

Ehrliche Antworten auf indiskrete Fragen - Randomized Response Technik

1. Einleitung

Auf „**indiskrete**“ Fragen wie:

- Sind Sie schon mal betrunken Auto gefahren?
- Haben Sie im letzten Jahr illegale Drogen (z.B. Haschisch, Heroin, Ecstasy, Kokain, ...) konsumiert?
- Haben Sie schon mal etwas an Ihrem Arbeitsplatz unerlaubt entwendet?

wird ein befragter Täter nur selten ehrlich antworten. Insbesondere dann, wenn seine Anonymität nicht hinreichend gewährleistet ist.

Das kann der Fall sein:

- bei Befragungen an der Haustür oder am Telefon;
- wenn der Interviewer den Befragten kennt;
- wenn zusätzliche Fragen (Beruf, Personenstand, Alter, ...) gestellt werden, mit denen ein Befragter eventuell ermittelt werden kann.

Bei indiskreten Fragen kann man davon ausgehen, dass eine **direkte** Befragung den tatsächlichen Anteil an „Tätern“ meistens deutlich unterschätzen wird.

Beispiel 1: direkte Befragung

Haben Sie schon einmal in einem Geschäft etwas gestohlen? ja nein

Beispiel 2: RRT- Umfrage

Haben Sie schon einmal etwas in einem Geschäft gestohlen?

Um Ihre Anonymität bei dieser Befragung auf **jeden Fall** zu gewährleisten, haben wir uns etwas überlegt.

Sie werfen zuerst verdeckt eine Münze.

Zeigt die Münze Wappen, so sagen Sie auf jeden Fall ja, ganz egal ob Sie schon einmal einen Ladendiebstahl begangen haben oder nicht.

Zeigt die Münze Zahl, so beantworten Sie die Frage wahrheitsgemäß.

Auf diese Weise kann Sie niemand bei einer Ja-Antwort belangen! Sie können ja immer sagen, dass Sie die Münze zu einem „ja“ gezwungen habe.

Antwort: ja nein

Unter RRT- Umfragen versteht man Umfragen, die mit einem **Zufallsexperiment** gekoppelt sind. Die Ausgänge des Zufallsexperiments, ihre Wahrscheinlichkeiten und die nachfolgenden Handlungen sind bekannt.

RRT kommt von **R**andomized - **R**esponse-**T**echnik.

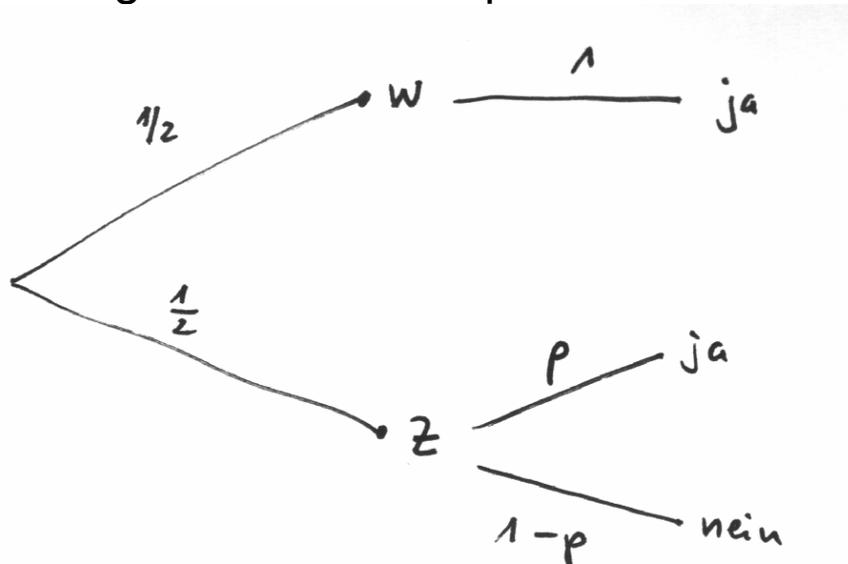
1. Auswertung:

Umfrage 1: ja Antworten; nein Antworten

Umfrage 2: ja Antworten; nein Antworten

1. Es gibt also wohl mindestens _____ unter uns, die einen Ladendiebstahl begangen haben.
2. Was fangen wir mit den Zahlen aus Umfrage zwei an? Wir müssen diese Zahlen korrigieren!

Lösung: Wir zeichnen einen Baum. Die gesuchte (unbekannte) Wahrscheinlichkeit einen Ladendieb in der Bevölkerung anzutreffen sei p .



Die Pfadregeln ergeben für die Wahrscheinlichkeit $P(ja)$

der ja - Antworten $P(ja) = \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot p$.

Setzen wir nun $P(ja) = h(ja)$, wobei $h(ja)$ die beobachtete relative Häufigkeit in unserer Stichprobe ist, so machen wir einen (kleinen?) Fehler, weil uns **nicht** die Häufigkeit von Ladendieben in dieser kleinen Gruppe, sondern in der ganzen Bevölkerung interessiert.

