Onlinebeispiel 10.5. Schätzung der Größe einer Geckopopulation, *Gehyra variegata*, durch Anpassung der Fangfrequenzen an die Geometrische Verteilung mit Programm FREQ

Kapitel 10.2 aus Henle, K., A. Grimm-Seyfarth & B. Gruber: Erfassung und Analyse von Tierpopulationen. Ulmer Verlag

Annegret Grimm-Seyfarth

28.03.2025

Im semi-ariden Kinchega Nationalpark, Australien, wurden 2016 Geckos der Art *Gehyra variegata* an einer Feldstation in sechs Fangnächsten per Hand gefangen. Per Fotoidentifikation können die Individuen bestimmt werden. Zusätzlich erhielten alle gefangenen Tiere Farbmarkierungen, die jedoch aufgrund von Häutungen nur wenige Tage erhalten blieben. Farbmarkierte Geckos wurden als Wiederfänge notiert, aber nicht erneut gefangen. Am Computer wurden zusätzlich Wiederfänge per Foto-ID zugeordnet, die ihre Farbmarkierung verloren hatten. Somit wurde der Fehler durch Verlust der Markierung beseitigt. Die Untersuchung war eine Fortsetzung eines 30-jährigen Populationsmonitorings, um die Auswirkungen des Klimawandels zu erforschen (Henle 1990, Grimm-Seyfarth et al. 2018). In diesem Beispiel wollen wir die Anwendung der Fangfrequenzmodelle zeigen.

```r
# check.packages function: install and load multiple R packages.
# Function from: https://gist.github.com/smithdanielle/9913897
check.packages <- function(pkg){
  new.pkg <- pkg[!(pkg %in% installed.packages()[, "Package"])]
  if (length(new.pkg))
    install.packages(new.pkg, dependencies = TRUE, type = "source")
  sapply(pkg, require, character.only = TRUE)
}

# benoetigte R pakete
#pakete <- c("FREQ")

# Prüfe und installiere
#check.packages(pakete)
# Anmerkung: Bis das Update von FREQ bei CRAN akzeptiert ist,
# laden wir den Quellcode direkt ein als:
UlmerBuch::beispiel.pfad() # Pfad zu den Beispieldaten
```

```
## Der Pfad zu den Beispieldaten wurde gesetzt auf:
##  C:/Users/grimm/AppData/Local/Programs/R/R-4.4.3/library/UlmerBuch/extdata
```

```r
source("FREQ2.2.r")
```

Weitere Informationen zur Nutzung des Paketes finden sich hier:

https://cran.r-project.org/web/packages/FREQ/FREQ.pdf

Daten einladen

Wir laden uns für dieses Beispiel die Fangzahlen der einzelnen Individuen (d. h., die Anzahl an Fanggelegenheiten, an denen ein Individuum gefangen oder per Farbmarkierung gesichtet wurde) in R ein. Es gab 6 Fanggelegenheiten, demnach können einzelne Individuen maximal sechs Mal gefangen worden sein.

```
UlmerBuch::beispiel.pfad() #Pfad zu den Beispieldaten
```

```
## Der Pfad zu den Beispieldaten wurde gesetzt auf:
##  C:/Users/grimm/AppData/Local/Programs/R/R-4.4.3/library/UlmerBuch/extdata
```

```
GV.2016 <- read.csv2("GV_Station_2016_capture_frequencies.csv")
head(GV.2016)
```

```
##   Capture_number
## 1              4
## 2              6
## 3              2
## 4              6
## 5              3
## 6              2
```

Die Anzahl gefangener Individuen beträgt:

```
nrow(GV.2016)
```

```
## [1] 115
```

Vektor der Fangfrequenzen erstellen

Die als Input benötigte Zusammenfassung aller einzelnen Fangfrequenzen erhalten wir folgendermaßen:

```
table(GV.2016)
```

```
## Capture_number
##  1  2  3  4  5  6
## 35 20 21 15 15  9
```

```
f <- as.vector(table(GV.2016))
```

Es wurden 35 Individuen einmal, 20 zweimal, 21 dreimal, 15 viermal, 15 fünfmal und 9 sechsmal gefangen. Nach der Empfehlung, bei größeren Populationen mindestens zehn Exemplare mindestens viermal gefangen zu haben, können wir also alle Verteilungen nutzen. Diese Fangfrequenzen haben wir im Vektor `f` gespeichert. Würde es mehr als sechs Fanggelegenheiten geben, müsste man für alle verbleibenden Fanggelegenheiten eine 0 hinzufügen.

