Hausübung 01 - Wahrscheinlichkeitstheorie

Baier Sebastian, Magdalena Figlmüller, Alice Schwarzböck

2023-11-29

Aufgabe 1

DNA ist bekanntermaßen aus 4 Nukleinsäuren (A, C, G, T) aufgebaut. Kodierende und nichtkodierende Regionen werden dabei unterschieden.

a) Wie wahrscheinlich ist es, in einer nicht-kodierenden Region (von der Länge 10.000 bp), bei der wir annehmen, dass die Basen zufällig angeordnet sind, die Sequenz des Stop-Codons einmal zu änden?

Anzahl der möglichen Triplets: 4^3 = 64
3 Stopp-Codons: –> p(Stop) = 3/64
Gegenwahrscheinlichkeit = 61/64

Es gibt in einer Sequenz mit 10.000 Basen 9998 mögliche Triplets:

Reading Frame 1: 3333 (Position 1 bis 9999)
Reading Frame 2: 3333 (Position 2 bis 10000)
Reading Frame 3: 3332 (Position 3 bis 9998)

Demnach kann die Wahrscheinlichkeit wie folgt berrechnet werden:

$$\binom{9998}{1} * (\frac{3}{64}) * (\frac{61}{64})^{9997}$$

```
stop <- choose(9998,1)*(3/64)*(61/64)**9997
```

Die Wahrscheinlichkeit einmal ein Stopkodon zu finden liegt bei $1.7061417 \times 10^{-204}$

b) Wie wahrscheinlich ist es, in derselben nicht-kodierenden Region die Folge A-A-A-A-C-C-G-A-T-C-G-A-T-T-G-A anzutreffen?

Sequenz AAAACCGATCGATTGA => Länge: 16 Basen
Wahrscheinlichkeit für ein bestimmte Sequenz mit 16 Basen: p(Seq) = (1/4)^16

Es gibt in der nicht-kodierenden Region mit der Länge 10.000 Basen 9985 Möglichkeiten, die Seqeunz anzuordnen (Sequenz der Länge 16 + 9984x um 1 verschieben = 10.000) Somit kann die Wahrsheinlichkeit wie folgt berechnet werden:

$$\binom{9985}{1} * (\frac{1}{4})^{16} * (\frac{3}{4})^{9969}$$

```r
seq16 <- choose(9985,1)*(1/4)**16*(3/4)**9969
```

Die Wahrscheinlichkeit diese Sequenz in der Region aufzufinden, liegt annähernd bei 0

Aufgabe 2

Ermitteln Sie, wie sich die Wahrscheinlichkeiten beim Lottospielen für 5er und 6er verändern wurden, wenn man die Kugel nach dem Ziehen wieder zurücklegen würde. Wie würde sich diese Veränderung auf die Gewinnwahrscheinlichkeiten von Lotterie und Spielern auswirken? Begründen Sie warum man nicht zulässt, dass 6-6-6-6-6-6 eine Möglichkeit für einen Lottosechser bietet?

1) Lotto-6er:

```r
# Möglichkeiten ohne Zurücklegen
(six1 <- choose(45,6))
```

```
## [1] 8145060
```

```r
# Wahrscheinlichkeit ohne Zurücklegen
1/six1
```

```
## [1] 1.227738e-07
```

```r
# Möglichkeiten mit Zurücklegen
(six2 <- factorial(50)/((factorial(6)*factorial(44))))
```

```
## [1] 15890700
```

```r
# Wahrscheinlichkeit ohne Zurücklegen
1/six2
```

```
## [1] 6.292989e-08
```

```r
# Unterschied Erfolgschancen
six2/six1
```

```
## [1] 1.950962
```

Lotto-6er: Für einen Lotto-6er exisiteren 8.14506×10^6 Möglichkeiten (–> Wahrscheinlichkeit: 1.227738×10^{-7}), für einen Lotto-6er mit Zurücklegen sind es 1.58907×10^7(–> Wahrscheinlichkeit: 6.292989×10^{-8}). Somit sind die Erfolgschancen für einen Spieler durch Zurücklegen etwa 1.9509617 x niedriger als ohne Zurücklegen.

–> Aufgrund der nochmals niedrigeren Wahrscheinlichkeit ist 6-6-6-6-6-6 als Lotto-6er nicht zulässig.

