratische Fehler

(Je kleiner, desto besser)

$$MSE(\theta, \hat{\theta}) = E[(\hat{\theta} - \theta)^2]$$

(Mean Square Error)

$$= E[\hat{\theta}^2] - 2\theta E[\hat{\theta}] + \theta^2$$

$$= E[\hat{\theta}^2] - E[\hat{\theta}]^2 + E[\hat{\theta}]^2 - 2\theta E[\hat{\theta}] + \theta^2$$

$$= V[\hat{\theta}] + (E[\hat{\theta}] - \theta)^2$$

$$MSE = \text{Varianz} + \text{Bias}^2$$

Für einen erwartungstreuen Schätzer gilt

$$MSE = \text{Varianz}$$

8.2 Herleitung von Schätzern

8.2.1 Methode der kleinsten Quadrate

Ein Schätzungsziel wäre die Minimierung von

$$MSE = E[(\hat{\theta} - \theta)^2]$$

wo θ uns unbekannt ist und wir das Minimum über alle möglichen Funktionen suchen. Das wird zu schwierig. Deshalb diese Methode.

Beobachtungen (x_1, \dots, x_n) und

$$E[X_i] = g_i(\theta)$$

$g_i(\theta)$ ist eine Prädiktorfunktion für X_i und wir wählen $\hat{\theta}$, um

$$Q = \sum_{i=1}^n (x_i - g_i(\theta))^2$$

zu minimieren.

$$\frac{\partial Q}{\partial \theta} = 0$$

$$\Rightarrow \sum_{i=1}^n (x_i - g_i(\theta)) \frac{\partial g_i}{\partial \theta} = 0$$

Beispiele

(1) Exponentialverteilung

$$X \sim E(\lambda)$$

$$\theta = \lambda$$

$$E[X_i] = \frac{1}{\lambda}$$

$$Q = \sum (x_i - \frac{1}{\lambda})^2$$

$$\frac{\partial Q}{\partial \lambda} = 0 \quad \Rightarrow \quad -\frac{2}{\lambda^2} \sum (x_i - \frac{1}{\lambda}) = 0$$

$$\frac{n}{\lambda} = \sum x_i \quad \Rightarrow \quad \hat{\lambda} = \frac{n}{\sum x_i}$$

(2) Ein statistisches Modell

$$Y_i = g_i(\theta; X_1, X_2, \dots) + \epsilon$$

wobei ϵ ein Meßfehler ist und $\epsilon \sim N(0, \sigma^2)$. Um $\sum \epsilon_i^2$ zu minimieren, wurden wir

$$Q = \sum_{i=1}^n (y_i - g_i(\theta))^2$$

minimieren.

(3) Erraten des Bereichs von Zufallszahlen

Zufallszahlen werden im Intervall $[0, \theta]$ erzeugt. Was ist der beste Schätzer für θ aus einer Stichprobe der Größe n , $\{u_1, \dots, u_n\}$?

$$U_i \sim U[0, \theta]$$

(a) Nach dem GGZ könnte man

$$\hat{\theta}_a = \frac{2}{n} \sum_{i=1}^n u_i$$

nehmen. $\hat{\theta}_a$ ist erwartungstreu und konsistent.

(b) Man könnte auch

$$\hat{\theta}_b = \max\{u_i\}$$

nehmen. $\hat{\theta}_b$ ist konsistent, aber nicht erwartungstreu.

$$E[\hat{\theta}_b] = \frac{n}{n+1}\theta$$

$$\text{N.B. } P(\hat{\theta}_b \leq c) = \left(\frac{c}{\theta}\right)^n$$

Vergleich der Varianzen

$$\begin{aligned}V[\hat{\theta}_a] &= \left(\frac{2}{n}\right)^2 V\left[\sum_{i=1}^n U_i\right] \\ &= \frac{4}{n} V[U_i] = \frac{\theta^2}{3n}\end{aligned}$$

$$\begin{aligned}V[\hat{\theta}_b] &= E[\hat{\theta}_b^2] - E[\hat{\theta}_b]^2 \\ &= \frac{n}{n+2}\theta^2 - \frac{n^2}{(n+1)^2}\theta^2 \\ &= \frac{n\theta^2}{(n+1)^2(n+2)}\end{aligned}$$

MSE von $\hat{\theta}_b$ ist

$$\frac{n\theta^2}{(n+1)^2(n+2)} + \left(\frac{n}{n+1}\theta - \theta\right)^2 = \frac{2\theta^2}{(n+1)(n+2)}$$

Besser wird es, den Schätzer

$$\hat{\theta}_c = \frac{n+1}{n} \max\{u_i\}$$

zu nehmen. $\hat{\theta}_c$ ist erwartungstreu, konsistent und hat die Varianz

$$\frac{\theta^2}{n(n+2)}$$

Für $n=10$, sind die MSEs

$$\frac{\theta^2}{30}, \quad \frac{\theta^2}{66}, \quad \frac{\theta^2}{120}$$

Ist $\hat{\theta}_c$ der beste Schätzer überhaupt?

8.2.2 Maximum Likelihood Schätzer (ML)

Das Likelihood einer vorliegenden Stichprobe

$$\underline{x} = (x_1, x_2, \dots, x_n)$$

unter einem bestimmten Modell mit multivariaten Dichte f und Parametervektor $\underline{\theta}$ ist

$$L(\underline{x}; \underline{\theta}) = f_{X_1, \dots, X_n}(\underline{x}; \underline{\theta})$$

Likelihood ist eine Funktion von $\underline{\theta}$ mit "Parametern" \underline{X} . Die gemeinsame Dichte $f_{\underline{X}}$ ist eine Funktion von \underline{X} mit Parametern $\underline{\theta}$.

Mit $f_{\underline{X}}$ geht es um die Wahrscheinlichkeit von Daten gegeben das Modell:

$$f_{\underline{X}}(\underline{x}; \underline{\theta}) = W(\text{Daten} | \text{Modell})$$

Mit $L(\underline{x}; \underline{\theta})$ geht es um das Likelihood des Modells gegeben die Daten:

$$L(\underline{x}; \underline{\theta}) = L(\text{Modell} | \text{Daten})$$

ML Schätzer maximieren das Likelihood.

(Vorsicht: Likelihoods sind keine Wahrscheinlichkeiten!)

Beispiele

(1) Exponentialverteilung

$$X \sim E(\lambda)$$

$$L(x_1, x_2, \dots, x_n) = \lambda^n e^{-\lambda \sum x_i}$$

Ein ML Schätzer maximiert L oder, der Einfachheit halber,

$$\log L$$

was auch mathematisch gleich ist.

$$\log L = n \log \lambda - \lambda \sum x_i$$

$$\frac{\partial \log L}{\partial \lambda} = 0$$

$$\Rightarrow \frac{n}{\lambda} - \sum x_i = 0$$

$$\hat{\lambda} = \frac{n}{\sum x_i}$$

(2) Krankheitsdauer

Angenommen

$$T \sim E(\lambda)$$

$$f(t) = \lambda e^{-\lambda t}$$

Aus einer Studie haben wir die Beobachtungen t_1, \dots, t_j und $(n - j)$ Fälle mit $T_i > a_i$

$$L(\underline{t}, \lambda) = \lambda^j e^{-\lambda \sum_{i=1}^j t_i} \prod_{i=j+1}^n e^{-\lambda a_i}$$

$$\log L = j \log \lambda - \lambda \sum_{i=1}^j t_i - \lambda \sum_{i=j+1}^n a_i$$

$$\frac{\partial \log L}{\partial \lambda} = \frac{j}{\lambda} - \sum_{i=1}^j t_i - \sum_{i=j+1}^n a_i = 0$$

$$\Rightarrow \hat{\lambda}_1 = \frac{j}{\sum_{i=1}^j t_i + \sum_{i=j+1}^n a_i}$$

Falls die Fälle mit unbekanntem Endpunkten vernachlässigt werden, bekommen wir

$$\hat{\lambda}_2 = \frac{j}{\sum_{i=1}^j t_i}$$

Oder wir nehmen an, dass unbekanntes $t_i = a_i$:

$$\hat{\lambda}_3 = \frac{n}{\sum_{i=1}^j t_i + \sum_{i=j+1}^n a_i}$$

Beide Schätzer sind offensichtlich schlechter als $\hat{\lambda}_1$, weil sie die verfügbare Information nicht voll ausnutzen.

