SEEC Toolbox Seminars

Estimating survival from capture-mark-recapture datasets

Res Altwegg

Statistics in Ecology, Environment and Conservation

05 March 2020

SEEC - Statistics in Ecology, Environment and Conservation

Capture-mark-recapture

Chris Oosthuizen

Mark Anderson

David Maphisa

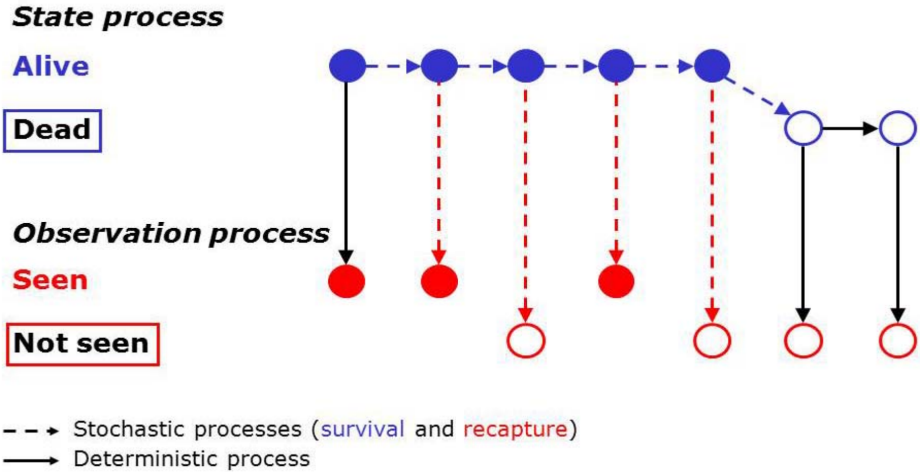
SEEC - Statistics in Ecology, Environment and Conservation

Non-detection is common

from: Gurney's guide to feathered friends 1969, by E. Gurney and N Gurney

SEEC - Statistics in Ecology, Environment and Conservation

Survival



State process

Alive

Dead

Observation process

Seen

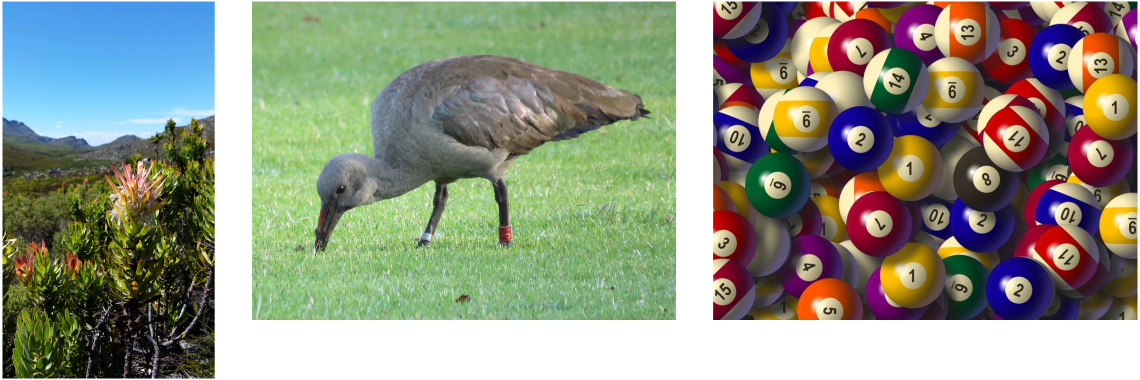
Not seen

- - → Stochastic processes (survival and recapture)