2) Lotto-5er:

```r
# Möglichkeiten ohne Zurücklegen
(five1 <- choose(45,5))
```

```
## [1] 1221759
```

```r
# Wahrscheinlichkeit ohne Zurücklegen
1/five1
```

```
## [1] 8.18492e-07
```

```r
# Möglichkeiten mit Zurücklegen
(five2 <- factorial(49)/((factorial(5)*factorial(44))))
```

```
## [1] 1906884
```

```r
# Wahrscheinlichkeit ohne Zurücklegen
1/five2
```

```
## [1] 5.244157e-07
```

```r
# Unterschied Erfolgschancen
five2/five1
```

```
## [1] 1.560769
```

Lotto-5er: Für einen Lotto-5er exisiteren 1.221759×10^6 Möglichkeiten (–> Wahrscheinlichkeit: 8.1849203×10^{-7}), für einen Lotto-5er mit Zurücklegen sind es 1.906884×10^6 (–> Wahrscheinlichkeit: 5.2441575×10^{-7}). Somit sind die Erfolgschancen für einen Spieler durch Zurücklegen etwa 1.5607693x niedriger als ohne Zurücklegen.

Aufgabe 3

Aminosäuren werden durch Codons, die sich aus 3 Nukleinsäuren zusammensetzen, kodiert. Wie viele Aminosäuren könnte man mit dieser Methode eindeutig kodieren? Wie viele sind es wirklich? Begründen Sie den Unterschied.

Mögliche Anzahl an eindeutig kodierten Aminosäuren beträgt: 4ˆ3 = 64

Diese 64 Codons kodieren nur für 20 (21) Aminosäuren aufgrund der Tatsache, dass der genetische Code degeneriert ist (–> mehrere Codons pro Aminosäure, jedoch vereinzelt Ausnahmen: z.B. Methionin).

Aufgabe 4

Ein neu entwickelter HIV-Test entdeckt 98% der tatsäachlich HIV Positiven. Allerdings schlägt er auch bei 5 % der HIV-Negativen an. In Österreich lebten 2009 etwa 15.000 infzierte Menschen, davon 4600 Frauen (laut UNAIDS 2010). Wie groß ist die Wahrscheinlichkeit, dass Sie an HIV erkrankt sind, wenn Sie einen positiven Testbescheid bekommen?

Bevölkerung Österreich 2010 [Quelle: Weltbank, Eurostat]: 8363000
Es wird näherungsweise von einer Verteilung von 50% Männern und 50% Frauen ausgegangen.

```
p_HIV <- 15000/8363000
p_HIVmale <- 10400/4181500
p_HIVfemale <- 4600/4181500
p_HIV_neg <- 1-p_HIV
p_HIVfemale_neg <- 1-p_HIVfemale
p_HIVmale_neg <- 1-p_HIVmale

#Wahrscheinlichkeit an HIV erkrankt zu sein und einen positiven Test zu erhalten
HIVpos_Testpos <- p_HIV * 0.98
HIVmalepos_Testpos <- p_HIVmale * 0.98
HIVfemalepos_Testpos <- p_HIVfemale * 0.98

#Wahrscheinlichkeit für positiven Test
posTest <- HIVpos_Testpos + p_HIV_neg * 0.05
posTestmale <- HIVmalepos_Testpos + p_HIVmale_neg * 0.05
posTestfemale <- HIVfemalepos_Testpos + p_HIVfemale_neg * 0.05

#Wahrscheinlichkeit an HIV zu erkranken bei positiven Test
HIV <- HIVpos_Testpos / posTest
HIVmale <- HIVmalepos_Testpos / posTestmale
HIVfemale <- HIVfemalepos_Testpos / posTestfemale
```

Die Wahrscheinlichkeit als Frau HIVpositiv zu sein bei positivem Testergebnis liegt bei rund 2.1129302 %. Die Wahrscheinlichkeit als Mann HIVpositiv zu sein bei positivem Testergebnis liegt bei rund 4.6592639 %. Die Wahrscheinlichkeit als Mann oder Frau HIVpositiv zu sein bei positivem Testergebnis liegt bei rund 3.4019903%.

Aufgabe 5

Die Krankheit Lupus erythematodes zeigt eine Inzidenz von 25 jährlich Neuerkrankten pro 100.000 Personen. Neun von zehn Betroffenen sind Frauen aufgrund des geschlechtsspezifischen Signalwegs der Interferon-alpha-Induktion. Im AKH Wien werden jährlich etwa 200000 PatientInnen, davon 80000 Patientinnen stationär aufgenommen. Wie wahrscheinlich ist es, dass eine dieser Frauen an Lupus leidet?

```
Lup_year_female <- 25 / 100000 * 0.9
AKH_female <- 80000 / 200000
female_lupAKH <- Lup_year_female / AKH_female * 100
```

Die Wahrscheinlichkeit, dass eine Frau an Lupus leidet liegt bei 0.05625%.

