Anwendungen der Binomialverteilung

Hansruedi Künsch
Seminar für Statistik, ETH Zürich
kuensch@stat.math.ethz.ch

26. Januar 2015

Dieses Dokument basiert auf einem Atelier am Kantonalen Fachschaftstag Mathematik an der Kantonsschule Alpenquai Luzern vom 21. Januar 2015. Es handelt sich um Beispiele, bei denen sich die Frage nach einem Test oder einem Vertrauensintervall für Wahrscheinlichkeiten in natürlicher Weise ergibt. Die Beispiele sollten auch einigermassen realistisch sein. Daher werfe ich manchmal auch die Frage auf, welche Schwierigkeiten auftreten können bei der praktischen Anwendung.

Es ist geplant, diese Liste noch etwas zu erweitern und auch Lösungen zu schreiben. Dafür brauche ich aber noch etwas Zeit.

Geschlecht von Neugeborenen und Zwillinge

Auf `http://www.bfs.admin.ch/bfs/portal/de/index/themen/01/06/blank/key/02/01.html` findet man Angaben über das Geschlecht von lebendgeborenen Kindern in der Schweiz von 2009 bis 2013:

Jahr	2009	2010	2011	2012	2013
Lebendgeburten	78'286	80'290	80'808	82'164	82'731
Knaben	40'407	41'111	41'626	42'435	42'595

Ferner wird die Häufigkeit von Zwillingsgeburten im Jahr 2009 mit 1415 unter 75'726 Entbindungen angegeben.

- Ist es plausibel, dass Knaben- und Mädchengeburten gleich wahrscheinlich sind ?
- Wie genau kann man die Wahrscheinlichkeit für eine Knabengeburt aus den Daten von einem, bzw. 4 Jahren schätzen ?
- Wie gross ist etwa die Wahrscheinlichkeit für eine Zwillingsgeburt ?

Bei Zwillingen gibt es bekanntlich eineiige und zweieiige. Die Wahrscheinlichkeit, dass ein Zwillingspaar eineiig ist, schwankt je nach Region und Periode. Wikipedia gibt Werte zwischen $\frac{1}{4}$ und $\frac{1}{2}$ an. Wenn wir annehmen, dass diese Wahrscheinlichkeit gleich $\frac{1}{4}$ ist, wie gross ist dann die Wahrscheinlichkeit, dass ein Zwillingspaar, von dem man nicht weiss, ob es ein- oder zweieiig ist, das gleiche Geschlecht hat ? In welchem Bereich schwankt etwa die Anzahl Zwillinge mit gleichem Geschlecht bei 1400 Geburten ?

Mendel'sche Versuche

Mendel hat in einem seiner Versuche zwei reine Stämme von gelben und grünen Erbsen gekreuzt. Da gelb dominant ist, sind in der ersten Generation alle Erbsen gelb und in der zweiten Generation erwartete er für die grünen Erbsen eine relative Häufigkeit von $\frac{1}{4}$. In seinem Versuch erhielt er in der zweiten Generation 8023 Erbsen, von denen 2001 grün waren.

- Wenn Mendel's Theorie richtig ist, in welchem Bereich schwankt dann mit Wahrscheinlichkeit 0.95 die beobachtete Häufigkeit grüner Erbsen bei einem Total von 8023 ?
- Was für Folgerungen ziehen Sie aus der beobachteten Häufigkeit von 2001 ?

Unterscheidung von echten und konstruierten Münzwürfen

Ein Run in einer binären Folge $x_1, x_2, \ldots, x_n$ ist ein Teilstück $x_i = x_{i+1} = \ldots = x_j$ mit den zusätzlichen Bedingungen $x_{i-1} \neq x_i$ falls $i > 1$ sowie $x_i \neq x_{i+1}$ falls $i < n$. Aus Erfahrung weiss man, dass die meisten Personen Folgen mit zuvielen Runs auswählen, wenn sie versuchen, eine typische zufällige Folge hinzuschreiben, ohne eine Münze zu werfen oder einen anderen Zufallsmechanismus zu benutzen.

