

## Autoversicherung

Versicherungsfirmen bitten ihre Kunden um viele Informationen: Geschlecht, Alter, Ehestand, Beruf, Adresse, Autotyp, Kilometer pro Jahr, Schadenfreiheitsstufe und mehr. Wieviele verschiedene Kombinationen gibt es? Die Kombinationen werden durch die Reihenfolge der Informationen definiert. Wieviele mögliche Reihenfolgen gibt es? Die Firma will Geschlecht, Schadenfreiheitsstufe, Autotyp und zwei andere Kriterien verwenden. Wieviel Möglichkeiten gibt es?

1. Wie könnte die Anzahl gemeldete Unfälle pro Jahr pro Versicherter modelliert werden?
2. Eine Firma hat  $N$  Versicherten in der Klasse A. Sei  $p$  die Unfallrate für Versicherte dieser Klasse. Wie hängt die beobachtete Rate von  $N$  ab?
3. Die Firma benutzt ein Callcenter in Indien. Es gibt kein Warteschlangensystem für die Anrufe und die Wahrscheinlichkeit, dass es besetzt ist, beträgt  $r$ . Wie oft muss man anrufen, um durchzukommen?

4. Gegeben vier Klassen A, B, C, D. Sei die Anzahl der Versicherten in Klasse  $i$   $N_i$  und die erwartete Unfallrate  $p_i$ , was ist die Verteilung der Anzahl Unfälle insgesamt?
5. Als Bonus am Ende des Jahres werden  $m$  Kunden ohne Unfall ausgewählt. Jeder gewinnt einen Preis. Was ist die Wahrscheinlichkeit, dass keiner aus Klasse A einen Preis gewinnt? Was ist die Wahrscheinlichkeit, dass doppelt so viele aus Klasse A einen Preis gewinnen als aus Klasse B? Gegeben, dass jemand einen Preis gewonnen hat, was ist die Wahrscheinlichkeit, dass er aus Klasse A kommt?

Wie sehen die Formeln aus, wenn man annimmt, dass die  $N_i$ 's und  $m$  groß sind?

## Wer hat es geschrieben?

Die Verfasserschaft gewisser Dokumente ist unsicher. Mehrere Wissenschaftler haben Methoden vorgeschlagen, wie man unter verschiedene Kandidaten entscheiden könnte. Eins davon ist die Häufigkeiten zu untersuchen, womit Autoren gewisse einfache Wörter verwenden.

1. Gegeben sei ein Dokument von  $M$  Wörtern. Welche Verteilung könnte man für die Anzahl “also” ’s bzw. “aber” ’s nehmen?
2. Sei das apriori Glauben an den Autoren A und B durch die Wahrscheinlichkeiten  $p_A$  und  $p_B = 1 - p_A$  gegeben. Aus bekannten Schriften der zwei Autoren, weißt man, dass A “also” und “aber” mit Raten  $\lambda_{A1}$  bzw.  $\lambda_{A2}$  pro 1000 Wörter verwendet.

Es gibt  $n_1$  mal “also” im Dokument und  $n_2$  mal “aber”. Was ist die Wahrscheinlichkeit, dass A das Dokument geschrieben hat?

3. Welche Annahmen sind hier gemacht worden?
4. Wie könnte man die Möglichkeit eines dritten unbekannten Autors im Modell berücksichtigen?

**Amazon ratings** Sei die Verkäufe pro Woche für Buch  $i$   $X_i \sim P(\lambda_i(t))$  in der Woche  $t$ . Wenn diese Zahlen den Ranking am Samstag Abend bestimmen und  $\lambda_i(t)$  und  $\lambda_j(t)$  konstant sind, was ist die Wahrscheinlichkeit, dass Buch  $i$  öfters höher als Buch  $j$  an den Samstag Abenden eines Jahres gerankt wird?

Was ist die Wahrscheinlichkeit, dass mehr Kopien des Buchs  $j$  als des Buchs  $i$  im Jahr verkauft werden?

Sei

$$\lambda_i(t) = \lambda_i e^{-\nu t} \quad t = 1, \dots, 52$$

und

$$\lambda_j(t) = \lambda_j e^{-\nu t} \quad t = 1, \dots, 52$$

wie sehen die Resultate aus?

## Jugendsportmodell

Sei die Dichte der Sportleistung der Kinder vom Alter  $a$

$$f_a(x) \quad 0 \leq x$$

wo  $a$  in Tagen gemessen wird. Die besten  $\alpha$  eines Jahrgangs (d.h. alle die in einem bestimmten Kalenderjahr geboren sind) werden für Spezialtraining ausgewählt. Wenn die Verteilung der Geburtstage gleichverteilt ist, was ist die relative Wahrscheinlichkeit, dass jemand mit einem Geburtstag im Januar ausgewählt wird, im Vergleich zu jemand mit einem Geburtstag im Dezember?

Da wir uns nur für die besten Sportler interessieren, ist die vollständige Form von  $f_a(x)$  unwesentlich. Welche Formen kommen für den rechten Schwanz der Verteilung in Frage?