Wir erhalten damit nur einen **Schätzwert** \hat{p} für p :

$$h(ja) = \frac{1}{2} + \frac{1}{2} \cdot \hat{p} \Rightarrow \hat{p} = (h(ja) - \frac{1}{2}) \cdot 2$$

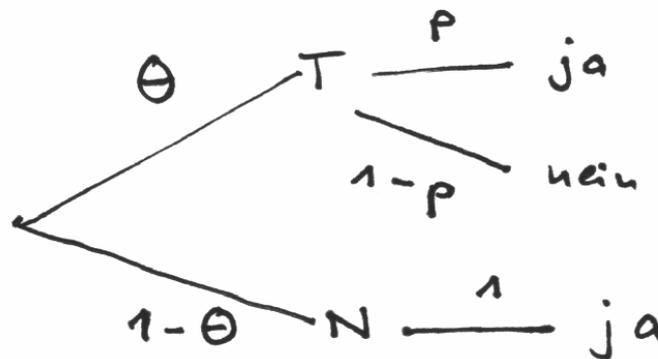
In unserm Beispiel also: $\hat{p} = 2 (h(ja) - 0,5) = \dots\dots\dots$

Fragen:

1. Was spricht dafür, dass Befragte bei Umfrage 2 eher ehrlicher antworten, als bei Umfrage 1?
2. Wann ist die Annahme $P(ja)=h(ja)$ gerechtfertigt?
3. Wie könnte man die Sicherheit für die Befragten steigern, und ihnen so die Antwort erleichtern?
4. Sehen Sie Nachteile bei der Umfragemethode 2?

2. Verallgemeinerung des einführenden Beispiels

Gegeben sei ein Zufallsgerät mit zwei Ausgängen N , T (Bernoulli-Experiment), wobei $P(T) = \theta$, $P(N) = 1 - \theta$ bekannt sei. Wird T gezogen, so wird eine **sensitive Frage** vom Befragten **wahrheitsgemäß beantwortet**, bei N wird auf **jeden Fall mit ja** - geantwortet. Es sei p die gesuchte Wahrscheinlichkeit eines Täters in der Grundgesamtheit. Es ergibt sich folgender Baum:



Die Pfadregeln ergeben:

$$(1) \quad P(ja) = \theta \cdot p + (1 - \theta) \cdot 1 \quad .$$

Nun setzt man $P(ja)=h(ja)$, wobei $h(ja)$ die beobachtete **relative Häufigkeit** von ja - Antworten in einer möglichst großen Stichprobe darstellt.

Damit errechnet man für p den Schätzwert:

$$(2) \quad \hat{p} = \frac{h(ja) - 1 + \theta}{\theta}$$

Bemerkung: In Übung A1, A2, A3, A4 → Übungsblatt finden Sie weitere - teilweise etwas anders geartete - Beispiele zu RRT - Verfahren.

Lösen Sie diese Aufgaben in Gruppenarbeit und führen Sie eine dieser Aufgaben an der Tafel vor.

3. Streuung des Schätzwertes \hat{p} experimentell

Man wird erwarten, dass bei RRT - Umfragen der Anteil der wahrheitsgemäß Antwortenden steigt. Hat der eingebaute Zufallsmechanismus also nur Vorteile?

Wenn $h(ja) = P(ja)$ wäre, dann wäre $p = \hat{p}$.

Dies können wir natürlich nicht erwarten, und im konkreten Fall auch nicht wissen.

Wenn die Stichprobe groß ist, werden allerdings größere Abweichungen zwischen $h(ja)$ und $P(ja)$ immer unwahrscheinlicher. Das ist ganz genau so wie beim Würfeln: auf lange Sicht wird die relative Häufigkeit z.B. für die Augenzahl 4 sich bei vielen Würfeln um $1/6$ stabilisieren.

Bei einer gegebenen Umfrage wird $h(ja)$ zufallsbedingt mehr oder weniger stark um $P(ja)$ streuen. Selbst bei derselben Gruppe von Befragten werden wir bei Wiederholung der Befragung einen anderen Wert von $h(ja)$ erhalten, einfach weil nun andere Personen sensible Fragen bzw. die vorgegebene Antwort „ja“ erhalten.

Davon können wir uns in einer Reihe von Zufallsexperimenten selbst überzeugen.

Übung: „Hast du schon mal einen Ladendiebstahl begangen?“

Unsere Arbeitsgruppe wird eingeteilt in 30% Ladendiebe und 70% ehrliche Menschen.

Jeder wirft eine Münze. Bei Wappen wird die Frage wahrheitsgemäß beantwortet, bei Zahl wird immer ja gesagt.