- Zeigen Sie, dass die Anzahl Runs gleich der Anzahl Wechsel plus eins ist. Ein Wechsel tritt auf an der Stelle $1 < i \leq n$ falls $x_{i-1} \neq x_i$.
- Wie gross ist bei einem fairen Münzwurf die Wahrscheinlichkeit, dass an einer festen Stelle i ein Wechsel auftritt ?
- Wie gross ist bei einem fairen Münzwurf die Wahrscheinlichkeit, dass sowohl an der Stelle i als auch an der Stelle $i + 1$ ein Wechsel auftritt ?
- Wie ist bei einem fairen Münzwurf die Anzahl Wechsel verteilt ?
- Geben Sie eine Grenze für die Anzahl runs an, ab der Sie eine binäre Folge der Länge $n = 50$ nicht mehr als erzeugt von einem fairen Münzwurf taxieren würden ?

Ist eine medizinische Behandlung wirkungslos ?

Medizinische Studien beurteilen die Wirksamkeit einer neuen Behandlung meist so, dass eine Behandlungsgruppe, welche die neue Behandlung erhält, verglichen wird mit einer Kontrollgruppe, welche die Standardbehandlung oder ein Placebo erhält. Dabei ist es wichtig, dass sich die beiden Gruppen nur durch die angewandte Behandlung und sonst durch keine anderen Eigenschaften (z.B. Alter, Lebensstil, genetische Voraussetzungen) unterscheiden. Um das zu garantieren, wird die Einteilung der Studienteilnehmer am besten durch das Los gemacht (sogenannte Randomisierung).

Wenn die neue Behandlung nicht besser oder schlechter ist als die Standardbehandlung oder das Placebo, dann müssen unterschiedliche Ergebnisse der beiden Behandlungen durch Glück, bzw. Pech bei der Einteilung in die beiden Gruppen erklärbar sein. Das Ergebnis hätte sich ja für jeden Teilnehmer nicht geändert, wenn er in die andere Gruppe gekommen wäre.

Als Beispiel betrachten wir die Salk Polio-Impfstudie von 1954 in den USA. Ungefähr 400'000 Kinder haben daran teilgenommen, von denen jedes mit Wahrscheinlichkeit $\frac{1}{2}$ in die Behandlungsgruppe eingeteilt wurde. 56 der Kinder aus der Behandlungsgruppe und 142 aus der Kontrollgruppe erkrankten an Polio, insgesamt also 198 Kinder. Wenn die Impfung überhaupt keine Wirkung gehabt hätte, wären diese 198 Kinder auf jeden Fall erkrankt, unabhängig davon, ob sie geimpft oder nicht geimpft waren. Das Los hätte also nur 56 statt der erwarteten 99 Kinder in die Behandlungsgruppe eingeteilt. Ist das noch plausiblel ?

Wunschkühe dank Spermasexing

Im Tages-Anzeiger vom 6. Dezember 2000 erschien ein Artikel mit obigem Titel. Darin wurde über ein Verfahren, genannt Spermasexing, berichtet, das es ermögliche, "vor der künstlichen Befruchtung die Samen nach dem Geschlecht [optisch] zu trennen und so die Rinderzucht entscheidend zu steuern". Bei einem Probelauf habe das Verfahren in elf von zwölf Fällen funktioniert ("Nur Hopplo war ein Stier"). Das Verfahren sei vom amerikanischen Landwirtschaftsdepartement und der US-Firma XY Inc. entwickelt worden, mit der Schweizer Biotechfirma BIG-X AG als Lizenznehmerin.

Eine Recherche im Internet im Januar 2015 ergab, dass über die Schweizer Lizenznehmerin BIG-X AG im Jahr 2005 das Konkursverfahren eröffnet wurde, welches 2011 abgeschlossen wurde. Die Firma XY Inc wurde 2007 von der Firma Sexing Technologies übernommen. Auf der Webseite dieser Firma (`http://www.sexingtechnologies.com/`) findet man unter anderem den folgenden Werbespruch "Would you like 90% of your calves to be male or female ? Repeatable, reliable and affordable results", sowie das folgende Testimonial von einem Kunden "The gender selection rate has been great. In the last AI program with female Sexed Semen, out of 62 calves born 61 were heifers."

Was für Schlussfolgerungen kann man aus dem Probelauf mit 11 von 12 Erfolgen, bzw. dem Testimonial mit 61 von 62 Erfolgen ziehen ? Könnte man vernünftigerweise die 11 von 12 Erfolgen noch als Zufallserfolg einer wirkungslosen Methode ansehen ? Wie gross ist etwa die Erfolgswahrscheinlichkeit, wenn die Behandlung in 61 von 62 Fällen das gewünschte Resultat zeigt ?