8.2.3 Eigenschaften von ML Schätzern

ML Schätzer sind nicht immer erwartungstreu. z.B.

$$X \sim N(\mu, \sigma^2)$$

$$\log L = l(\mu, \sigma) = -n \log \sigma - \frac{n}{2} \log 2\pi - \frac{1}{2\sigma^2} \sum_{i=1}^n (X_i - \mu)^2$$

$$\frac{\partial l}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu)$$

$$\frac{\partial l}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^n (X_i - \mu)^2$$

Die ML Schätzer sind $\hat{\mu} = \bar{X}$ und

$$\hat{\sigma} = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2}$$

Es kann gezeigt werden, dass ML Schätzer konsistent sind.

8.2.4 Weitere ML Beispiele

(1) Gleichverteilung

$$f(x) = \frac{1}{b-a} \quad a \leq x \leq b$$

$$L(\underline{x}; a, b) = \left(\frac{1}{b-a} \right)^n$$

u.d.B. $b \geq \max x_i$ und $a \leq \min x_i$

Das Maximum kann man mit Lagrange Methoden finden:

$$\hat{a} = \min x_i \quad \text{und} \quad \hat{b} = \max x_i$$

(2) Negativ Binomial Verteilung

$$p(n) = \binom{n-1}{r-1} p^r q^{n-r}$$

Daten n_1, n_2, \dots, n_m

(a) r bekannt

$$L = \prod \binom{n_i-1}{r-1} p^r q^{n_i-r}$$

$$\log L = \sum \log \binom{n_i-1}{r-1} + r \sum \log p + \sum (n_i-r) \log(1-p)$$

$$\frac{\partial \log L}{\partial p} = \frac{rm}{p} - \frac{\sum n_i - rm}{1-p}$$

$$\Rightarrow \hat{p} = \frac{rm}{\sum n_i}$$

(b) r unbekannt

r muß ganzzählig sein und $1 \leq r \leq \min n_i$

Wir berechnen $\max L$ für $r = 1, \dots, \min n_i$ und nehmen den größten.

8.2.5 ML Intervallschätzer

Konfidenz Intervalle (KIs)

Bis jetzt haben wir nur Punktschätzer besprochen (d.h. es wird nur einen Schätzwert angegeben. Unsere Unsicherheit um den Wert wird nicht berichtet.) Jetzt führen wir Konfidenzintervalle ein — Intervalle, die mit einer bestimmten Wahrscheinlichkeit den wahren, unbekanntem Wert enthalten.

Falls wir die Verteilung unseres Schätzers kennen, können wir vor der Ziehung Vorhersageintervalle berechnen:

$$P_{\hat{\theta}}(a < \hat{\theta} < b) = 1 - \alpha$$

$$\text{z.B. } \bar{X} \sim N(\mu, \sigma^2/n)$$

$$P_{\bar{X}}\left(\mu - 1.96 \frac{\sigma}{\sqrt{n}} < \bar{X} < \mu + 1.96 \frac{\sigma}{\sqrt{n}}\right) = 0.95$$

Diese Wahrscheinlichkeit kann umgeschrieben werden:

$$P_{\bar{X}}\left(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95$$

aber es handelt sich als Wahrscheinlichkeitsintervall immer noch um \bar{X} und nicht um μ .

Wenn wir den Stichprobenwert \bar{x} für \bar{X} einsetzen, bekommen wir ein sogenanntes Konfidenzintervall, und wir sagen, dass

$$\left(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}\right)$$

ein 95% Konfidenzintervall für μ ist.

Wenn wir viele Stichproben ziehen und Konfidenzintervalle aus jeder berechnen, werden wir finden, dass ungefähr 95% der Intervalle den wahren Wert μ enthalten werden.

Falls σ unbekannt ist aber wir $X \sim N(\mu, \sigma^2)$ annehmen dürfen, dann können wir mit $t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$ arbeiten.

Weitere Beispiele von KIs

(1) Gleichverteilung

$$X \sim G(0, b)$$

Wir beobachten $x = 7$, was ist eine 95% KI für b ?

$$P_X(0 < X < 0.95b) = 0.95 \quad \Rightarrow \quad \left(\frac{x}{0.95}, \infty \right)$$

$$P_X(0.05b < X < b) = 0.95 \quad \Rightarrow \quad (x, 20x)$$

$$P_X(0.025b < X < 0.975b) = 0.95 \quad \Rightarrow \quad \left(\frac{x}{0.975}, 40x \right)$$

Meistens versuchen wir, das kürzeste Intervall zu finden.
In diesem Fall haben wir die Intervalle

$$(7.37, \infty) \quad (7, 140) \quad (7.18, 280)$$

und wir würden das zweite nehmen.

(2) Proportionen — Binomial Verteilung

$$X \sim B(N, p)$$

Für N groß gilt

$$Y = \frac{X}{N} \sim N\left(p, \frac{p(1-p)}{N}\right)$$

und ein 95% Vorhersageintervall für Y wäre

$$P_Y\left(p - 1.96\sqrt{\frac{p(1-p)}{N}} < Y < p + 1.96\sqrt{\frac{p(1-p)}{N}}\right)$$

$$P_Y\left((Y - p)^2 < 1.96^2 \frac{p(1-p)}{N}\right) = 0.95$$

Y und N sind bekannt. Für welche Werte von p gilt

$$(Y - p)^2 < 1.96^2 \frac{p(1-p)}{N} \quad ?$$

Drei Möglichkeiten

(a) Als Approximation für die Varianz setzen wir $p = 0.5$ und

$$\frac{p(1-p)}{N} \leq \frac{1}{4N} \quad \forall p \in [0, 1]$$

$$\Rightarrow P_Y\left(Y - 1.96\frac{1}{\sqrt{4N}} < p < Y + 1.96\frac{1}{\sqrt{4N}}\right) \geq 0.95$$

(b) “Plug-in Schätzer”

Wir benutzen $\frac{y(1-y)}{N}$ als Approximation für die Varianz

$$P_Y \left(Y - 1.96 \sqrt{\frac{y(1-y)}{N}} < p < Y + 1.96 \sqrt{\frac{y(1-y)}{N}} \right) \approx 0.95$$

Dieser Schätzer ist besser als (a). Er ist am Rand (p nahe 0 oder 1) schlecht und die Coverage (Überdeckung) ist wegen der Normalapproximation unregelmäßig.