Dieses Experiment führen wir mehrfach durch und zählen jeweils die Anzahl der ja - Antworten in der Gruppe:

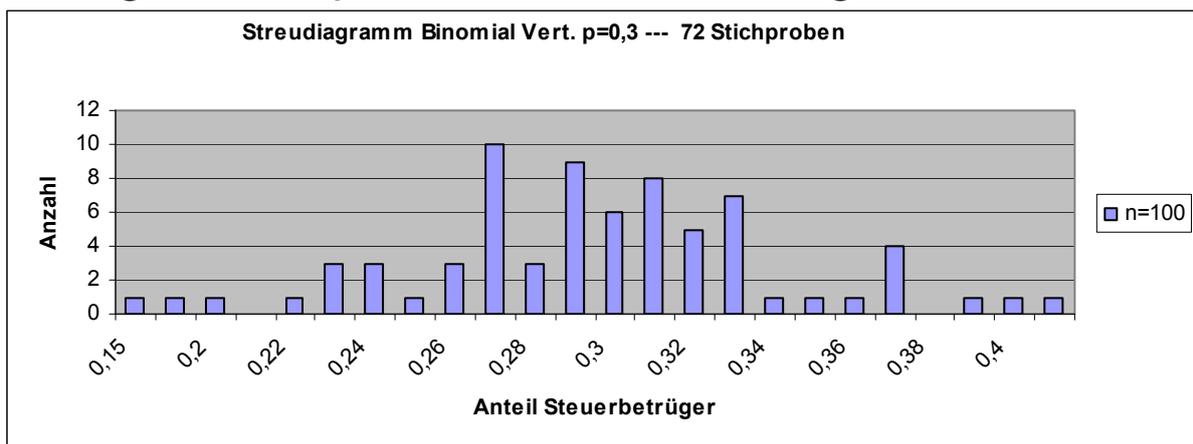
| Ex | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|-----------|---|---|---|---|---|---|---|---|---|----|
| ja | | | | | | | | | | |
| \hat{p} | | | | | | | | | | |

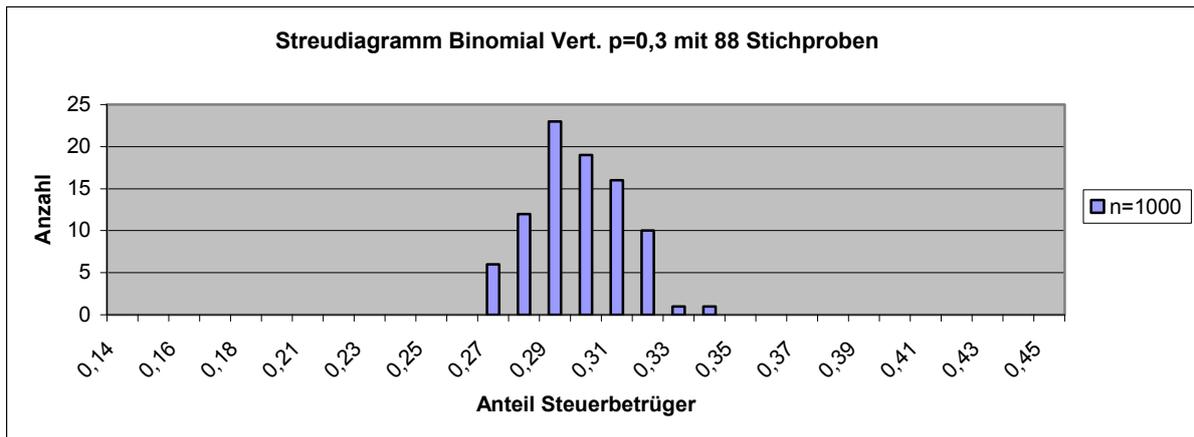
Wir sehen deutlich die Streuung der Schätzwerte \hat{p} .
Wie reagiert die Streuung auf eine Vergrößerung des Stichprobenumfangs?

Große Stichproben erzeugt man sinnvoll mit Computerhilfe → Excel Dateien RRT1.

Wir erzeugen zufällig Stichproben vom Umfang 100 bzw. 1000 mit festem Anteil $p=0,7$ von ehrlichen Steuerzahlern.