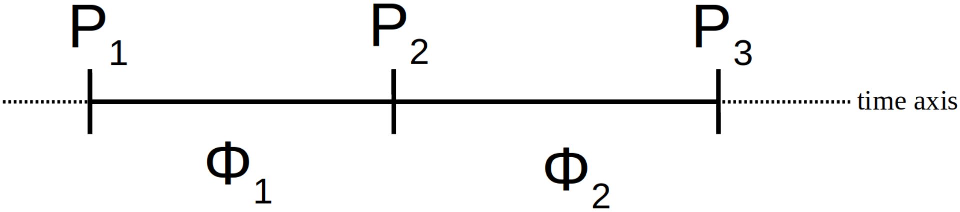
→ Deterministic process

Schaub and Kéry 2012, Academic Press

Real world vs model world



CMR experiment



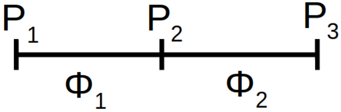
P: recapture probability

Φ: survival probability

SEEC - Statistics in Ecology, Environment and Conservation

Capture histories

Capture history	Meaning	Probability
111	ringed at occasion 1, bird survived first period, seen at occasion 2, bird survived second period, seen at occasion 3	$\Phi_1 P_2 \Phi_2 P_3$
101	ringed at occasion 1, bird survived first period, not seen at occasion 2, bird survived second period, seen at occasion 3	$\Phi_1 (1-P_2) \Phi_2 P_3$
110	ringed at occasion 1, bird survived first period, seen at occasion 2, bird survived second period, not seen at occasion 3 or bird did not survive second period	$\Phi_1 P_2 [\Phi_2 (1-P_3)+(1- \Phi_2)]$ $= \Phi_1 P_2 (1- \Phi_2 P_3)$

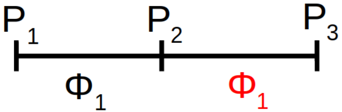


Parameters P and Φ estimated by maximum likelihood

P_1 not estimated!

Capture histories

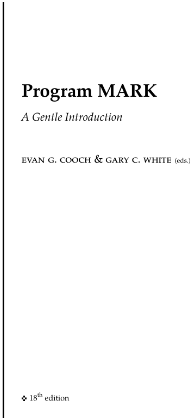
Capture history	Meaning	Probability
111	ringed at occasion 1, bird survived first period, seen at occasion 2, bird survived second period, seen at occasion 3	$\Phi_1 P_2 \Phi_1 P_3$
101	ringed at occasion 1, bird survived first period, not seen at occasion 2, bird survived second period, seen at occasion 3	$\Phi_1(1-P_2)\Phi_1 P_3$
110	ringed at occasion 1, bird survived first period, seen at occasion 2, bird survived second period, not seen at occasion 3 or bird did not survive second period	$\Phi_1 P_2[\Phi_1(1-P_3)+(1-\Phi_1)]$ $= \Phi_1 P_2(1-\Phi_1 P_3)$



Parameters P and Φ estimated by maximum likelihood

P_1 not estimated!

‘MARK book’



Program MARK

A Gentle Introduction

EVAN G. COOCH & GARY C. WHITE (eds.)

18th edition

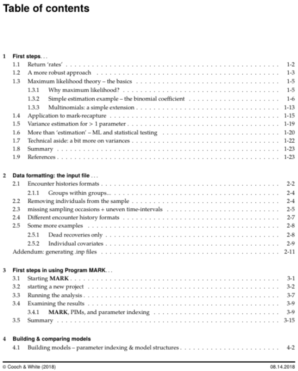


Table of contents

1 First steps. . .

2 Data formatting: the input file . . .

3 First steps in using Program MARK. . .

4 Building & comparing models

http://www.phidot.org/software/mark/docs/

Data wrangling: hadeda data

```
head(ch) # here we've got our capture history
```

```
## 
##        2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017
##   -       0    0    1    0    0    0    0    0    0    0    0    0
##   AA      1    0    0    0    0    0    0    0    0    0    0    0
##   AB      1    0    0    0    0    0    0    0    0    0    0    0
##   AC      1    0    0    0    0    0    0    0    0    0    0    0
##   AD      1    0    0    0    0    0    0    0    0    0    0    0
##   AF      1    0    0    0    0    0    0    0    0    0    0    0
```

```
dim(ch)  # we have data on 261 individuals for 12 years
```

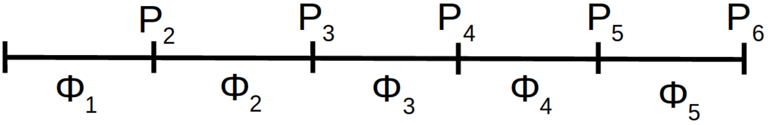
```
## [1] 261  12
```

Input format for MARK

```r
write.table(cbind(ch," ",1,";"),"hadeda.inp",sep="",
              row.names=F,col.names=F,quote=F)
```