(c) Eine genauere Methode

Lösen wir die quadratische Gleichung

$$(y - p)^2 = 1.96^2 \frac{p(1-p)}{N}$$

um Wurzeln p_1 und p_2 zu finden

$$\Rightarrow P_Y(p_1 < p < p_2) = 0.95$$

(Für eine eingehende Studie der vielen Alternativen, siehe Brown L, Cal, T.T., DasGupta, A. (2001) Interval Estimation for a Binomial Proportion. *Statistical Science* 16: 101-133)

Beispiele von Binomial KIs

(i) (Die Zeit 13. Februar, 2003 s30)

Paare sind auf Flughäfen und Bahnhöfen beobachtet worden (vermutlich in Deutschland, da ein Bochumer Wissenschaftler die Daten gesammelt hat). Aus 124 küssenden Paaren haben 80 den Kopf beim Küssen nach rechts und 44 den Kopf nach links gedreht.

$$\hat{p} = 80/124 = 0.645$$

95% KIs

(a) Approx $p = 0.5$ (0.557, 0.733)

(b) "Plug-in" $p = \hat{p}$ (0.561, 0.729)

(c) Normal genau (0.558, 0.724)

(ii) Sonntagsfrage in der politischen Meinungsforschung
z.B. 1000 Befragte CSU 55%, SPD 25%, FDP 3%

(a) (0.519, 0.581) (0.219, 0.281) (-0.001, 0.061)

(b) (0.519, 0.581) (0.223, 0.277) (0.019, 0.041)

(c) (0.519, 0.581) (0.224, 0.278) (0.021, 0.043)

Browser Statistiken

Schätzungen von HitsLink, netapplications.com (April 2006):

Internet Explorer 83,88%

Mozilla Firefox 10,68%

Schätzungen von OneStat.com (Mai 2006):

Internet Explorer 85,17%

Mozilla Firefox 11,79%

“Wie bei anderen Erhebungen erklären sich die Unterschiede aber nicht nur aus dem Zufallsfehler der Stichprobe, sondern auch durch unterschiedliche Erhebungsmethoden. Während Net Applications die Statistiken von 40.000 besuchten Webseiten auswertet, gibt OneStat an, zwei Millionen Pageimpressions in 20.000 Webseiten-Besuchen aufzudröseln.” (macnews.de 12.5.06)

Welche KIs würden Sie berechnen?

8.3 Bayes Schätzer

Wir nehmen an, dass wir unsere Unsicherheit über unbekannte Parameter mit einer Wahrscheinlichkeitsdichte beschreiben können — die apriori Dichte

$$p(\underline{\theta})$$

Die apriori Dichte widerspiegelt unser Wissen/Glauben/Vorurteile über unbekannte Parameter. Nach dem ZDF Politbarometer war die Unterstützung der CDU/CSU in der Sonntagsfrage vom 18 Juni 2010 33%. Was sollte man als apriori Verteilung dieses Parameters für eine Umfrage morgen nehmen? Könnte es sein, dass Mitarbeiter von Frau Merkel und Herrn Gabriel andere apriori Verteilungen vorschlagen würden?

Die Formel von Bayes wird angewendet, um die Resultate einer Stichprobe (neue Informationen) auszuwerten und eine aposteriori Dichte zu finden.

$$\begin{aligned} p(\underline{\theta}|\underline{x}) &= \frac{f(x_1, \dots, x_n|\underline{\theta})p(\underline{\theta})}{\int f(x_1, \dots, x_n|\underline{\theta})p(\underline{\theta})d\underline{\theta}} \\ &= \frac{f(x_1, \dots, x_n|\underline{\theta})p(\underline{\theta})}{f(\underline{x})} \end{aligned}$$

$$= \frac{L_{\theta}(\underline{\theta}, \underline{x})p(\underline{\theta})}{\int L_{\theta}(\underline{\theta}, \underline{x})p(\underline{\theta})d\underline{\theta}}$$

Apriori Glauben + Information \rightarrow aposteriori Glauben

und für eine neue Stichprobe \underline{y} gilt die Vorhersage

$$f(\underline{y}|\underline{x}) = \int f(\underline{y}|\underline{\theta})p(\underline{\theta}|\underline{x})d\underline{\theta}$$

Falls neue Daten \underline{y} beobachtet werden, würden wir die alte aposteriori Verteilung als apriori nehmen und eine aktualisierte aposteriori Verteilung berechnen.

$$p(\underline{\theta}|\underline{y}) \propto f(\underline{y}|\underline{\theta})p(\underline{\theta}|\underline{x})$$

8.3.1 Bayes Beispiel mit der Binomial Verteilung

$$X \sim B(N, p)$$

Unter der Annahme der apriori Dichte

$$p \sim G(0, 1)$$

ist die aposteriori Dichte

$$f(p|x) = \frac{\binom{N}{x} p^x (1-p)^{N-x} * 1}{\int_0^1 \binom{N}{x} p^x (1-p)^{N-x} * 1 dp}$$

Da

$$\int_0^1 p^x (1-p)^{N-x} * 1 dp = \frac{\Gamma(x+1)\Gamma(N-x+1)}{\Gamma(N+2)}$$

haben wir

$$\begin{aligned} f(p|x) &= (N+1) \binom{N}{x} p^x (1-p)^{N-x} \quad 0 \leq p \leq 1 \\ &= \frac{\Gamma(N+2)}{\Gamma(x+1)\Gamma(N-x+1)} p^{(x+1)-1} (1-p)^{(N-x+1)-1} \end{aligned}$$

eine Betaverteilung mit Parametern

$$a = x + 1 \quad b = N - x + 1$$

$$E[p|x] = \frac{a}{a+b} = \frac{x+1}{N+2}$$

Wenn wir diesen Wert als unseren Schätzer für p nehmen, haben wir

$$\hat{p} = \frac{x+1}{N+2} = \frac{2}{N+2} \frac{1}{2} + \frac{N}{N+2} \frac{x}{N}$$

eine gewichtete Mischung des apriori Erwartungswert $\frac{1}{2}$ und des ML Schätzers $\frac{x}{N}$.

In diesem Fall hatten wir
apriori + Information \rightarrow aposteriori

$$G(0, 1) + B(N, p) \rightarrow \text{Beta}$$

aber die Gleichverteilung ist eine Betaverteilung mit $a = 1$ $b = 1$. Im allgemeinen gilt

$$\text{Beta}(a, b) + B(N, p) \rightarrow \text{Beta}(a', b')$$

$$a' = a + x \quad b' = b + N - x$$

$$\mu_{apriori} = \frac{a}{a + b}$$

$$\mu_{aposteriori} = \frac{a'}{a' + b'} = \frac{a + x}{a + b + N}$$

$$= \left(\frac{a + b}{a + b + N} \right) \left(\frac{a}{a + b} \right) + \left(\frac{N}{a + b + N} \right) \left(\frac{x}{N} \right)$$

eine gewichtete Summe von $\mu_{apriori}$ und dem ML Schätzer.

Die Betaverteilung ist eine konjugierte apriori Familie für die Binomialverteilung.

Statt den aposteriori Erwartungswert als Schätzer zu nehmen, könnten wir auch den Modalschätzer wählen, d.h. das Maximum aus der aposteriori Dichte. Um einen Intervallschätzer zu berechnen, benutzen wir die aposteriori Dichte und nehmen, aus verständlichen Gründen, das kürzeste 95% Intervall (was nicht immer leicht zu bestimmen sein mag):

$$P_{aposteriori}(c < \theta < d) = 0.95$$

$$\min_{c < \theta < d} f(\theta) \geq \max_{\theta \leq c, \theta \geq d} f(\theta)$$

8.3.2 Poisson Verteilung

$$Y \sim P(\theta)$$

$$p(y|\theta) \propto \theta^{\sum y_i} e^{-n\theta}$$

Eine konjugierte apriori Verteilung wäre dann

$$p(\theta) \propto \theta^{\alpha-1} e^{-\beta\theta}$$

eine $\Gamma(\alpha, \beta)$ Verteilung

$$\Rightarrow \theta|y \sim \Gamma(\alpha + n\bar{y}, \beta + n)$$

Für $n = 1$ haben wir ein interessantes Resultat:

$$\begin{aligned} p(y) &= \frac{p(y|\theta)p(\theta)}{p(\theta|y)} \\ &= \frac{e^{-\theta} \frac{\theta^y}{y!} \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta}}{\frac{(\beta+1)^{\alpha+y}}{\Gamma(\alpha+y)} \theta^{\alpha+y-1} e^{-(\beta+1)\theta}} \\ &= \frac{\Gamma(\alpha+y)}{\Gamma(\alpha)y!} \left(\frac{\beta}{(1+\beta)} \right)^\alpha \left(\frac{1}{1+\beta} \right)^y \end{aligned}$$