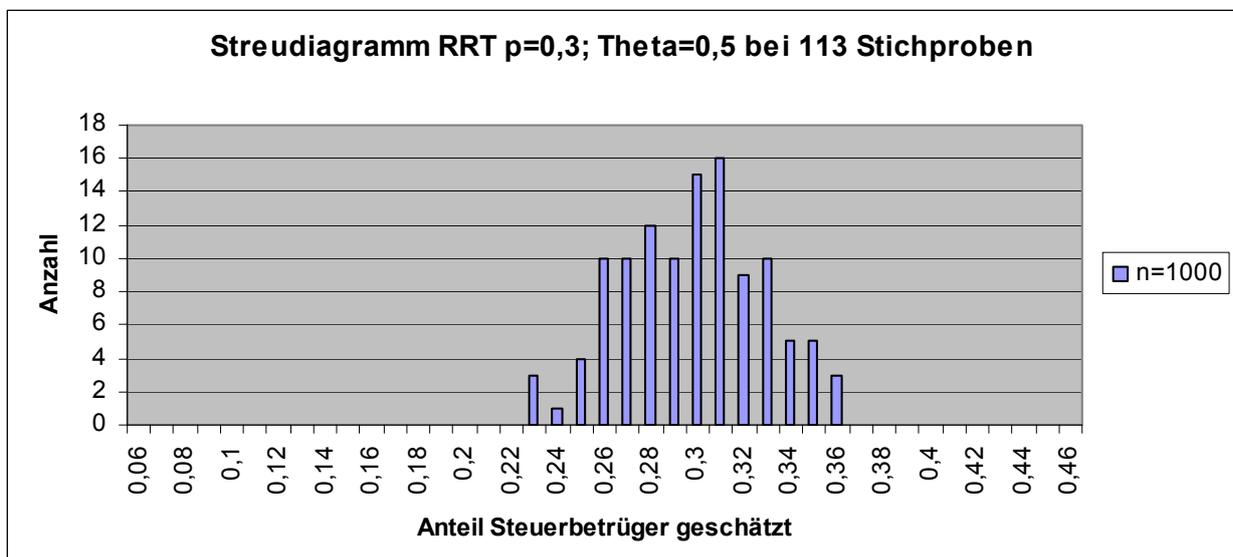
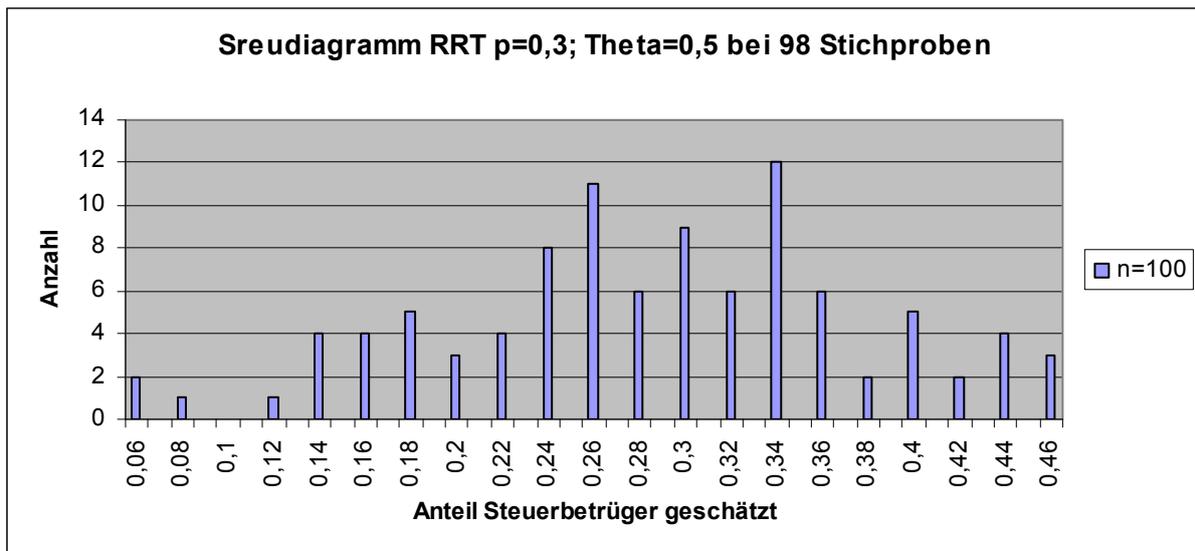
Wegen des Zufalls schwanken die tatsächlichen Anzahlen der Steuerbetrüger in den jeweiligen zufällig erzeugten Stichproben mehr oder weniger stark.





Wir erkennen, dass bei größerem Stichprobenumfang die Streuung des Anteils an Steuerbetrügern geringer wird.

Nun bauen wir jede erzeugten Stichprobe zu einem RRT - Experiment mit Wahrscheinlichkeit θ aus \rightarrow RRT3.



Wir erkennen, dass die Schätzwerte für \hat{p} nun leider viel stärker streuen wie vorher ohne RRT- Experiment.

Stichproben mit $n=100$ sind zu klein: \hat{p} schwankt zwischen 0,06 und 0,46! Aus einer einzelnen solchen Befragung können wir also recht wenig schließen.

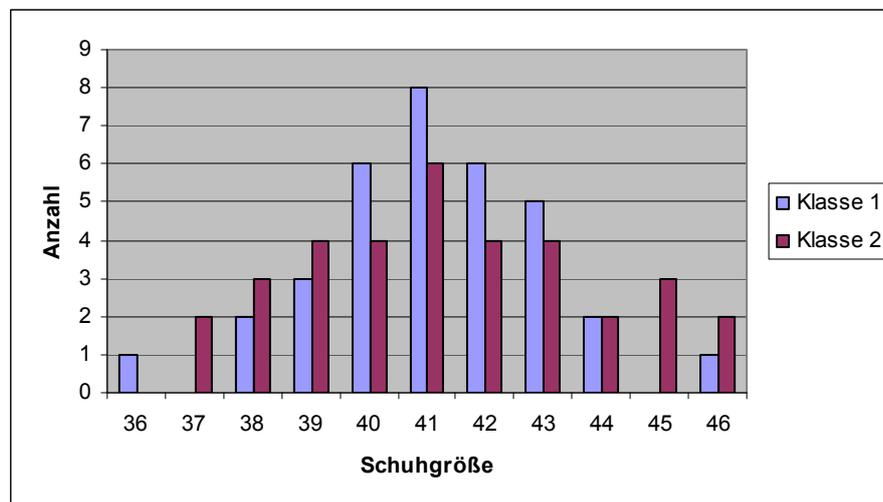
Allerdings nimmt zum Glück die Streuung für größere n wieder ab. Bei $n=1000$ schwankt \hat{p} nur noch zwischen 0,23 und 0,36. Solche Werte geben den tatsächlichen Wert doch recht genau wieder.

