

Présentation données TD1

Marie-Pierre Etienne, Etienne Rivot

11/19/2019

Section 1

Contenu du TD

Contexte

Dans une étude publiée en 2016, (un extrait est disponible ici, les auteurs s'intéressent à l'évolution de la population de phoques sur l'île Marion entre 1995 et 2015.

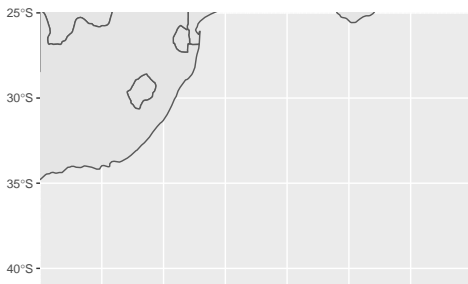
```
marion_coord <- data.frame(longitude = c(37.7), latitude =  
ggplot(data = world) +  
  geom_sf() +  
  geom_point(data = marion_coord, aes(x = longitude, y = latitude,  
    shape = 23, fill = "darkred")) +  
  coord_sf(xlim = c(5, 60), ylim = c(-62, 20), expand = FALSE)
```



Contexte

Dans une étude publiée en 2016, (un extrait est disponible [ici](#)), les auteurs s'intéressent à l'évolution de la population de phoques sur l'île Marion entre 1995 et 2015.

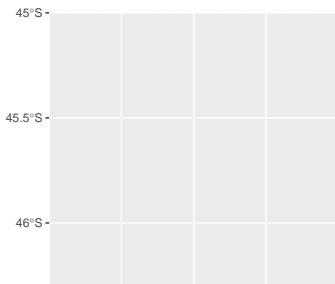
```
ggplot(data = world) +  
  geom_sf() +  
  geom_point(data = marion_coord, aes(x = longitude, y = latitude,  
    shape = 23, fill = "darkred")) +  
  coord_sf(xlim = c(20, 55), ylim = c(-62, -25), expand = FALSE)
```



Contexte

Dans une étude publiée en 2016, (un extrait est disponible [ici](#)), les auteurs s'intéressent à l'évolution de la population de phoques sur l'île Marion entre 1995 et 2015.

```
ggplot(data = world) +  
  geom_sf() +  
  geom_point(data = marion_coord, aes(x = longitude, y = latitude,  
    shape = 23, fill = "darkred")) +  
  coord_sf(xlim = c(37, 39), ylim = c(-48, -45), expand = FALSE)
```



Les données de Cape Davis

Les données de marquage sont disponibles ici

```
CMR_Davis <- read.table('data/CMR_Davis.csv', sep = ',', header = TRUE)
CMR_Davis %>% group_by(Year) %>% summarise(n_transect = n())
```

```
## # A tibble: 10 x 3
##   Year n_transect n_rep
##   <int>      <int> <int>
## 1  1995          3      2
## 2  2007          8      1
## 3  2008          8      1
## 4  2009          8      3
## 5  2010          9      1
## 6  2011          6      3
## 7  2012          6      3
## 8  2013          7      3
## 9  2014          7      3
## 10 2015         10      0
```

Modélisation d'une observation

On ne considère dans un premier temps qu'une seule opération de recapture pour une année donnée

```
CMR_Davis %>% filter(Year == 2007, Transect == 1) -> CMR_2007_T1
```

##	Year	Transect	Repeat	M	NM	n	R	Site
##	1	2007	1	109	11	19	8	MM046M

On a marqué $m=109$ individus, 8 ont été capturés marqués et 11 individus non marqués ont été capturés.

But : Estimer la taille de la population.

Proposer un modèle permettant de le faire et le faire tourner dans jags.

Modélisation d'une observation

On ne considère dans un premier temps qu'une seule opération de recapture pour une année donnée

```
CMR_Davis %>% filter(Year == 2007, Transect == 1) -> CMR_2007_T1
```

```
##   Year Transect Repeat    M NM   n R   Site
## 1 2007         1       1 109 11 19 8 MM046M
```

On a marqué $m=109$ individus, 8 ont été capturés marqués et 11 individus non marqués ont été capturés.

