Scaling

BIO401-01/598-02

2021-03-22 Mon

REVIEW : Git

Concepts

- created by *Linus Torvalds* in 2005
- distributed version control : each directory as a full-fledged repo
- used for changes tracking and work coordination among collaborators

REVIEW : Git

Concepts

- created by *Linus Torvalds* in 2005
- distributed version control : each directory as a full-fledged repo
- used for changes tracking and work coordination among collaborators

Basic Practice

- create a local copy of your SE.data repo
- git sync. with SE server
- local sync. with rsync

REVIEW : Git

Concepts

- created by *Linus Torvalds* in 2005
- distributed version control : each directory as a full-fledged repo
- used for changes tracking and work coordination among collaborators

Basic Practice

- create a local copy of your SE.data repo
- git sync. with SE server
- local sync. with rsync

Advanced

- git repo setup
- good for professional development
- good for collaboration

Git Procedure

Local Synchronisation

- routine after the first time

```
$ cd ~/SE_data
$ git pull # (sync. w/ server)

$ rsync -hvrPt --ignore-existing  ~/SE_data/* \
  /media/sf_LVM_Shared/my_SE_data
#(sync. only new files)

$ cd /media/sf_LVM_Shared/my_SE_data # (work here)
```

Git Procedure

Replacement (only when needed)

```
example : replacing the exising copy

$ cp ~/SE_data/exercise/05_R_Intro.ipynb \
  /media/sf_LVM_Shared/my_SE_data/exercise

or, if you want to keep the old copy

$ myPath=/media/sf_LVM_Shared/my_SE_data/exercise
$ mv ${myPath}/05_R_Intro.ipynb \
  ${myPath}/05_R_Intro.ipynb
$ cp ~/SE_data/exercise/05_R_Intro.ipynb ${myPath}
```

about the Course

- methodology instead of recipe

about the Course

- methodology instead of recipe
- active role

about the Course

- methodology instead of recipe
- active role
- efficiency and effectiveness

CW Case : setup

objectives

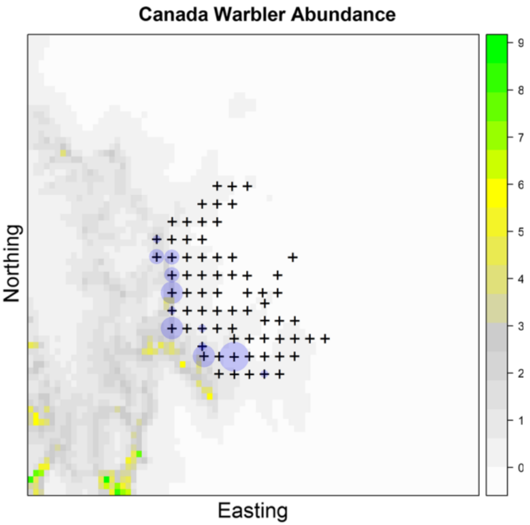
- modelling Canada Warbler (CW) abundance
- scaling effects

landscape variables

- elevation : CW high elevation
- evergreen coverage (NDVI) : understory thickets

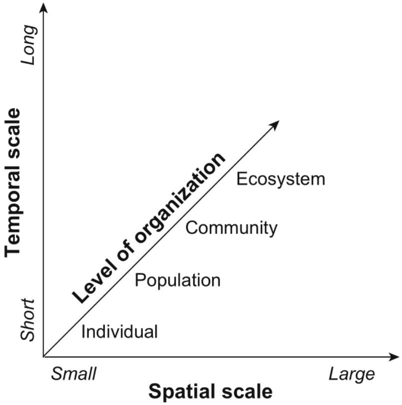
R. Chandler and J. Hepinstall-Cymerman, Landscape.Ecol. (2016) doi : 10.1007/s10980-016-0380-z

CW Case : output



Scale in Ecology

- describes the spatiotemporal dimension of a pattern or process



Spatial Ecology and Conservation Modeling Applications with R, Robert Fletcher, Marie-Josée Fortin (2018)