\Rightarrow Die Negativbinomial Verteilung kann man als eine Poisson Verteilung betrachten, wo der Parameter eine Γ apriori Dichte hat. (α ganzzahlig)

8.3.3 Bayes und die Normalverteilung

$$X \sim N(\mu, \sigma^2) \quad \sigma^2 \text{ bekannt}$$

Sei meine apriori Dichte für μ

$$\mu \sim N(\mu_0, \sigma_0^2) \quad \mu_0, \sigma_0^2 \text{ bekannt}$$

$$f(\mu|x_1, \dots, x_n) =$$

$$\frac{\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{1}{2\sigma^2} \sum (x_i - \mu)^2} \frac{1}{\sigma_0\sqrt{2\pi}} e^{-\frac{1}{2\sigma_0^2} (\mu - \mu_0)^2}}{\int_{-\infty}^{\infty} (\text{Zähler}) d\mu}$$

$$\propto e^{-\frac{1}{2\sigma^2} \sum (x_i - \mu)^2 - \frac{1}{2\sigma_0^2} (\mu - \mu_0)^2}$$

\Rightarrow die aposteriori Dichte von μ ist eine Normaldichte.

$$\mu_{\text{aposteriori}} \sim N(\mu_1, \sigma_1^2)$$

$$\sigma_1^2 = \frac{1}{\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}}$$

$$\mu_1 = \sigma_1^2 * \left(\frac{\sum x_i}{\sigma^2} + \frac{\mu_0}{\sigma_0^2} \right)$$

Die Normalverteilung ist eine konjugierte Familie für sich selbst (σ^2 bekannt).

Für den Sonderfall $\sigma_0^2 = \sigma^2$ haben wir

$$\sigma_1^2 = \frac{\sigma^2}{n+1} \text{ und } \mu_1 = \frac{n\bar{x} + \mu_0}{n+1}$$

Der apriori Erwartungswert bekommt ein Gewicht von 1 und der Stichprobenmittelwert ein Gewicht von n .

Im allgemeinen Fall mit $\sigma_0^2 \gg \frac{\sigma^2}{n}$ (d.h. große apriori Unsicherheit oder eine große Stichprobe):

$$\sigma_1^2 = \frac{\sigma^2}{n} \text{ und } \mu_1 = \bar{x}$$

$$\mu \sim N\left(\bar{x}, \frac{\sigma^2}{n}\right)$$

$$P_\mu\left(\bar{x} - 1.96 \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.96 \frac{\sigma}{\sqrt{n}}\right) = 0.95$$

ein echtes 95% Wahrscheinlichkeitsintervall, kein KI.

Für $\sigma_0^2 \ll \frac{\sigma^2}{n}$ gilt

$$\mu \sim N(\mu_0, \sigma_0^2)$$

Beim festen Glauben, ändert neue Information nichts.

8.3.4 Schwierigkeiten

Was passiert, wenn wir auch σ^2 nicht kennen? Wie sollten wir eine gemeinsame apriori Dichte für μ und σ^2 wählen?

8.3.5 Nichtkonjugierte Verteilungen

z.B.

$$X \sim P(\lambda) \quad \lambda \sim G(5, 10)$$

$$f(\lambda|x) = \frac{e^{-\lambda} \lambda^x \frac{1}{x!} \frac{1}{5}}{\int_5^{10} e^{-\lambda} \lambda^x \frac{1}{x!} \frac{1}{5} d\lambda} \quad 5 \leq \lambda \leq 10$$

In diesem Fall geht es nur mit einer direkten Auswertung. In komplexeren Fällen ist die Auswertung nicht möglich. Wenn analytische und numerische Methoden nicht klappen, simuliert man mit klugen rechnerintensiven Methoden:

MCMC Monte Carlo Markov Chain.

K9 Statistische Tests

Beispiel: Exemestane und Brustkrebs

Tamoxifen über 5 Jahre ist für gewisse Frauen mit Brustkrebs eine Standardbehandlung. Wissenschaftler haben eine neue Behandlung vorgeschlagen:

2 bis 3 Jahre Tamoxifen und danach Exemestane.

In einer Doppelblindstudie von Coombes et al sind folgende Resultate berichtet worden (NEJM 350(11) 1081-92).

4742 Patientinnen insgesamt

2362 bekamen nach 2 bis 3 Jahren Exemestane

2380 bekamen noch Tamoxifen

Tote

93 in der Exemestane Gruppe %

106 in der Tamoxifen Gruppe %

Ereignisse

266 in der Exemestane Gruppe %

449 in der Tamoxifen Gruppe %

Zeigen diese Resultate, dass die Behandlung mit Exemestane besser ist?

9.1 Statistische Tests im allgemeinen

z.B. Hilft Vitamin C gegen Erkältungen?

Ist der neue Ferrari schneller als der alte?

Sind finnische Schüler besser als deutsche?

- Was ist ein statistischer Test?
- Warum werden statistische Tests durchgeführt?
- Wann werden statistische Tests durchgeführt?
- Wie werden statistische Tests durchgeführt?
- Wer führt statistische Tests durch?
- Welche Folgen haben statistische Tests?

9.1.1 Terminologie für Test-Theorie

Hypothese

H_0 Nullhypothese

H_1 Alternativhypothese

Teststatistik (Prüfgröße)

Verteilung der Teststatistik

Testniveau α

Annahmereich/Ablehnungsbereich

Entscheidungsregel (kritischer Wert)

Statistische Signifikanz

p -Wert

Fehler Typ I $P(H_0 \text{ verwerfen} | H_0 \text{ wahr})$

Fehler Typ II $P(H_0 \text{ akzeptieren} | H_1 \text{ wahr})$

Gütefunktion (beim Test eines Parameters)

$g(\theta) = P(H_0 \text{ verwerfen} | \theta \text{ der wahre Parameter})$

Operationscharakteristik eines Tests $OC = 1 - g(\theta)$

9.1.2 Statistische Testprozeduren

Theorie

Nullhypothese H_0

Studie entwerfen

Definition der Grundgesamtheit

Stichprobenverfahren

Stichprobengröße

Daten holen

Daten erheben

Daten organisieren

Häufigkeitstabellen, Histogramme usw.

Resultate zusammenfassen

Teststatistik berechnen

Teststatistik mit erwarteten Werten vergleichen

Verteilung der Teststatistik

p -Wert, Signifikanz

9.1.3 Signifikanz der Teststatistik

Wir berechnen den p -Wert

$$P_{H_0}(|\text{Teststatistik}| > |\text{beobachtete Wert der Statistik}|)$$

(Im allgemeinen sucht man die Wahrscheinlichkeit des Wertebereichs, worin alle Werte eine kleinere Wahrscheinlichkeit als der beobachtete Wert haben.)

Traditionelle Deutung

(nach Fisher nur als Richtlinie, aber ...):

p -Wert $\geq 0.05 \Rightarrow H_0$ sollte nicht verworfen werden

p -Wert $< 0.05 \Rightarrow H_0$ kann vielleicht verworfen werden

p -Wert $< 0.01 \Rightarrow H_0$ kann verworfen werden

Vieles hängt von den nachfolgenden Wirkungen ab, wie man p -Werte interpretiert.

z.B.

neues Medikament genehmigt

oder

alte Behandlung verboten

9.2 Binomial Test (eine Stichprobe)

Nullhypothese

$$p = p_0$$

Teststatistik

Anzahl Erfolge R aus n

Verteilung unter H_0

$$P(R = r) = \binom{n}{r} p_0^r (1 - p_0)^{n-r}$$

Annahmebereich

$$[a, b] \text{ um } np_0$$

Ablehnungsbereich

$$[0, a - 1] \cup [b + 1, n]$$

Typ I Fehler (Test-Niveau)

$$P(R \notin [a, b] | H_0)$$

Das Test-Niveau wird für einen diskreten Test selten genau α sein. Eine Möglichkeit ist ein randomisierter Test: falls $R = a$ oder $R = b$, akzeptiert man H_0 mit Wahrscheinlichkeiten, die so gewählt sind, dass das Test-Niveau $= \alpha$. Das wird mathematisch genau sein, ist aber in der Praxis unzufriedenstellend.