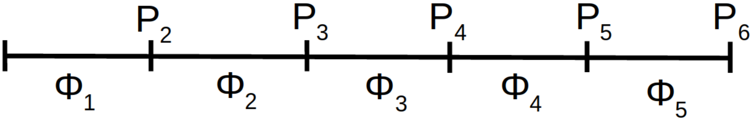
```
001000000000 1;
100000000000 1;
100000000000 1;
100000000000 1;
100000000000 1;
100000000000 1;
```

Parameter Index Matrix



$$\begin{array}{ccccccl}
\Phi_1 & \Phi_2 & \Phi_3 & \Phi_4 & \Phi_5 & \longrightarrow & \text{cohort 1} \\
 & \Phi_2 & \Phi_3 & \Phi_4 & \Phi_5 & \longrightarrow & \text{cohort 2} \\
 & & \Phi_3 & \Phi_4 & \Phi_5 & \longrightarrow & \text{cohort 3} \\
 & & & \Phi_4 & \Phi_5 & \longrightarrow & \text{cohort 4} \\
 & & & & \Phi_5 & \longrightarrow & \text{cohort 5}
\end{array}$$

Parameter Index Matrix



$$\underbrace{\begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \\ & 2 & 3 & 4 & 5 \\ & & 3 & 4 & 5 \\ & & & 4 & 5 \\ & & & & 5 \end{array}}_{\text{“PIM”}} \qquad \begin{array}{l} \longrightarrow \text{cohort 1} \\ \longrightarrow \text{cohort 2} \\ \longrightarrow \text{cohort 3} \\ \longrightarrow \text{cohort 4} \\ \longrightarrow \text{cohort 5} \end{array}$$

MARK analysis using RMark

```r
library(RMark)

# MARK input file with capture histories
ch1 <- apply(ch, 1 , paste , collapse = "" )
hade <- data.frame(ch=ch1, group = 1)
hade$ch <- as.character(hade$ch)

# run default model Phi()P()
model1 <- mark(hade)
```

```
##
## Output summary for CJS model
## Name : Phi(~1)p(~1)
##
## Npar :  2
## -2lnL:  964.9276
## AICc :  968.9553
```

Output

```
Beta
                     estimate        se        lcl        ucl
Phi:(Intercept)  0.5803299 0.1004607  0.3834269  0.7772329
p:(Intercept)   -0.3789708 0.1363474 -0.6462117 -0.1117299


Real Parameter Phi

          1         2         3         4         5         6         7
1 0.6411433 0.6411433 0.6411433 0.6411433 0.6411433 0.6411433 0.6411433 0
2           0.6411433 0.6411433 0.6411433 0.6411433 0.6411433 0.6411433 0
3                     0.6411433 0.6411433 0.6411433 0.6411433 0.6411433 0
4                               0.4063752 0.4063752 0.4063752 0.4063752 0
5                                         0.4063752 0.4063752 0.4063752 0
```

Take a bit more control

```
hade.process <- process.data(hade,model="CJS", begin.time=2006)
hade.ddl <- make.design.data(hade.process)
head(hade.ddl$Phi)
```

```
##   par.index model.index group cohort age time occ.cohort Cohort Age Tim
## 1         1           1     1   2006   0 2006          1      0   0
## 2         2           2     1   2006   1 2007          1      0   1
## 3         3           3     1   2006   2 2008          1      0   2
## 4         4           4     1   2006   3 2009          1      0   3
## 5         5           5     1   2006   4 2010          1      0   4
## 6         6           6     1   2006   5 2011          1      0   5
```

Working with constraints

“all different” PIM

1	2	3	4	5
	6	7	8	9
		10	11	12
			13	14
				15

$$logit(\Phi) = \beta_0 + \sum \beta_i \times x_i$$

x_i are covariates

- time, age, sex
- environmental, e.g. rain
- individual, e.g. size at fledging

β are parameters to be estimated

“Design Data”

- Mapping x_i onto “all different” PIM
- Parameter number in rows
- x in columns

```r
head(hade.ddl$Phi)
```

```
##   par.index model.index group cohort age time occ.cohort Cohort Age Tim
## 1         1           1     1   2006   0 2006          1      0   0
## 2         2           2     1   2006   1 2007          1      0   1
## 3         3           3     1   2006   2 2008          1      0   2
## 4         4           4     1   2006   3 2009          1      0   3
## 5         5           5     1   2006   4 2010          1      0   4
## 6         6           6     1   2006   5 2011          1      0   5
```