Befragungen von 1000 Menschen sind aber aufwändig! Es ergeben sich nun folgende Fragen:

- Was bedeutet Streuung genau?
- Wie können wir diesen Begriff mathematisch fassen?
- Kann man unsere empirischen Aussagen über die Streuung mathematisch nachweisen?

4. Einschub Theorie

Beispiel: In zwei 11. Klassen mit jeweils 34 Schülern werden die Schuhgrößen der Schüler ermittelt. Die mittlere Schuhgröße in Klasse 1 ist 41,1 und unterscheidet sich kaum von der in Klasse 2: 41,3.



Die Schuhgrößen in Klasse 1 sind deutlicher um den Mittelwert konzentriert, die Größen in Klasse 2 streuen stärker. Wie soll man dies quantifizieren, also in einer leicht berechenbaren Größe fassen?

Unter der **empirischen Varianz** einer Stichprobe x_1, x_2, \dots, x_n mit Mittelwert \hat{x} versteht man die mittlere quadratische Abweichung vom Mittelwert:

$$(3) \quad V_{emp} = \frac{(x_1 - \hat{x})^2 + (x_2 - \hat{x})^2 + \dots + (x_n - \hat{x})^2}{n}$$

Die Wurzel aus der Varianz nennt man **empirische Standardabweichung** (4) $\sigma_{emp} = \sqrt{V_{emp}}$

Wenn der Stichprobenwert x_j mit der relativen Häufigkeit h_j auftaucht, so kann man die Varianz rechentechnisch einfacher wie folgt berechnen:

$$(5) \quad V_{emp} = (x_1 - \hat{x})^2 \cdot h_1 + (x_2 - \hat{x})^2 \cdot h_2 + \dots + (x_n - \hat{x})^2 \cdot h_n$$

Es ergibt sich im Schuhbeispiel:

$$\text{Klasse 1: } V_{emp} \approx 3,8, \quad \sigma_{emp} \approx 2$$

$$\text{Klasse 2: } V_{emp} \approx 6,2, \quad \sigma_{emp} \approx 2,5$$

Durch Nachrechnen bestätigt man:

1. Addiert man zu jedem Stichprobenwert eine feste Konstante c , so ändert sich die Varianz nicht.
2. Multipliziert man jeden Stichprobenwert mit einer festen Konstanten k , so wird die Varianz k^2 mal so groß.

Bei glockenförmigen Verteilungen gilt eine Faustregel: Ungefähr **2/3** aller Stichprobenwerte liegen im Intervall

$$[\hat{x} - \sigma_{emp}, \hat{x} + \sigma_{emp}]$$

Die Begriffe Mittelwerte und Varianz lassen sich von konkreten Stichproben auf Zufallsvariable übertragen.

Gegeben seien endlich viele Ausgänge a_1, a_2, \dots, a_n eines Wahrscheinlichkeitsraumes.

Unter einer **Zufallsvariablen** X versteht man eine Funktion, die jedem Ausgang a_i eine reelle Zahl $X(a_i)$ zuordnet.

Beispiel: 1. Man wirft dreimal eine Münze.

Dann besteht der Wahrscheinlichkeitsraum aus acht 3-Tupeln: (w|w|w) (w|w|z) (w|z|w) (z|w|w) (z|z|w) (z|w|z) (w|z|z) (z|z|z).

Eine sinnvolle Zufallsvariable X ist hier Anzahl von Zahl:

$$X((z|w|w))=1 \quad X((z|z|z))=3 \quad X((w|w|w))=0$$

2. Man würfelt zweimal. Der Wahrscheinlichkeitsraum besteht aus 36 2-Tupeln (1|1), (1|2), ..., (6|6).

Eine sinnvolle Zufallsvariable Y ist die Augensumme $Y((3|4))=7, Y((1|3))=4$.

Man kann sich fragen, welchen Wert eine Zufallsvariable im Mittel annimmt. Dabei lässt man sich vom Mittelwert einer Stichprobe inspirieren:

gegeben Stichprobe: 3,3,3,4,4,2,2,1

$$\Rightarrow \hat{x} = \frac{3+3+3+4+4+2+2+1}{8} = 3 \cdot \underbrace{\frac{3}{8}}_{\text{Häufigkeit 3}} + 4 \cdot \underbrace{\frac{2}{8}}_{\text{Häufigkeit 4}} + 2 \cdot \underbrace{\frac{2}{8}}_{\text{Häufigkeit 2}} + 1 \cdot \underbrace{\frac{1}{8}}_{\text{Häufigkeit 1}}$$

Die Zufallsvariable X nehme die Werte x_1, \dots, x_m an. Dann wird der **Erwartungswert** von X festgelegt durch:
 $E(X) := x_1 \cdot P(X = x_1) + x_2 \cdot P(X = x_2) + \dots + x_m \cdot P(X = x_m)$

Beispiel 1: Dreimaliger Münzwurf, mittlere Anzahl an W:

$$E(X) = 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = \frac{12}{8} = 1,5$$

Bemerkung: Der dreimalige Münzwurf ist ein Spezialfall einer Bernoullikette der Länge 3.