Alternativhypothese

$$p = p_1$$

Typ II Fehler

$$P(R \in [a, b] | H_1)$$

Gütefunktion

$$g(p_1) = P(R \notin [a, b] | p = p_1)$$

Binomial Test Beispiel
Außersinnliche Wahrnehmung
Farbe einer Karte erraten
(cf. u.a. Jung und Pauli)

Nullhypothese $p = 0.5$

Teststatistik Anzahl Erfolge R aus 6

Verteilung unter H_0
$$P(R = r) = \binom{6}{r} 0.5^r (1 - 0.5)^{6-r}$$

Annahmebereich $[1, 5]$

Ablehnungsbereich $[0] \cup [6]$

Typ I Fehler (Test-Niveau)

$$P(R \notin [1, 5] | H_0) = 0.03125$$

p -Wert von $R = 5$ $P(R \notin [2, 4] | H_0) = 0.2188$

Alternativhypothese $p = p_1 = 0.8$

Typ II Fehler $P(R \in [1, 5] | H_1) = 0.74$

Gütefunktion $g(p_1) = P(R \notin [1, 5] | p = p_1)$

Binomial Test (zwei Stichproben)

Nullhypothese $p_1 = p_2$

Teststatistik $X/n_1 - Y/n_2$

Verteilung unter H_0

$$E[T] = 0, V[T] = p(1 - p)(1/n_1 + 1/n_2)$$

Annahmebereich ?

Ablehnungsbereich ?

Typ I Fehler (Test-Niveau) ?

Alternativhypothese $p_1 \neq p_2$

Typ II Fehler ?

Gütefunktion ?

Oder wir nehmen an, dass n_1, n_2 groß genug sind, dass die Normalapproximation gilt. Dann haben wir:

$$X/n_1 \sim N(p, p(1-p)/n_1) \text{ und } Y/n_2 \sim N(p, p(1-p)/n_2)$$

$$\Rightarrow T = X/n_1 - Y/n_2 \sim N(0, p(1-p)(1/n_1 + 1/n_2))$$

9.3 Normal Test (z Test, Gauß Test)

Nullhypothese

$$\mu = \mu_0$$

Teststatistik

Stichprobenmittelwert \bar{x}

Verteilung unter H_0

$$N(\mu_0, \sigma^2/n) \quad \text{für } n \text{ groß, } \sigma^2 \text{ (un)bekannt}$$

Annahmehbereich

$$\left[\mu_0 - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}, \mu_0 + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \right]$$

$$\text{wo } P(Z > z_{1-\alpha/2}) = 1 - \Phi(z_{1-\alpha/2}) = \alpha/2$$

$$\text{und } P(Z < z_{\alpha/2}) = \Phi(z_{\alpha/2}) = \alpha/2 \quad \text{für } Z \sim N(0, 1)$$

Ablehnungsbereich

$$\left(-\infty, \mu_0 - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \right) \cup \left(\mu_0 + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}, \infty \right)$$

Typ I Fehler (Test-Niveau)

$$\alpha = 1 - (\Phi(z_{1-\alpha/2}) - \Phi(z_{\alpha/2}))$$

Alternativhypothese

$$\mu = \mu_1$$

Typ II Fehler

$$P \left(\mu_0 - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} < \bar{x} < \mu_0 + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \mid \mu = \mu_1 \right)$$

$$\Phi \left((\mu_0 - \mu_1) \frac{\sqrt{n}}{\sigma} + z_{1-\alpha/2} \right) - \Phi \left((\mu_0 - \mu_1) \frac{\sqrt{n}}{\sigma} - z_{1-\alpha/2} \right)$$

Gütefunktion

$$g(\mu_1) = 1 - \text{Typ II Fehler}$$

9.4 t-Verteilung

Sei X eine Zufallsvariable mit Erwartungswert μ und Varianz σ^2 . Für n gross und σ bekannt gilt

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Für n noch grösser (was immer "gross" heissen mag) gilt es auch für σ unbekannt.

Für n klein gibt es das Resultat von Student.

Sei X eine normalverteilte Zufallsvariable mit unbekanntem Parametern μ und σ :

$$X \sim N(\mu, \sigma^2)$$

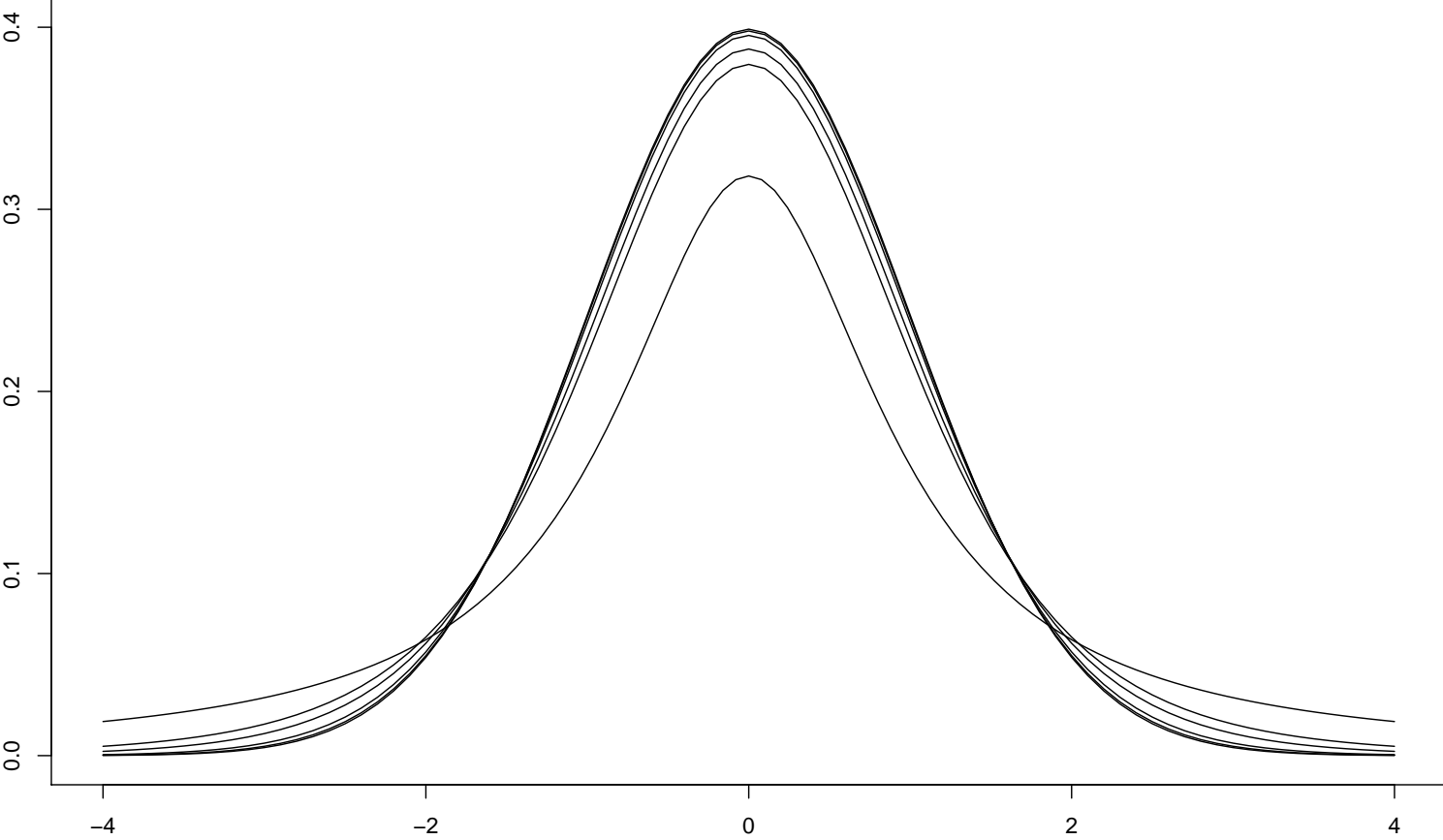
dann gilt, dass die Zufallsvariable T aus einer Stichprobe von n

$$T = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

eine t-Verteilung mit $(n - 1)$ Freiheitsgraden hat. Damit kann man Tests durchführen für kleine Stichproben, solange die Daten normalverteilt sind.