Prepare some model structures and run the fully time-dependent model

```
dot <- list(formula=~1)
time <- list(formula=~time)
age <- list(formula=~age)
timeage <- list(formula=~time + age)

# runs the fully time-dependent model
Phi.time.P.time <- mark(hade.process,hade.ddl,
                  model.parameters=list(Phi=time, p=time))
```

Summary output

```r
summary(Phi.time.P.time)
```

```
Output summary for CJS model
Name : Phi(~time)p(~time)

Npar :  22  (unadjusted=21)
-2lnL:  930.9508
AICc :  977.4012  (unadjusted=975.18273)

Beta
                  estimate        se        lcl        ucl
Phi:(Intercept)  1.0822175 1.3642741 -1.5917597  3.7561947
...
```

Have a look at the parameter estimates

```r
Phi.time.P.time$results$real   # parameter estimates
```

```
                        estimate        se       lcl       ucl
Phi g1 c2006 a0 t2006  0.7469134 0.2578939 0.1691365 0.9771613
Phi g1 c2006 a1 t2007  0.3558946 0.0828915 0.2138338 0.5288463
Phi g1 c2006 a2 t2008  0.8116639 0.1377452 0.4242693 0.9618373
...
Phi g1 c2006 a10 t2016 0.7236854 0.0000000 0.7236854 0.7236854

p g1 c2006 a1 t2007    0.3966944 0.1692755 0.1411761 0.7245293
p g1 c2006 a2 t2008    0.3959044 0.1101140 0.2099931 0.6177118
...
p g1 c2006 a11 t2017   0.5922065 0.0000000 0.5922065 0.5922065
```

SEEC - Statistics in Ecology, Environment and Conservation

Counting parameters...

- MARK tries to work out how many parameters were estimable
- RMark assumes all specified parameters are estimable
- You need to check and adjust if necessary

```
Phi.time.P.time <- adjust.parameter.count(Phi.time.P.time, 21)
```

```
##
## Number of parameters adjusted from 22 to 21

## Adjusted AICc = 975.182724057971
## Unadjusted AICc = 975.18273
```

Run some more models

```
Phi.dot.P.dot <- mark(hade.process,hade.ddl,
                 model.parameters=list(Phi=dot, p=dot))
Phi.age.P.dot <- mark(hade.process,hade.ddl,
                 model.parameters=list(Phi=age, p=dot))
Phi.age.P.time <- mark(hade.process,hade.ddl,
                  model.parameters=list(Phi=age, p=time))
Phi.timeage.P.dot <- mark(hade.process,hade.ddl,
                     model.parameters=list(Phi=timeage, p=dot))
Phi.timeage.P.time <- mark(hade.process,hade.ddl,
                      model.parameters=list(Phi=timeage, p=time))
```

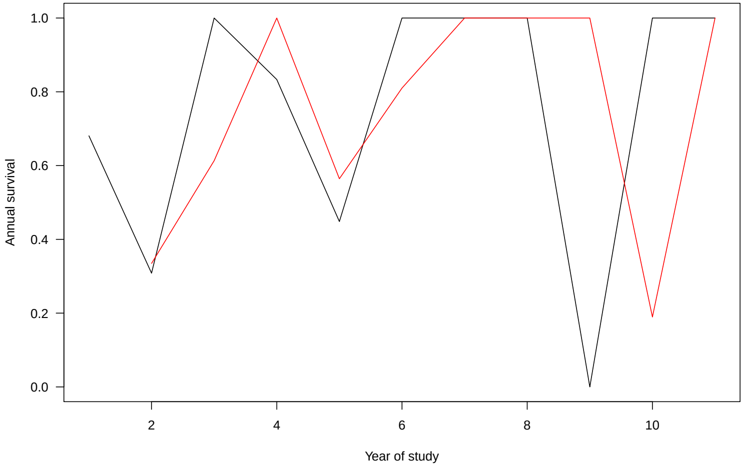
Model selection

```
hade.cjs.results <- collect.models()
hade.cjs.results
```

```
##                         model npar     AICc DeltaAICc       weight Deviance
## 5    Phi(~time + age)p(~1)   22 950.0441   0.00000 9.955527e-01 250.3671
## 1           Phi(~age)p(~1)   12 961.5801  11.53601 3.112092e-03 283.6159
## 6 Phi(~time + age)p(~time)   32 963.4889  13.44478 1.198304e-03 241.0215
## 3             Phi(~1)p(~1)    2 968.9553  18.91115 7.790406e-05 311.7009
## 2        Phi(~age)p(~time)   22 969.6318  19.58769 5.554582e-05 269.9548
## 4       Phi(~time)p(~time)   21 975.1827  25.13858 3.461711e-06 277.7241
```