Wie genau sind Ergebnisse von Umfragen ?

Wir betrachten eine Umfrage, die aus einer einzigen Frage besteht, welche nur mit "Ja" oder mit "Nein" beantwortet werden kann. Das Ziel ist, den unbekannten Anteil p von Personen in der Population, die mit "Ja" antworten, zu bestimmen. Dazu wird eine Stichprobe von n Personen zufällig ausgewählt. Damit ist die Anzahl Personen in der Stichprobe, die mit "Ja" antworten, eine Zufallsvariable X.

- Wie ist X verteilt, wenn jede Person in der Population die gleiche Wahrscheinlichkeit hat, in die Stichprobe aufgenommen zu werden ?
- Scheint es Ihnen gerechtfertigt, X genähert als binomialverteilt anzunehmen ? Wovon hängt das ab ?
- Wenn bei einer Stichprobe von $n = 1000$ Personen 542 mit "Ja" antworten, was für Schlussfolgerungen kann man dann über den Anteil p der Befürworter in der Population ziehen ? Nehmen Sie dazu an, dass X binomialverteilt ist ?

- Diskutieren Sie, was für Schwierigkeiten auftreten, wenn man solche Umfragen in der Praxis durchführt.

Ändern sich die erwarteten Unfallhäufigkeiten mit der Zeit ?

Gemäss der Webseite des Bundesamts für Strassen
`http://www.astra.admin.ch/unfalldaten/04343/04351/index.html?lang=de`
gab es 2012 in der Schweiz 301 Unfälle mit mindestens einem Todesopfer, und 2013 noch 257 solche Unfälle. Handelt es sich dabei um eine zufällige Schwankung, oder kann man daraus schliessen, dass die erwartete Häufigkeit von solchen Unfällen abgenommen hat ?

Mit Hilfe des folgenden Arguments kann man diese Frage auf einen Binomialtest zurückführen: Wenn es sich nur um eine zufällige Schwankung handelt, dann hätte jeder der $257 + 301 = 558$ Unfälle mit Wahrscheinlichkeit $\frac{1}{2}$ im Jahr 2012 und mit Wahrscheinlichkeit $\frac{1}{2}$ im Jahr 2013 stattfinden können. Damit wären die 257 Unfälle im Jahr 2013 das Ergebnis eines Binomialexperiments mit $n = 558$ und "Erfolgs"wahrscheinlichkeit $\frac{1}{2}$. Ist der Unterschied zwischen der beobachteten relativen Häufigkeit $\frac{257}{558} = 0.461$ und der erwarteten Häufigkeit $\frac{1}{2}$ plausibel ?

Schätzung der Grösse einer Population mit Fang/Wiederfang

Wenn man die Anzahl N von Fischen in einem Teich bestimmen möchte, kann man folgende Methode verwenden: Man fängt zuerst M Fische, markiert sie und setzt sie wieder aus. In einem zweiten Schritt fängt man n Fische und zählt die Anzahl markierter Fische X in dieser Stichprobe.

- Wie ist X verteilt unter der Annahme, dass jeder Fisch beim zweiten Schritt die gleiche Wahrscheinlichkeit hat, gefangen zu werden (unabhängig davon, ob er markiert ist oder nicht) ?
- Wann ist es gerechtfertigt, X als genähert binomialverteilt anzusehen ? Leiten Sie daraus eine Schätzung von N ab, indem Sie erwartete Häufigkeit mit der beobachteten Häufigkeit gleichsetzen.
- Überlegen Sie, wie man ein Vertrauensintervall für N aus dem Vertrauensintervall für die Erfolgswahrscheinlichkeit bei der Binomialverteilung erhalten kann.
- Illustrieren Sie das Vorgehen für $M = 50$, $n = 30$, $X = 6$, bzw. $M = 100$, $n = 60$, $X = 12$.
- Diskutieren Sie mögliche Fehlerursachen bei dieser Methode.

Ein Test zur Feststellung von unehrlichem Verhalten

Folgender Ausschnitt stammt aus einer Studie über die Ehrlichkeit von Bankangestellten (Cohn, Fehr und Maréchal, Nature 516, p. 86–89, 04 December 2014).