Eine **Bernoullikette der Länge n** besteht in der n - maligen Durchführung eines Zufallsexperiments mit zwei Ausgängen T (Treffer) und N (Niete) mit den Wahrscheinlichkeiten $P(T)=p$ und $P(N)=1-p$.

Man kann zeigen:

Die Zufallsvariable X = Anzahl der Treffer in einer Bernoullikette ist **binomial-verteilt** mit dem Erwartungswert $E(X) = np$.

Beweis: s. [1], für den Spezialfall des dreimaligen Würfels haben wir das Ergebnis bereits ermittelt .

Beispiel 2: Mittlere Augensumme bei 2 Würfeln:

$$E(Y) = 2 \cdot \frac{1}{36} + 3 \cdot \frac{2}{36} + 4 \cdot \frac{3}{36} + 5 \cdot \frac{4}{36} + 6 \cdot \frac{5}{36} + 7 \cdot \frac{6}{36} + 8 \cdot \frac{5}{36} + \dots + 12 \cdot \frac{1}{36} = \frac{252}{36} = 7$$

Analog zu Stichproben definiert man:

Unter der **Varianz** einer Zufallsvariable versteht man die Zahl:

$$(6) V(X) = (x_1 - E(X))^2 \cdot P(X = x_1) + (x_2 - E(X))^2 \cdot P(X = x_2) + \dots + (x_m - E(X))^2 \cdot P(X = x_m)$$

Die **Standardabweichung** ist definiert durch

$$(7) \sigma(X) = \sqrt{V(X)}$$

Beispiel 1: Die Varianz der Zufallsvariablen Anzahl von Zahl beim dreimaligen Münzwurf ist:

$$V(X) = (0 - 1,5)^2 \cdot \frac{1}{8} + (1 - 1,5)^2 \cdot \frac{3}{8} + (2 - 1,5)^2 \cdot \frac{3}{8} + (3 - 1,5)^2 \cdot \frac{1}{8} = 0,75 .$$

Auch hier kann man wieder auf beliebige Bernoulliketten verallgemeinern. Man erhält:

Die Varianz der Zufallsvariable X = Anzahl der Treffer in einer Bernoulli Kette der Länge n beträgt: $V(X) = np(1-p)$

Beweis: s.[1], unser Spezialfall $n=3$ stützt dieses Ergebnis.

Auch für die Varianz von Zufallsvariablen kann man leicht zeigen:

$$V(X+c)=V(X) \text{ und } V(kX)=k^2 V(X) \text{ f\u00fcr konstante Werte } c,k.$$

\u00dcbung: Berechnen Sie die Varianz der Augensumme beim zweimaligen W\u00fcrfeln.

5. Berechnung der Varianz von RRT- Umfragen

Nehmen wir an, wir stellen eine sensitive Frage in einer RRT - Befragung vom Umfang n . Wir interessieren uns f\u00fcr die Streuung der Zufallsvariablen

$$\hat{p} = \frac{h(ja) - 1 + \theta}{\theta} \approx \frac{\frac{X}{n} - 1 + \theta}{\theta}.$$

Dabei ist X die Zahl der ja- Antworten bei n Befragten. F\u00fcr deren Streuung gilt nach dem Vorigen:

$$(8) \quad V\left(\frac{\frac{X}{n} - 1 + \theta}{\theta}\right) = \frac{V\left(\frac{X}{n}\right)}{\theta^2} = \frac{V(X)}{n^2 \theta^2}.$$

Die Zufallsvariable X ist binomialverteilt mit den Parametern n und $P(ja)$, also gilt:

$$V(X) = P(ja) \cdot (1 - P(ja)) \cdot n.$$

F\u00fcr $P(ja)$ gilt nach (1) $P(ja) = \theta \cdot p + (1 - \theta) \cdot 1$.

Das setzen wir in (8) ein:

$$V(\hat{p}) = \frac{(\theta p + (1 - \theta))(1 - \theta p - (1 - \theta))}{n\theta^2} = \frac{(\theta p + 1 - \theta)(\theta - \theta p)}{n\theta^2} =$$

$$\frac{(1 + \theta p - \theta)(1 - p)}{n\theta}$$

$$\Rightarrow V(\hat{p}) = \underbrace{\frac{p(1-p)}{n}}_{\text{Varianz direkte Befragung}} + \underbrace{\frac{(1-\theta)(1-p)}{n\theta}}_{\text{zus\u00e4tzlicher Anteil RRT}}$$

Bei direkter Befragung würden wir eine Bernoullikette mit den Parametern n und p erzeugen.

Die Varianz der uns dabei interessierenden Zufallsvariablen

$\frac{X}{n}$ wäre $V\left(\frac{X}{n}\right) = \frac{np(1-p)}{n^2}$, also gerade der erste

Summand.