Plotting results

```r
plot(1:11, Phi.timeage.P.dot$results$real$estimate[1:11], type='l', las=1,
     ylab = "Annual survival", xlab = "Year of study")
lines(2:11, Phi.timeage.P.dot$results$real$estimate[12:21], col='red')
```



Revisiting age

```
head(hade.ddl$Phi)
```

```
##   par.index model.index group cohort age time occ.cohort Cohort Age Time
## 1         1           1     1   2006   0 2006          1      0   0    0
## 2         2           2     1   2006   1 2007          1      0   1    1
## 3         3           3     1   2006   2 2008          1      0   2    2
## 4         4           4     1   2006   3 2009          1      0   3    3
## 5         5           5     1   2006   4 2010          1      0   4    4
## 6         6           6     1   2006   5 2011          1      0   5    5
```

```
ageclass <- NA
ageclass[hade.ddl$Phi$Age == 0] <- "juvenile"
ageclass[hade.ddl$Phi$Age ==1] <- "immature"
ageclass[hade.ddl$Phi$Age > 1] <- "adult"
hade.ddl$Phi$ageclass <- as.factor(ageclass)
```

Look at the design data

```r
head(hade.ddl$Phi, n=15)
```

```
   par.index model.index group cohort age time occ.cohort Cohort Age Time ageclass
1          1           1     1   2006   0 2006          1      0   0    0 juvenile
2          2           2     1   2006   1 2007          1      0   1    1 immature
3          3           3     1   2006   2 2008          1      0   2    2    adult
4          4           4     1   2006   3 2009          1      0   3    3    adult
5          5           5     1   2006   4 2010          1      0   4    4    adult
6          6           6     1   2006   5 2011          1      0   5    5    adult
7          7           7     1   2006   6 2012          1      0   6    6    adult
8          8           8     1   2006   7 2013          1      0   7    7    adult
9          9           9     1   2006   8 2014          1      0   8    8    adult
10        10          10     1   2006   9 2015          1      0   9    9    adult
11        11          11     1   2006  10 2016          1      0  10   10    adult
12        12          12     1   2007   0 2007          2      1   0    1 juvenile
13        13          13     1   2007   1 2008          2      1   1    2 immature
14        14          14     1   2007   2 2009          2      1   2    3    adult
15        15          15     1   2007   3 2010          2      1   3    4    adult
```

Run models with age class

```r
ageclass <- list(formula=~ageclass)
timeageclass <- list(formula=~time + ageclass)

Phi.ageclass.P.dot <- mark(hade.process,hade.ddl,
                      model.parameters=list(Phi=ageclass, p=dot))
Phi.timeageclass.P.dot <- mark(hade.process,hade.ddl,
                          model.parameters=list(Phi=timeageclass, p=dot))
```