Für $n \rightarrow \infty$ wird die t-Verteilung eine Normalverteilung.

t - Dichten, df=(1,5,9,29,99,99999)



9.5 Chi-Quadrat Tests für diskrete Daten

Gegeben seien Zähldaten wie $X_i =$ Anzahl Fälle in der Klasse i . Wie testen wir, ob diese Daten mit einem Modell konsistent sein?

9.5.1 Beispiele

(1) Ist die Anzahl der Tore in einem Fußballspiel eine poissonverteilte ZV mit $\lambda = 3$? Daten aus 18 Spielen aus der 1. und 2. Bundesliga stehen zur Verfügung:

7, 9, 2, 1, 7, 3, 5, 3, 6 und 2, 3, 5, 3, 5, 2, 3, 0, 5

(2) Gibt es eine geschlechtsspezifische Abneigung gegen das Studium an der Mathematisch-Naturwissenschaftlichen Fakultät der Uni Augsburg? Im Sommersemester 1999 waren die Zahlen

	Frauen	Männer
MNF	433	693
Rest	5505	4641

Methode

Wir berechnen die unter dem Modell zu erwartenden Werten und vergleichen sie mit den beobachteten Werten.

Beispiel (1) Anzahl Tore

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

Sei e_i die erwartete Anzahl Spiele mit i Toren unter den 18 Spielen und o_i die beobachtete Zahl:

Tore	0	1	2	3	4	5	6	≥ 7
o_i	1	1	3	5	0	4	1	3
e_i	0.89	2.69	4.03	4.03	3.02	1.81	0.91	0.60

Heuristische Idee. Die (quadrierten) Abweichungen

$$(o_i - e_i)^2$$

messen den Unterschied zum Modell für eine Klasse, aber sie sind nicht gut untereinander vergleichbar. Wir wollen nicht die (o_i, e_i) Paare

$$(100, 110) \quad \text{und} \quad (10, 20)$$

gleich bewerten. Deshalb arbeiten wir mit den relativen Abweichungen

$$\frac{(o_i - e_i)^2}{e_i}$$

Wir summieren sie über alle Klassen, und verwenden die Teststatistik

$$T = \sum_i \frac{(o_i - e_i)^2}{e_i}$$

Für H_0 stellt sich heraus, dass unter gewissen Annahmen

$$T \sim \chi_7^2$$

wo 7 die Anzahl der Klassen minus 1 ("Anzahl Freiheitsgrade") ist. Wir werden H_0 verwerfen, wenn T groß ist.

$$P(T > 14.1) = 0.05 \quad P(T > 18.5) = 0.01$$

Für die Daten finden wir $T = 16.79$, so dass wir H_0 zu einem Testniveau von 5% verwerfen können, aber nicht zu einem Niveau von 1%.

Der Durchschnitt der Tore pro Spiel war $71/18 = 3.944$. Wenn wir einen Test mit diesem λ ausführen, finden wir $T = 7.03$, aber wir müssen die Benutzung des Schätzers berücksichtigen. Da ein Parameter geschätzt worden ist, verlieren wir einen Freiheitsgrad und nehmen $T \sim \chi_6^2$ an.

$$P(T > 12.6) = 0.05$$

H_0 wird in diesem Fall akzeptiert.

Beispiel (2) Fakultäten

H_0 : kein Geschlechtseinfluß — der Anteil in MNF entspricht dem Gesamtanteil der Frauen an Studierenden.

Insgesamt gibt es 5938 Frauen und 5334 Männer mit einem Frauenanteil von $\frac{5938}{11272}$. Unter H_0 erwarten wir

$$\frac{5938}{11272} * 1126$$

Frauen in MNF. Die erwarteten Werte für die Tabelle sind

	Frauen	Männer
MNF	593.17	532.83
Rest	5344.83	4801.17

Aus demselben heuristischen Argument wie vorher nehmen wir als Teststatistik

$$X^2 = \sum_i \sum_j \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$

Approximativ gilt

$$X^2 \sim \chi_{(r-1)(c-1)}^2$$

wo r die Anzahl von Zeilen ("rows") sei und c die Anzahl von Spalten ("columns"). $r = 1, c = 1 \Rightarrow X^2 \sim \chi_1^2$

$$X^2 = 101.5 \Rightarrow H_0 \text{ verwerfen.}$$

9.5.2 (Zweidimensionale) Kontingenztafeln im allgemeinen

Gegeben seien zwei Klassifikationen der Daten, $i = 1, \dots, r$ und $j = 1, \dots, c$. Diese Information kann in einer $r \times c$ Tabelle mit Einträgen

$$n_{ij} = \text{Anzahl F\u00e4lle in } i \text{ und } j$$

zusammengefasst werden. Wir interessieren uns meistens f\u00fcr die Hypothese

H_0 : keine Assoziation zwischen den Klassifikationen.

Sei

$$p_{ij} = P(\text{Zeilenklasse} = i, \text{Spaltenklasse} = j)$$

$$p_{i.} = P(\text{Zeilenklasse} = i)$$

$$p_{.j} = P(\text{Spaltenklasse} = j)$$

$$\text{dann } H_0 \Rightarrow p_{ij} = p_{i.} p_{.j}$$

Wir schätzen $p_{i.}, p_{.j}$ mit

$$\hat{p}_{i.} = \frac{n_{i.}}{n} \quad \hat{p}_{.j} = \frac{n_{.j}}{n}$$

$$\Rightarrow e_{ij} = n \left(\frac{n_{i.}}{n} \right) \left(\frac{n_{.j}}{n} \right)$$

$$= \frac{(\text{Zeilensumme } i)(\text{ Spaltensumme } j)}{n}$$

Als Teststatistik haben wir

$$X^2 = \sum_i \sum_j \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$

Für $n \rightarrow \infty$ gilt $X^2 \sim \chi^2_{(r-1)(c-1)}$.

Als Faustregel (Cochran 1952!) hat man $e_{ij} \geq 5 \quad \forall ij$,
aber das ist sicherlich zu streng.

Freiheitsgrade

es sind rc Zellen (Kombinationen) rc

n wird festgelegt -1

$p_{i.}$ werden geschätzt $-(r - 1)$

$p_{.j}$ werden geschätzt $-(c - 1)$

Es bleiben $(r - 1)(c - 1)$

$$Y = X^2 \sim \chi_k^2$$

$$f(y) = \frac{1}{\Gamma(k/2)} \left(\frac{1}{2}\right)^{k/2} y^{k/2-1} e^{-y/2} \quad x \geq 0$$

$$E[Y] = k$$

$$V[Y] = 2k$$

Für $k = 1, 2$ hat die Dichte ein Maximum bei $y = 0$.

Für $k \geq 3$ ist $f(0) = 0$ und das Maximum liegt bei $y = k - 2$.

9.5.3 Interpretationen

In 1949 hatten Kieser und Schäfer aus *Who's Who* eine Tabelle für 1436 Frauen zusammengestellt, die all mindestens einmal verheiratet worden waren.