Die Varianz der RRT - Umfrage vergrößert sich also immer um den 2. Term.

Beispiel: Angenommen 10% aller Eltern schlagen ihre Kinder.

Bei direkter Befragung in einer Stichprobe vom Umfang $n=30$ würde sich unter der (unrealistischen) Annahme, dass alle darauf ehrlich antworten eine Streuung von $0,1 \cdot 0,9 / 30 = 0,0037$ für den interessierenden Anteil ergeben.

Bei einer RRT - Umfrage mit $\theta = \frac{1}{2}$ käme dazu aus dem

zweiten Term noch $\frac{(1-0,5) \cdot (1-0,1)}{30 \cdot 0,5} = 0,0371$ dazu.

Die Gesamtvarianz wäre somit $0,041$.

Damit ist die Standardabweichung $0,20$.

Selbst wenn also der ermittelte Schätzwert $\hat{p} = 0,11$

wäre, könnten wir lediglich sagen, dass etwa $\frac{2}{3}$ der

möglichen Werte im Intervall $[-0,09 ; 0,31]$ lägen!

Das nützt uns wenig!

Steigern wir aber den Umfang der Stichprobe auf

$n=1000$, so ist bei gleichen Gegebenheiten die

Gesamtvarianz $0,00009+0,0009=0,00099$ und die

Standardabweichung $0,031$.

Nun wüssten wir, dass bei einem Schätzwert $\hat{p} = 0,11$

etwa $\frac{2}{3}$ der Werte im Intervall $[0,08|0,14]$ lägen, also mit

ziemlicher Sicherheit zwischen 8% und 14% der Eltern ihre Kinder schlagen würden.
Dies wäre ein brauchbares Ergebnis.

Diese Abhängigkeit von n wird sehr gut durch die experimentellen Werte bestätigt.

Übung: A5 Aufgabenblatt

Fazit: Die RRT- Umfrage verbessert zwar die Anonymität, führt aber zu einer stärkeren Streuung des Schätzers \hat{p} , die man nur durch eine Vergrößerung der Stichprobe $\sigma(\hat{p}) \propto \frac{1}{\sqrt{n}}$ in den Griff bekommen kann.
Größere Befragungen kosten aber deutlich mehr Geld.

6. Das 3 - Gruppenmodell

Wir sind bisher davon ausgegangen, dass wegen der garantierten Anonymität alle Befragten ehrlich antworten und die Anweisungen des Testverfahrens befolgen.

Das wird in der Realität wohl kaum der Fall sein.

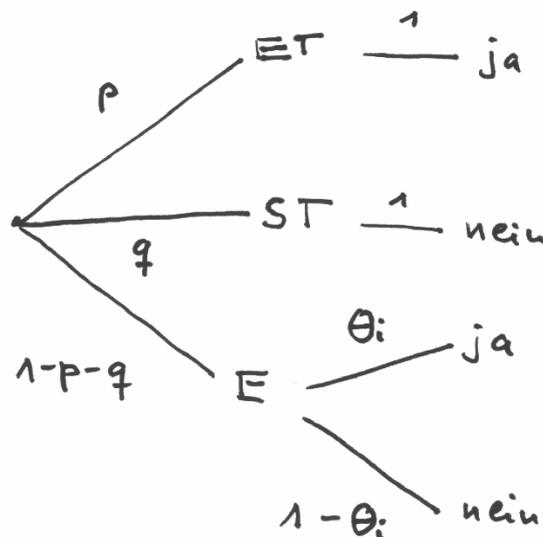
Es ist denkbar, dass manche „Täter“ ein Interesse daran haben, dass die Befragung falsche Ergebnisse liefert.

Wir wollen deswegen unser Modell um eine neue Gruppe, die **Testsaboteure** erweitern.

Es gibt nun **drei Gruppen** von Befragten.

- ET (ehrliche Täter) - Personen, auf die nachgefragte Frage zutrifft und die sich im Test ehrlich verhalten. Ihr Anteil sei p .
- ST (sabotierende Täter) - Personen, auf die die Frage zutrifft, die aber den Test boykottieren. Bei echter Frage, sagen sie wahrheitswidrig „nein“, bei vorgegebener Antwort sagen sie ebenfalls „nein“. Ihr Anteil sei q .
- E (ehrliche Personen) - Personen, auf die die sensitive Frage nicht zutrifft. Sie sollen sich im Test in jeder Hinsicht ehrlich verhalten (einmal ehrlich - immer ehrlich!). Ihr Anteil ist dann $1-p-q$.

Man erhält folgenden Baum für die RRT - Umfrage:



Die Pfadregeln ergeben: (9) $P(ja) = p \cdot 1 + (1 - p - q) \cdot \theta_i$.

Wieder werden wir $P(ja) \approx h(ja)$ als Ergebnis unserer Umfrage setzen.