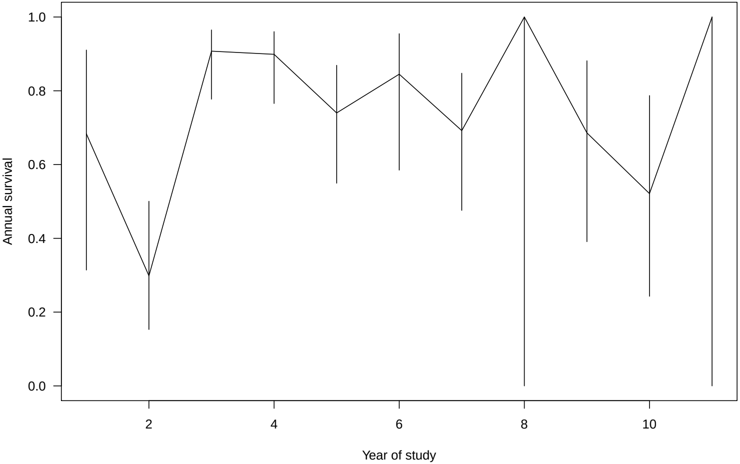
Model selection

```r
hade.cjs.results <- collect.models()
hade.cjs.results
```

```
##                         model npar     AICc DeltaAICc       weight Deviance
## 8 Phi(~time + ageclass)p(~1)   14 947.8054  0.000000 5.936613e-01 265.5811
## 3        Phi(~ageclass)p(~1)    4 949.8681  2.062723 2.116529e-01 288.5486
## 6      Phi(~time + age)p(~1)   22 950.0441  2.238758 1.938200e-01 250.3671
## 1             Phi(~age)p(~1)   12 961.5801 13.774764 6.058802e-04 283.6159
## 7   Phi(~time + age)p(~time)   32 963.4889 15.683540 2.332928e-04 241.0215
## 4               Phi(~1)p(~1)    2 968.9553 21.149909 1.516681e-05 311.7009
## 2          Phi(~age)p(~time)   22 969.6318 21.826448 1.081398e-05 269.9548
## 5         Phi(~time)p(~time)   21 975.1827 27.377339 6.739459e-07 277.7241
```

SEEC - Statistics in Ecology, Environment and Conservation

Plotting results



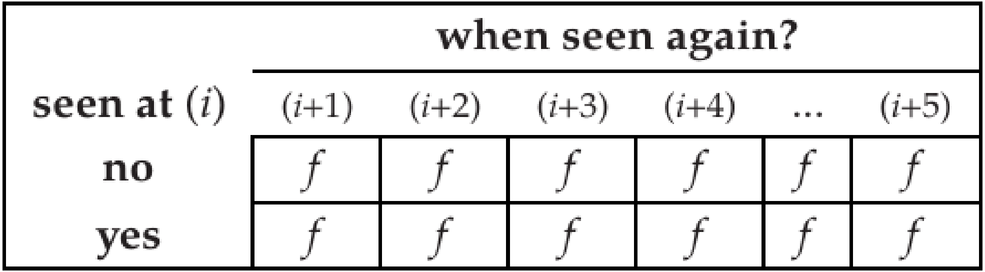
SEEC - Statistics in Ecology, Environment and Conservation

Assumptions

1. every marked animal present in the population at time (i) has the same probability of recapture (p_i)
2. every marked animal in the population immediately after time (i) has the same probability of surviving to time (i+1)
3. marks are not lost or missed.
4. all samples are instantaneous, relative to the interval between occasion (i) and (i+1), and each release is made immediately after the sample.

Goodness-of-fit test: ‘Test 2’

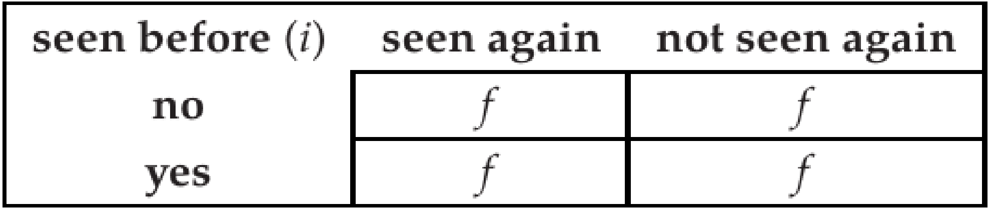
- Assumption: all marked individuals have same recapture probability.
- ‘of those marked animals not seen at $i+1$, but known to be alive at $i+1$ (since they were captured after $i+1$, does when they were next seen $(i+2, i+3\ldots)$ depend on whether or not they were seen at i?’



	when seen again?					
seen at (i)	$(i+1)$	$(i+2)$	$(i+3)$	$(i+4)$	...	$(i+5)$
no	f	f	f	f	f	f
yes	f	f	f	f	f	f

Goodness-of-fit test: ‘Test 3’

- Assumption: all marked individuals have same survival probability.
- ‘Of those individuals seen at occasion i, how many were ever seen again?’