"... all subjects anonymously performed a coin tossing task that has been shown to reliably measure dishonest behaviour in an unobtrusive way and to predict rule violation outside the laboratory. The rules required subjects to take any coin, toss it ten times, and report the outcomes online. For each coin toss they could win an amount equal to approximately

USD 20 (as opposed to USD 0) depending on whether they reported "heads" or "tails". Subjects knew in advance whether heads or tails would yield the monetary payoff for a specific coin toss."

- Geben Sie an, wie die Anzahl Antworten, die zu einem Gewinn führen, verteilt ist, i) bei einer Person, die immer ehrlich antwortet, bzw. ii) wenn alle $n = 128$ Personen in der Studie stets ehrlich antworten.
- Begründen Sie, weshalb dieser Test Aussagen ermöglicht über die Ehrlichkeit in der Gruppe aller Personen in der Studie, ohne dass man einem einzelnen Teilnehmer Unehrlichkeit nachweisen kann.
- Überlegen Sie, wie man eventuell abschätzen kann, wie viele Teilnehmer unehrlich waren.

Börsenkurse und American Football

Am 24. Januar 1998 erschien im Tages-Anzeiger ein Artikel mit dem Titel "Alle Börsianer schauen nach San Diego", auf einem ähnlichen Artikel in New York Times basierte. Darin war zu lesen "Steigt die Börse in den nächsten 12 Monaten, fällt sie ? Das hängt davon ab, wer am Sonntag die XXXII. Super-Bowl [den Final der amerikanischen Football-Meisterschaft] gewinnt. ... Jeder der sich an der Börse auskennt, weiss, dass es keinen besseren Kurs-Indikator gibt als die Super-Bowl ... Die Geschichte liefert den Beweis. In der Super-Bowl treffen seit 32 Jahren das beste Team der American Football League (AFL) und der National Football League (NFL) aufeinander. Gewinnt der NFL-Champion auch den ultimativen nationalen Titel,steigt die Börse. Passiert das Unglück, dass die AFL-Mannschaft triumphiert, bröckeln die Kurse. Nach 27 der bisher 31 Finals traf der Super-Bowl-Indikator zu. Das kann kein Zufall sein ..."

- Halten Sie es für plausibel, dass die Börse vom Ausgang der Super-Bowl beeinflusst wird ?
- Wie wahrscheinlich ist es, dass jemand der blind rät, in 27 von 31 Versuchen korrekt vorhersagen kann, ob die Börse im kommenden Jahr steigt oder fällt ? Wie wahrscheinlich ist es, dass diese Person 27 oder mehr richtige Vorhersagen macht ?
- Sind Sie einverstanden mit der Aussage "Das kann kein Zufall sein ?"

Seit 1998 ist der Super-Bowl-Indikator wesentlich weniger erfolgreich: Im Zeitraum 1998-2011 entwickelte sich die Börse ungefähr 10% besser, wenn die AFL-Mannschaft gewann, als wenn die NFL-Mannschaft gewann – genau das Gegenteil von dem, was der Super-Bowl-Indikator besagt. Woran kann dieser scheinbare Widerspruch liegen ?

Aussersinnliche Wahrnehmung

Im Jahr 1973 führte Charles Tart ein Experiment an der University of California in Davis ein Experiment durch, um aussersinnliche Wahrnehmung zu demonstrieren. Er verwendete einen Zufallszahlengenerator auf einem Computer, der eines von vier möglichen Symbolen auswählte. Die Versuchsperson sah das Resultat nicht und musste das richtige Symbol herausfinden. Jede Versuchsperson wiederholte den Versuch 500 mal, und es wurden 15

Personen getestet, welche behaupteten, hellseherische Fähigkeiten zu besitzen. Insgesamt wurde das richtige Symbol in 2006 von $15 \cdot 500 = 7'500$ Versuchen herausgefunden.

- In welchem Bereich schwankt mit Wahrscheinlichkeit 99% die Anzahl richtig geratener Symbole, wenn es keine hellseherischen Fähigkeiten gibt ?
- Wie gross ist die Wahrscheinlichkeit, dass Personen ohne hellseherische Fähigkeiten bei 7500 Versuchen rein zufällig in 2006 oder mehr Fällen das richtige Symbol erraten ?
- Bestimmen Sie ein 99%-Vertrauensintervall für die Erfolgswahrscheinlichkeit der 15 Versuchspersonen im Experiment von Tart.