Aufgabe 6

Erklären Sie anhand der beiden Grafken die Begriffe: Laplace-Wahrscheinlichkeit, Variation, Kombination, Permutation, bedingte Wahrscheinlichkeit, statistische Unabhängigkeit.

LAPLACE-WAHRSCHEINLICHKEIT:

Beschreibt die Wahrscheinlichkeit des Eintritts eines bestimmten Ereignisses in Verhältnis zu allen möglichen Ereignissen.

z.B. Wahrscheinlichkeit für grünen Käfer:

$$p(green) = \frac{4}{9}$$

VARIATION:

Wenn für die Berechnung der Wahrscheinlichkeit die Reihenfolge relevant ist, wird von Variation gesprochen.

OHNE ZURÜCKLEGEN (–> gezogene Elemente werden aus der Grundgesamtheit entfernt):
Die Möglichkeiten der Variation von k Elementen aus einer Grundgesamtheit n, werden durch folgende Formel beschrieben:

$$\frac{n!}{n-k!}$$

Werden beispielsweise 2 der 9 Käfer gezogen, exisiteren 9*8 = 72 verschiedene Möglichkeiten, Käfer zu ziehen.

MIT ZURÜCKLEGEN (–> gezogene Elemente werden zur Grundgesamtheit zurückgelegt):

$$n^k$$

Werden hier 2 Käfer gezogen, wird der 1.Käfer nach dem Zug zurückgelegt und kann erneut gezogen werden. Folglich exisiteren mit Zurücklegen mehr Möglichkeiten –> 9*9 = 81.

KOMBINATION:

Wenn die Reihenfolge für die Berechnung der Wahrscheinlichkeit nicht relevant ist, wird von Kombination gesprochen.

OHNE ZURÜCKLEGEN (–> gezogene Elemente werden aus der Grundgesamtheit entfernt):
Die Möglichkeiten der Kombination von k Elementen aus einer Grundgesamtheit n, werden durch folgende Formel beschrieben:

$$\binom{n}{k}$$

Werden unter diesen Bedingungen 2 der 9 Käfer gezogen, exisiteren 36 verschiedene Möglichkeiten. Dies entspricht genau der Hälfte an Möglichkeiten gegenüber Variation ohne Zurücklegen, denn da es unerheblich ist, ob bei einer bestimmten Käferkombination Käfer1 oder Käfer2 zuerst gezogen wird, reduzieren sich die Möglichkeiten um 50%.

MIT ZURÜCKLEGEN (–> gezogene Elemente werden zur Grundgesamtheit zurückgelegt):

$$\frac{(n+k-1)!}{k! * (n-1)!}$$

Es werden erneut 2 der 9 Käfer gezogen, wobei der 1.Käfer zurückgelegt wird und erneut gezogen werden kann. Mit Zurücklegen exisiteren daher mehrere Möglichkeiten, für das konkrete Beispiel sind es 45.

PERMUTATION

Permutation beschreibt die Anzahl an Möglichkeiten die Reihenfolge von Objekten zu verändern. Die Grundformel lautet:

$$n!$$

Für die 9 abgebildeten Käfer gebe es demnach 9! = 362880 mögliche Reihenfolgen (9 für den ersten Käfer, 8 für den 2.Käfer, 7 für den 3.Käfer,. . . ..)

BEDINGTE WAHRSCHEINLICHKEIT

Die bedingte Wahrscheinlichkeit beschreibt die Häufigkeit eines Ereignisses B unter der Voraussetzung, dass Ereignis A eingetreten ist. Ein Beispiel wäre hier die Wahrscheinlichkeit, dass ein gezogener grüner Käfer gepunktet ist.

$$p(spotty|green) = \frac{4}{9} * \frac{1}{4} = \frac{1}{9}$$

STATISTISCHE UNABHÄNGIGKEIT

An letzterem Beispiel kann auch gut statistische Unabhängigkeit demonstriert werden, welche einerseits symmetrisch ist (es ist somit egal, welches der beiden Ereignisse zuerst eintritt) und weiters besagt, dass Ereignisse voneinander unabhängig sind, wenn ihre Eintrittswahrscheinlichkeit miteinander multipliziert der Wahrscheinlichkeit des Eintritts beider Ereignisse zum selben Zeitpunkt entspricht

$$p(spotty + green) = \frac{4}{9} * \frac{1}{4} = \frac{3}{9} * \frac{1}{3} = \frac{1}{9}$$