seen before (i)	seen again	not seen again
no	f	f
yes	f	f

GOF testing

```r
require(R2ucare) # load U-CARE

ch.gof <- as.matrix(ch)

freq <- rep(1, length=dim(ch.gof)[1])

overall_CJS(ch.gof,freq)
```

```
##                            chi2 degree_of_freedom p_value
## Gof test for CJS model: 83.231                31       0
```

SEEC - Statistics in Ecology, Environment and Conservation

Detailed tests

```r
test3sr(ch.gof,freq)
```

```
$test3sr
     stat        df     p_val sign_test
   17.894    10.000     0.057     2.513

$details
   component   stat p_val signed_test   test_perf
1          2 0.182  0.67      -0.427      Fisher
2          3 4.027 0.045       2.007 Chi-square
...
```

```r
test3sm(ch.gof,freq)
```

```
$test3sm
 stat    df p_val
5.614 6.000 0.468

$details
   component   stat df p_val   test_perf
1          2 0.936  1 0.333      Fisher
2          3 3.155  1 0.076 Chi-square
...
```

```r
test2ct(ch.gof,freq)
```

```
$test2ct
     stat        df     p_val sign_test
   53.996     9.000     0.000    -6.705

$details
  component dof   stat p_val signed_test   test_perf
1         2   1  3.452 0.063      -1.858      Fisher
2         3   1  8.155 0.004      -2.856 Chi-square
...
```

```r
test2cl(ch.gof,freq)
```

```
$test2cl
 stat    df p_val
5.727 6.000 0.454

$details
  component dof  stat p_val   test_perf
1         2   1 1.913 0.167      Fisher
2         3   1 0.748 0.387 Chi-square
...
```

M-array

```r
marray(ch.gof, freq)
```

```
$R           $m                                                                   $never
      [,1]           [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11]           [,1]
 [1,]   27     [1,]     8    0    1    2    0    0    0    0    0     0     0   [1,]    16
 [2,]   73     [2,]     0   12    0    2    1    0    1    0    1     0     1   [2,]    55
 [3,]   59     [3,]     0    0   19    3    3    2    0    1    1     0     0   [3,]    30
 [4,]   95     [4,]     0    0    0   32    5    0    0    2    1     0     0   [4,]    55
 [5,]   65     [5,]     0    0    0    0   17    2    1    3    0     0     0   [5,]    42
 [6,]   29     [6,]     0    0    0    0    0   15    1    1    0     0     0   [6,]    12
 [7,]   29     [7,]     0    0    0    0    0    0    9    1    3     0     0   [7,]    16
 [8,]   16     [8,]     0    0    0    0    0    0    0   11    0     0     0   [8,]     5
 [9,]   20     [9,]     0    0    0    0    0    0    0    0    8     0     1   [9,]    11
[10,]   16    [10,]     0    0    0    0    0    0    0    0    0     6     1  [10,]     9
[11,]    7    [11,]     0    0    0    0    0    0    0    0    0     0     3  [11,]     4
```

Measure of overdispersion

```
chi2 <- overall_CJS(ch.gof,freq)$chi2  # test statistic
df <- overall_CJS(ch.gof,freq)$degree_of_freedom # degrees of freedom
chi2 / df
```

```
## [1] 2.684871
```

Reasons for lack of fit and possible solutions

- Test 2 significant:
  - ‘trap response’ → ‘time since capture’ models
  - non-random temporary emigration → multi-state models
- Test 3 significant:
  - age effect → age models
  - cohorts differ → cohort models
  - transients → ‘time since first capture’ models
- Non-systematic deviations, general overdispersion → apply overdispersion factor, $\hat{C}$ if below ~ 3

Toolbox Seminars

SEEC STATS TOOLBOX 2020 SCHEDULE

Date	Topic	Speaker
13 February	Spatio-temporal point pattern analysis	Mzabalazo Ngwenya
5 March	Capture-recapture	Res Altwegg
16 April	Reproducible research	Glenn Moncrieff
7 May	TBD	TBD
18 June	TBD	TBD
9 July	TBD	TBD
30 July	Time-to-first detection (TTD) occupancy models	Dominic Henry
20 August	TBD	TBD
10 September	MCMC	Allan Clark
1 October	Spatial GLMMs	David Maphisa
12 November	Species richness and diversity	Jasper Slingsby

SEEC - Statistics in Ecology, Environment and Conservation