Ausbildung	Einmal Verheiratet	Mehrmals
College	550	61
Kein College	681	144

H_0 : keine Assoziation

$$\chi^2 = 16.0 \Rightarrow H_0 \text{ verwerfen, da } \chi_{1,0.99}^2 = 6.63$$

Heißt das, dass Frauen ohne Ausbildung heirateten öfters (17% v. 10%)?

Oder, dass Frauen die öfters heirateten eher nicht zum College gingen (30% v. 45%)?

Erstens haben wir nur die Hypothese von keiner Assoziation verworfen, ohne Kausalität zu besprechen. Zweitens wird es andere Faktoren geben.

Die Daten sind keine zufällige Stichprobe aus einer Grundgesamtheit, kann man trotzdem den Test durchführen und die Resultate verallgemeinern?

Ein komplexeres Beispiel — Intelligenz und Kleidung

Gilbey (Biometrika **8** 94) hat die Verteilung von 1725 Kindern besprochen, die nach Intelligenz

A: geistig mangelhaft

B: langsam und dumm

C. dumm

D: langsam aber intelligent

E: ziemlich intelligent

F: ausgesprochen fähig

G: sehr fähig

und nach Kleidung

sehr gut gekleidet

gut gekleidet

schlecht, aber passend

sehr schlecht

klassifiziert worden sind.

Was kann man daraus lesen?

Wie könnten die Daten graphisch dargestellt werden?

Könnte man Untergruppen testen oder Kategorien kombinieren?

Table 2. *Distribution of 1725 school children according to their standard of clothing and their intelligence: Kendall & Stuart (1967, p. 558) after Gilby.*

Standard of clothing	Intelligence class							Total
	A, B	C	D	E	F	G		
Very well clad	33	48	113	209	194	39	636	
Well clad	41	100	202	255	138	15	751	
Poor but passable	39	58	70	61	33	4	265	
Very badly clad	17	13	22	10	10	1	73	
Total	130	219	407	535	375	59	1725	

9.5.4 Warum die Chi-Quadrat Verteilung?

(1) Binomial und Chi-Quadrat Tests für 2 x 2 Kontingenztafeln

Als Beispiel nehmen wir die Erkältungsdaten, die Pauling berühmt gemacht hat:

279 Skifahrer in einer Doppelblindstudie (Ritzel [1961])

	Placebo	Ascorbic Acid	
Erkältet	31	17	48
Nein	109	122	231
	140	139	279

$$X^2 = 4.81$$

also, signifikant ($p < 0.05$), aber nicht sehr ($p > 0.01$).

Wir betrachten den allgemeinen Fall:

	G1	G2	
Ja	x_1	x_2	$x_1 + x_2$
Nein	$n_1 - x_1$	$n_2 - x_2$	$n_1 + n_2 - x_1 - x_2$
	n_1	n_2	$n_1 + n_2$

(a) Binomial Test

Wir vergleichen die Erfolgsquoten, p_1, p_2 für die zwei Gruppen, G1 und G2. Für groß n_1, n_2 sind die Schätzer $\hat{p}_j = \frac{x_j}{n_j}$ approximativ normalverteilt.

Sei $\{X_{ij}, i = 1, \dots, n_j \quad j = 1, 2\}$ u.i.v. Bernoulli ZV, d.h.

$$P(X_{ij} = 1) = p_j \quad P(X_{ij} = 0) = 1 - p_j$$

$$E[X_{ij}] = p_j \quad V[X_{ij}] = p_j(1 - p_j)$$

$$\Rightarrow \sum_{i=1}^{n_j} X_{ij} \sim B(n_j, p_j)$$

und für n_j groß gilt

$$\sum_{i=1}^{n_j} X_{ij} \sim N(n_j p_j, n_j p_j(1 - p_j))$$

bzw.

$$\hat{p}_j = \bar{X}_j = \sum_{i=1}^{n_j} X_{ij}/n_j \sim N\left(p_j, \frac{p_j(1 - p_j)}{n_j}\right)$$

$$\Rightarrow \hat{p}_1 - \hat{p}_2 \sim N\left(p_1 - p_2, \frac{p_1(1 - p_1)}{n_1} + \frac{p_2(1 - p_2)}{n_2}\right)$$

Unser H_0 wird $p_1 = p_2 = p$ sein. Für die Varianz setzen wir den Plugin-Schätzer für p ein

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{48}{279} = 17.2\% \text{ beim Pauling}$$

und betrachten die Teststatistik

$$T = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.2214 - 0.1223}{0.0452} = 2.193$$

$$T \sim N(0, 1) \Rightarrow T^2 \sim \chi_1^2$$

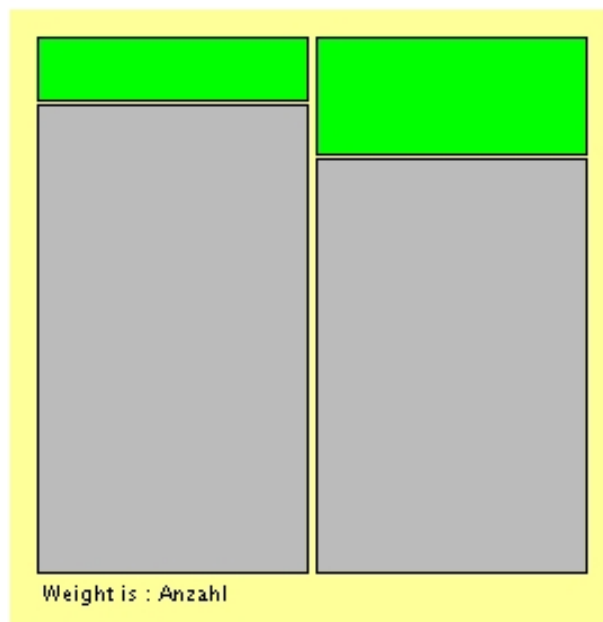
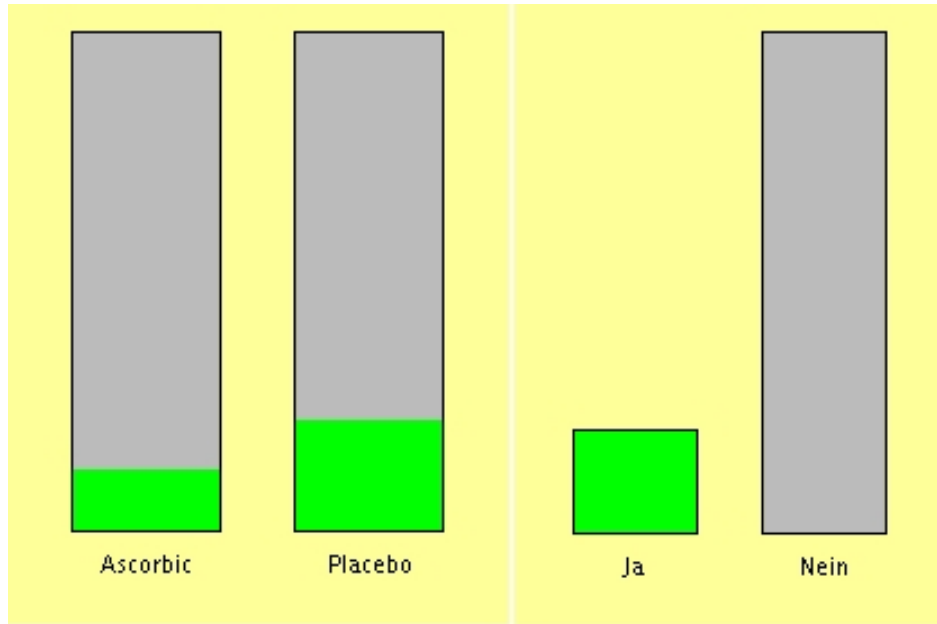
und wir können zeigen, dass

$$T^2 = \frac{(n_2x_1 - n_1x_2)^2(n_1 + n_2)}{n_1n_2(x_1 + x_2)(n_1 + n_2 - x_1 - x_2)}$$