Berechnung der Erwartungswerte

Die verschiedenen Erwartungswerte bei Anpassung an die verschiedenen Verteilungen erhält man folgendermaßen:

```
f.freq <- freq(f)
f.freq
```

```
## $expected_values
##   original data Geometric Distribution Truncated Geometric Distribution
## 1            35              40.214724                        29.061536
## 2            20              26.151906                        23.036500
## 3            21              17.006761                        18.260575
## 4            15              11.059612                        14.474794
## 5            15               7.192140                        11.473881
## 6             9               4.677097                         9.095117
##   Poisson Distribution Negative Binomial Distribution
## 1            23.318630                      29.790113
## 2            30.788841                      30.132815
## 3            27.101442                      23.131757
## 4            17.891746                      14.937055
## 5             9.449374                       8.552732
## 6             4.158839                       4.480059
##
## $estimations
##                                Geometric Distribution
## number of individuals captured                    115
## parameter q or lambda                          0.6503
## population size N                              176.84
## standard error                                   7.15
## lower symmetric 95%-CI                         162.82
## upper symmetric 95%-CI                         190.86
## lower asymmetric 95%-CI                        164.34
## upper asymmetric 95%-CI                        192.51
##                                Truncated Geometric Distribution
## number of individuals captured                              115
## parameter q or lambda                                    0.7927
## population size N                                        145.08
## standard error                                             3.15
## lower symmetric 95%-CI                                    138.9
## upper symmetric 95%-CI                                   151.25
## lower asymmetric 95%-CI                                  139.51
## upper asymmetric 95%-CI                                  151.91
##                                Poisson Distribution
## number of individuals captured                  115
## parameter q or lambda                        2.6407
## population size N                            123.83
## standard error                                 6.38
## lower symmetric 95%-CI                       111.33
## upper symmetric 95%-CI                       136.33
## lower asymmetric 95%-CI                      117.48
## upper asymmetric 95%-CI                      146.43
##                                Negative Binomial Distribution
```

```
## number of individuals captured                              115
## parameter q or lambda                                    2.4208
## population size N                                        135.08
## standard error                                             7.09
## lower symmetric 95%-CI                                   121.19
## upper symmetric 95%-CI                                   148.97
## lower asymmetric 95%-CI                                  125.26
## upper asymmetric 95%-CI                                  154.31
```

Der erste Datensatz (`expected_values`) beinhaltet die Erwartungswerte der Fangfrequenzen nach den verschiedenen Verteilungen. Der zweite Datensatz (`estimations`) beinhaltet die entsprechenden Schätzwerte. Zunächst müssen wir aber die optimalste Verteilung finden.

Dazu vergleichen wir die tatsächlichen Fangfrequenzen mit den Erwartungswerten. Wir nutzen den Fisher-Test als Alternative zum Chi-Quadrat-Test, da dieser hier exaktere Werte liefern sollte.

```r
f.freq$expected_values
```

```
##   original data Geometric Distribution Truncated Geometric Distribution
## 1            35              40.214724                        29.061536
## 2            20              26.151906                        23.036500
## 3            21              17.006761                        18.260575
## 4            15              11.059612                        14.474794
## 5            15               7.192140                        11.473881
## 6             9               4.677097                         9.095117
##   Poisson Distribution Negative Binomial Distribution
## 1            23.318630                      29.790113
## 2            30.788841                      30.132815
## 3            27.101442                      23.131757
## 4            17.891746                      14.937055
## 5             9.449374                       8.552732
## 6             4.158839                       4.480059
```

```r
# mit geometrischer Verteilung
fisher.test(round(f.freq$expected_values[,c(1,2)]))
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  round(f.freq$expected_values[, c(1, 2)])
## p-value = 0.3297
## alternative hypothesis: two.sided
```

```r
# mit doppelt abgeschnittener geometrischer Verteilung
fisher.test(round(f.freq$expected_values[,c(1,3)]))
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  round(f.freq$expected_values[, c(1, 3)])
## p-value = 0.9549
## alternative hypothesis: two.sided
```

```r
# mit Poissonverteilung
fisher.test(round(f.freq$expected_values[,c(1,4)]))
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  round(f.freq$expected_values[, c(1, 4)])
## p-value = 0.1016
## alternative hypothesis: two.sided
```

```r
# mit Negativ Binomialverteilung
fisher.test(round(f.freq$expected_values[,c(1,5)]))
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  round(f.freq$expected_values[, c(1, 5)])
## p-value = 0.3316
## alternative hypothesis: two.sided
```

Kein Test ist signifikant, doch die Poissonverteilung weist den geringsten p-Wert auf. Es besteht also womöglich ein Unterschied zwischen beobachteten und erwarteten Fangfrequenzen einer Poisonverteilung, was bedeutet, es liegt wahrscheinlich individuelle Heterogenität vor (vgl. auch Onlinebeispiele 9.1 und 9.2).