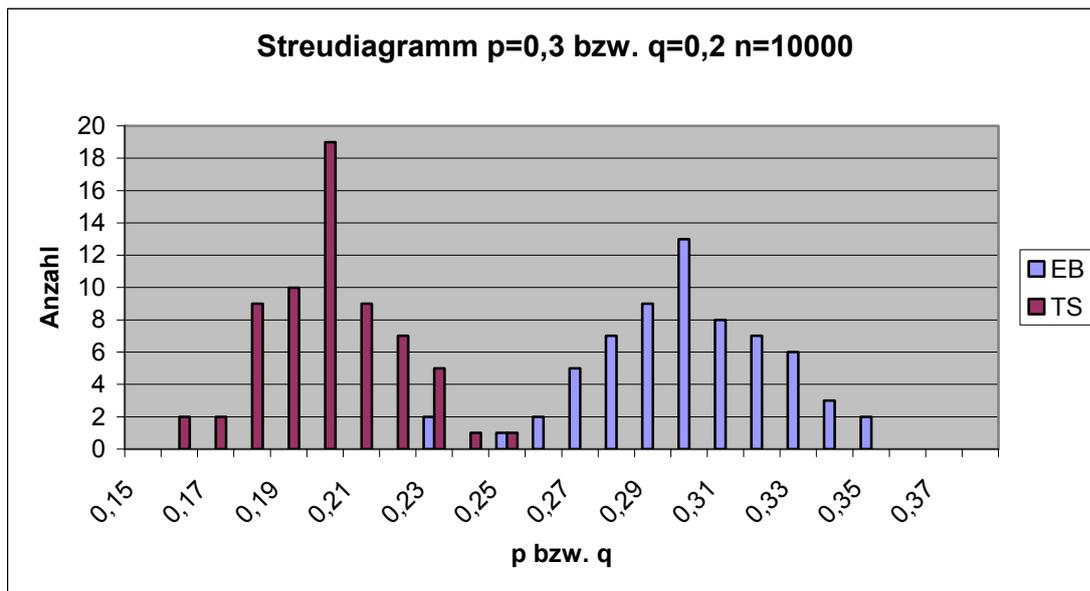
Damit bekommen wir aus jeder Umfrage eine lineare Gleichung für zwei Unbekannten \hat{p} und \hat{q} .

Wir müssen also, um beide Variable schätzen zu können, zwei Umfragen mit unterschiedlichem θ durchführen.

Dabei erhalten wir ein 2x2 lineares Gleichungssystem für \hat{p} und \hat{q} , das wir lösen können.

Übung: A6 → Übungsblatt

Wenn wir das Verfahren wieder mit Excel simulieren (→ RRT3Gruppen), so erkennen wir, dass nun die beiden Schätzwerte noch stärker streuen, und wir sehr große n (im Beispiel $n=10\,000$) wählen müssen, damit wir einigermaßen stabile Werte erhalten. Das macht den Einsatz in der Praxis leider sehr teuer.



Literatur:

- [1] Lambacher - Schweizer: Stochastik - Leistungskurs, Klett-Verlag
- [2] Engel, A.: Wahrscheinlichkeitsrechnung und Statistik, Band 1, Klett Studienbücher
- [3] Krüger, K.: Wahrheit oder Pflicht, Mathematik lehren, H125, S.50-54

Laufblatt: Bezeichnungen, Begriffe:

p - unbekannte Wahrscheinlichkeit eines Täters in der Grundgesamtheit

\hat{p} - Schätzer für p

$P(ja)$ - unbekannte Wahrscheinlichkeit für eine ja -Antwort in der Grundgesamtheit

$h(ja)$ - relative Häufigkeit von ja- Antworten in der Stichprobe

θ - Wahrscheinlichkeit für echte Fragen des vorgeschalteten Zufallsexperiments

\hat{x} - arithmetischer Mittelwert einer Reihe von Stichprobenwerte

$V_{emp} = \frac{(x_1 - \hat{x})^2 + (x_2 - \hat{x})^2 + \dots + (x_n - \hat{x})^2}{n}$ - empirische Varianz einer

Stichprobe

$V_{emp} = (x_1 - \hat{x})^2 \cdot h_1 + (x_2 - \hat{x})^2 \cdot h_2 + \dots + (x_n - \hat{x})^2 \cdot h_n$ - andere Form der empirischen Varianz

$\sigma_{emp} = \sqrt{V_{emp}}$ - die empirische Standardabweichung einer Stichprobe

X, Y - Zufallsvariable

$E(X) := x_1 \cdot P(X = x_1) + x_2 \cdot P(X = x_2) + \dots + x_m \cdot P(X = x_m)$ - Erwartungswert einer diskreten Zufallsvariablen

$V(X) = (x_1 - E(X))^2 \cdot P(X = x_1) + (x_2 - E(X))^2 \cdot P(X = x_2) + \dots + (x_m - E(X))^2 \cdot P(X = x_m)$ -

Varianz einer Zufallsvariable

$\sigma(X) = \sqrt{V(X)}$ - Standardabweichung einer Zufallsvariablen

ET - Täter, aber im Test ehrlich

ST - Täter, die den Test sabotieren

q - unbekannte Wahrscheinlichkeit einen Testsaboteur in der Grundgesamtheit anzutreffen

E - keine Täter, auch im Test ehrlich