But : Estimer la taille de la population.

Proposer un modèle permettant de le faire et le faire tourner dans jags.

Le modèle est donc

$$NM \sim \mathcal{B}(\text{Size}_{pop} - m, p), \quad R \sim \mathcal{B}(m, p)$$

Modélisation pour une année, plusieurs transects

```
CMR_Davis %>% filter(Year == 2007) -> CMR_2007
CMR_2007
```

##	Year	Transect	Repeat	M	NM	n	R	Site
## 1	2007	1	1	109	11	19	8	MM046M
## 2	2007	2	1	109	61	66	5	MM046M
## 3	2007	3	1	109	77	91	14	MM046M
## 4	2007	4	1	109	106	120	14	MM046M
## 5	2007	5	1	109	44	47	3	MM046M
## 6	2007	6	1	109	57	61	4	MM046M
## 7	2007	7	1	109	18	29	11	MM046M
## 8	2007	8	1	109	54	55	1	MM046M

Modélisation pour une année, plusieurs transects

```
model{  
  
  for(i in 1:NObs){  
    R[i] ~ dbin(p, M[i])  
    NM[i] ~ dbin(p, Size_pop - M[i])  
  }  
  
  ## loi a priori  
  p ~ dbeta(a, b)  
  Size_pop ~ dpois(mean_pop_prior)  
}
```

Modélisation pour plusieurs années, pour Cape Davis

Options à discuter - Même probabilité de capture ? - Les probabilité de capture se ressemblent ? - Que faire des répétitions ?

Pour des raisons pratiques, créer une colonne année, qui vaut 1 en 1995, 13 en 2007, 14 en 2008, etc ...

Modélisation pour une année, plusieurs années

```
model{  
  
  for(i in 1:NObs){  
    R[i] ~ dbin(p, M[i])  
    NM[i] ~ dbin(p, Size_pop[annee[i]] - M[i])  
  }  
  
  ## loi a priori  
  p ~ dbeta(a, b)  
  for( an in 1:NAnnee){  
    Size_pop[an] ~ dpois(mean_pop_prior)  
  }  
  
}
```

Mise en oeuvre

- Estimer les paramètres du modèle
- Vérifier la convergence des algorithmes
- Comparer les priors et les posteriors
- Réaliser un graphique permettant de voir l'évolution de la taille de la population sur Cape Davis

Comptages complets Cape Davis

Chaque année des comptages directs sont réalisés pour Cape Davis.
Les données sont dans le fichier CDDirectCount_Davis.csv

Proposer un modèle permettant d'intégrer ces données

Comment ces données font-elles évoluer la connaissance disponible
sur la taille de la population ?

Comptages depuis la falaise

Le fichier `ClifftopCounts_Davis.csv` contient des comptages effectués depuis la falaise, sans descendre sur la plage. Ces comptages sont disponibles pour Cape Davis mais également d'autres plages difficilement accessibles.

Ces données permettent d'estimer une probabilité de détecter p_{detect} un individu de puis la falaise pour Cape Davis. En supposant que cette probabilité est la même pour toute les plages, on peut donner une estimation de la taille des populations sur toutes les plages comptées depuis la falaise.

Section 2

Et ensuite

Comptages complets pour les plages faciles d'accès

Chaque plage accessible a été comptée directement au cours des dernières années de suivi. (fichier `AnnualCountData.csv`).

Prendre en compte ces nouvelles données pour obtenir le modèle final.

Comptages depuis la falaise

Le fichier `ClifftopCounts.csv` contient des comptages effectués depuis la falaise, sans descendre sur la plage. Ces comptages sont disponibles pour Cape Davis mais également d'autres plages difficilement accessibles.

Ces données permettent d'estimer une probabilité de détecter p_{detect} un individu de puis la falaise pour Cape Davis. En supposant que cette probabilité est la même pour toute les plages, on peut donner une estimation de la taille des populations sur toutes les plages comptées depuis la falaise.

Modifier le modèle en conséquence