Schauen wir uns die Differenz von beobachteter und erwarteter Fangfrequenz an:

```r
# mit geometrischer Verteilung
summary(f.freq$expected_values[,1] - f.freq$expected_values[,2])
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -6.152  -2.926   3.967   1.450   4.240   7.808
```

```r
# mit doppelt abgeschnittener geometrischer Verteilung
summary(f.freq$expected_values[,1] - f.freq$expected_values[,3])
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -3.03650  0.05996  1.63232  1.59960  3.32945  5.93846
```

```r
# mit Poissonverteilung
summary(f.freq$expected_values[,1] - f.freq$expected_values[,4])
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -10.7888  -5.2990   0.9747   0.3819   5.3733  11.6814
```

```r
# mit Negativ Binomialverteilung
summary(f.freq$expected_values[,1] - f.freq$expected_values[,5])
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -10.1328  -1.5831   2.2914   0.6626   5.0374   6.4473
```

Die geringste Streuung weist die doppelt abgeschnittene geometrische Verteilung auf.

Populationsgrößenschätzung

Die größte Übereinstimmung (höchster p-Wert, geringste Streuung) bestand mit der doppelt abgeschnittenen geometrischen Verteilung, daher schauen wir hier einmal genauer auf die Schätzwerte.

```
f.freq$estimations
```

```
##                                Geometric Distribution
## number of individuals captured                    115
## parameter q or lambda                          0.6503
## population size N                              176.84
## standard error                                   7.15
## lower symmetric 95%-CI                         162.82
## upper symmetric 95%-CI                         190.86
## lower asymmetric 95%-CI                        164.34
## upper asymmetric 95%-CI                        192.51
##                                Truncated Geometric Distribution
## number of individuals captured                              115
## parameter q or lambda                                    0.7927
## population size N                                        145.08
## standard error                                             3.15
## lower symmetric 95%-CI                                    138.9
## upper symmetric 95%-CI                                   151.25
## lower asymmetric 95%-CI                                  139.51
## upper asymmetric 95%-CI                                  151.91
##                                Poisson Distribution
## number of individuals captured                  115
## parameter q or lambda                        2.6407
## population size N                            123.83
## standard error                                 6.38
## lower symmetric 95%-CI                       111.33
## upper symmetric 95%-CI                       136.33
## lower asymmetric 95%-CI                      117.48
## upper asymmetric 95%-CI                      146.43
##                                Negative Binomial Distribution
## number of individuals captured                            115
## parameter q or lambda                                  2.4208
## population size N                                      135.08
## standard error                                           7.09
## lower symmetric 95%-CI                                 121.19
## upper symmetric 95%-CI                                 148.97
## lower asymmetric 95%-CI                                125.26
## upper asymmetric 95%-CI                                154.31
```

Laut der Schätzung der doppelt abgeschnittenen geometrischen Verteilung (`Truncated Geometric Distribution`) liegt die Populationsgröße bei 145 Individuen, und mit 95 %-Wahrscheinlichkeit zwischen 139 und 151 bzw. 140 und 152 Individuen, womit die symmetrische und asymetrische Schätzung sehr nahe beieinander liegen.

Neben der Populationsgröße N und dem Konfidenzintervall (`CI`) gibt die Ergebnistabelle noch den Wert für die Verteilung (`q` bzw. `lambda`) sowie die Anzahl insgesamt gefangener Individuen an.

Literaturverzeichnis

Grimm-Seyfarth, A., Mihoub, J.-B., Gruber, B., Henle, K. (2018): Some like it hot: from individual to population responses of an arboreal arid-zone gecko to local and distant climate. — Ecological Monographs 88: 336—352.

Henle, K. (1990): Population ecology and life history of the arboreal gecko *Gehyra variegata* in arid Australia. — Herpetological Monographs 4: 30–60.