(b) Chi-Quadrat Test

Für die angegebene 2x2 Kontingenztafel gilt

$$\begin{aligned} X^2 &= \left(x_1 - \frac{(x_1 + x_2)n_1}{n_1 + n_2}\right)^2 \left(\frac{1}{e_{11}} + \frac{1}{e_{21}} + \frac{1}{e_{12}} + \frac{1}{e_{22}}\right) \\ &= \frac{(n_2x_1 - n_1x_2)^2(n_1 + n_2)}{n_1n_2(x_1 + x_2)(n_1 + n_2 - x_1 - x_2)} \\ &= T^2 \end{aligned}$$



Erkältung gegen Ascorbic acid/Placebo
Oben: Spineplot und Säulendiagramm mit
Erkältungen selektiert
Unten: Mosaicplot mit Erkältungen selektiert

(2) Chi-Quadrat Tests für u. Poisson Verteilungen

Sei die Anzahl Beobachtungen in der Zelle i

$$X_i \quad \text{u.} \quad \sim P(\lambda_i)$$

Für λ_i groß gilt approximativ

$$X_i \sim N(\lambda_i, \lambda_i)$$

$$\Rightarrow \frac{X_i - \lambda_i}{\sqrt{\lambda_i}} \sim N(0, 1)$$

$$\Rightarrow \frac{(X_i - \lambda_i)^2}{\lambda_i} \sim \chi_1^2$$

$$\Rightarrow X^2 = \sum_{i=1}^m \frac{(X_i - \lambda_i)^2}{\lambda_i} \sim \chi_m^2$$

Aber im allgemeinen sind die Zellen nicht unabhängig und wir müssen die λ_i 's schätzen. Für jeden geschätzten Parameter nimmt man einen Freiheitsgrad weg. Der Beweis für die asymptotische Richtigkeit dieser Vorgehensweise ist aufwändig.

9.5.5 Beispiel — Tests von π

Gibt es Muster in der Dezimaldarstellung von π ?

1) Die Anzahl von Nullen testen

$$H_0 : P(0) = 0.1$$

$$H_A : P(0) \neq 0.1$$

Sei X_{0n} die Anzahl Nullen in den ersten n Stellen

$$X_{0n} \sim B(n, 0.1)$$

$$n \text{ groß} \Rightarrow X_{0n} \sim N(0.1n, 0.09n)$$

Ein Test des Niveaus $\alpha = 0.05$ hat in diesem Falle den Annahmebereich

$$0.1n \pm 1.96 * 0.3\sqrt{n}$$

z.B. $n = 1000000 \Rightarrow [99412, 100588]$ und in der Tat gibt es 99959 Nullen in den ersten Million Stellen.

Aber was ist mit den Zahlen 1, ..., 9?

2) Multinomial Test

Sei n_i die Anzahl von 'i', $i = 0, \dots, 9$

$$P(\{n_i\}) = \frac{n!}{\prod_{i=0}^9 n_i!} \prod_{i=0}^9 p_i^{n_i}$$

$$p_i = P(i)$$

$$H_0 : p_i = 0.1 \quad \forall i$$

$$H_A : p_i \neq 0.1 \text{ für einige } i$$

Annahmebereich B:

Alle Zustände (n_0, \dots, n_9) die unter H_0 eine Gesamtwahrscheinlichkeit $\geq 1 - \alpha$ haben und wo

$$P_{H_0}(\{n_i\} \in B) \geq P_{H_0}(\{n_i\} \notin B)$$

d.h. die Wahrscheinlichkeit von jedem Ereignis im Annahmebereich ist größer gleich die Wahrscheinlichkeit von jedem Ereignis im Ablehnungsbereich

3) Ein anderes Modell

Sei $n_i \sim P(np_i) \quad \forall i$ unabhängig von den anderen n_j 's und zuerst ohne der Einschränkung $\sum n_i = n$.

$$P(\{n_i\}) = \prod_{i=0}^9 e^{-np_i} \frac{(np_i)^{n_i}}{n_i!} = e^{-n} n^n \prod \frac{p_i^{n_i}}{n_i!}$$

Unter der Bedingung, daß $\sum n_i = n$, muß dann

$$P_p(\{n_i\} | \sum n_i = n) \propto \prod \frac{p_i^{n_i}}{n_i!}$$

Aus einem Vergleich mit dem Resultat für die Multinomiale Verteilung muß

$$P_p(\{n_i\} | \sum n_i = n) = n! \prod \frac{p_i^{n_i}}{n_i!}$$

Daraus schliessen wir, dass beide Ansätze zum selben Test führen. Aber der Zustandsraum ist zu groß. n identische Kugeln auf m verschiedene Zellen bedeutet

$$\binom{n + m - 1}{m - 1}$$

$$m = 10, n = 100 \Rightarrow \binom{109}{9} \gg 12^9$$

4) χ^2 Test

Für np_i groß gilt

$$n_i \sim N(np_i, np_i)$$
$$\Rightarrow \left(\frac{n_i - np_i}{\sqrt{np_i}} \right)^2 \sim \chi_1^2$$

Für $\{n_i\}$ unabhängig

$$X^2 = \sum_{i=0}^9 \frac{(n_i - np_i)^2}{np_i} \sim \chi_{10}^2$$

Unter der Einschränkung, dass $\sum n_i = n$, hat X^2 eine χ_9^2 Verteilung, weil wir einen Freiheitsgrad verloren haben.

$$\chi_{9,0.95}^2 = 16.92, \chi_{9,0.99}^2 = 21.67$$

Bei den Tests von Kanada war der höchste X^2 Wert 9.32.

Natürlich gibt es viele andere Tests auf die Zufälligkeit von Zahlenreihen — u.a. den Poker Test und den Lücken Test.

Frage: gibt es Muster in den ersten Dezimalstellen von e ?

9.5.6 Eine wichtige Eigenschaft von X^2

Für eine Kontingenztafel gilt

$$X^2 = \sum_i \sum_j \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$

und unter H_0

$$= \sum_i \sum_j \frac{(n\hat{p}_{ij} - n\hat{p}_i\hat{p}_j)^2}{n\hat{p}_i\hat{p}_j}$$

$$= n \sum_i \sum_j \frac{(\hat{p}_{ij} - \hat{p}_i\hat{p}_j)^2}{\hat{p}_i\hat{p}_j}$$

d.h. der Wert von X^2 steigt direkt mit n für gegebene beobachtete Proportionen.

Wenn ein Resultat nicht signifikant ist, kann man leicht ausrechnen, wie groß die Stichprobe hätte sein müssen, um Signifikanz zu erreichen, gegeben die selben Proportionen.

9.5.7 Zusammenfassendes zum χ^2 Test

χ^2 Tests werden sehr oft verwendet. Man muss überprüfen, welche Annahmen gemacht worden sind, welche Nullhypothese getestet worden ist, und wieviele Tests gemacht worden sind. (Es gibt jede Menge Möglichkeiten, Kategorien einer Kontingenztafel zu kombinieren.)

Der p-Wert allein genügt nicht. Man sollte untersuchen, wo die Abweichungen von der Nullhypothese erscheinen. Dafür sind die χ^2 Beiträge der einzelnen Zellen aufschlussreich.

χ^2 Tests berücksichtigen nicht ordinale Struktur. Deshalb sind andere Testverfahren besser, wenn es um Anpassungstests geht. Trotzdem können χ^2 Tests als erster Versuch nützlich sein.

Nur große Werte von χ^2 werden als signifikant betrachtet, nicht kleine Werte, sogar wenn sie sehr unwahrscheinlich sind. ("Die beobachtete Werte sind zu gut.")