CW Case : methods

Scale effect

- spatial extent at which each landscape covariate mostly highly correlated with the response variable
- landscape-level covariate covering the entire region while abundance data only available at sampled sites

Goal

- coordinates of a site : $\boldsymbol{x}$
- abundance data at the site : $N(\boldsymbol{x})$
- $E(N(\boldsymbol{x})) = \lambda(\boldsymbol{x}) = z(\boldsymbol{x})$

Spatial smoothing

- unknown : scale at which the surrounding landscape effecting $N(\boldsymbol{x})$
- effects diminish with distance

$$s(z(\boldsymbol{x}), \sigma) = \sum_{(\boldsymbol{x}_j \neq \boldsymbol{x}_i) \in S} z(\boldsymbol{x}_j) w(\boldsymbol{x}_i, \boldsymbol{x}_j, \sigma)$$

CW Case : methods (contd.)

weighting function w

$$w(\boldsymbol{x}_i, \boldsymbol{x}_j, \sigma) = \frac{\exp(-||\boldsymbol{x}_i - \boldsymbol{x}_j||^2/(2\sigma)^2)}{\sum_{(\boldsymbol{x}_j \neq \boldsymbol{x}_i) \in S} \exp(-||\boldsymbol{x}_i - \boldsymbol{x}_j||^2/(2\sigma)^2)}$$

log-linear model

$$log(\lambda(\boldsymbol{x})) = \sum_{i=1}^{p} \beta_i s_i(z_i(\boldsymbol{x}_i), \sigma_i)^{i-1}$$

Poisson model : $N(\boldsymbol{x}_i) \sim Poisson(\lambda(\boldsymbol{x}_i))$

$$\mathcal{L}(\boldsymbol{\beta}, \sigma; \{N(x_i)\}) = \prod_{i=1}^{R} \frac{\lambda(\boldsymbol{x}_i)^{N(\boldsymbol{x}_i)} \exp(-\lambda(\boldsymbol{x}_i))}{N(\boldsymbol{x}_i)!}$$

Probability : Basic Definitions

Sample Space

The set of all possible outcomes in a random experiment is called the sample space, denoted S.

Example : a six-sided dice
$S = \{1, 2, 3, 4, 5, 6\}$

Event

A subset of S ($A \subseteq S$) is called an event.

Example (continued)
$A = \{1, 2\}$
$A = \{3, 4, 5, 6\}$
$A = \{1, 3, 5\}$

Probability : Axioms and Propositions

Axioms

A probability measure is a function P, which assigns to each event A a number P(A) satisfying

(1) $0 \leq P(A) \leq 1$

(2) $P(S) = 1$

(3) let A_i be a sequence of disjoint events

$$P(\bigcup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} P(A_k)$$

Propositions

Let P be a probability measure in some sample space S and let A and B be events

(1) $P(A^c) = 1 - P(A)$

(2) $P(A \setminus B) = P(A) - P(A \cap B)$

(3) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Exercises

Please point out the error in the statement below.

"There will be 50% chance of rain on Sat. and 50% chance of rain on Sun. Therefore, there will be 100% chance of rain this weekend."

Conditional Probability

Definition

Given a non-empty event B, the probability of A given B is defined as

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Example :
Suppose the chance of rain for this Sat. and Sun. is 50% and the chance of two rainy days in a row is 60%. What's the probability of rain this weekend?

Solution :
$P(A \cup B) = P(A) + P(B) - P(A \cap B)$
$P(A \cap B) = P(B|A)P(A) = 0.3$
So, $P(A \cup B) = 0.7$

Independent Events

Definition

If A and B are two events satisfying

$$P(A \cap B) = P(A)P(B)$$

then they are said to be independent.

Example :
Suppose the chance of rain for this Sat. and Sun. is 50% and rainy days are independent. What's the probability of rain this weekend?

Solution :
$\mathrm{P}(\mathrm{A} \cap \mathrm{B}) = \mathrm{P}(\mathrm{A})\mathrm{P}(\mathrm{B}) = 0.25$
So, $\mathrm{P}(\mathrm{A} \cup \mathrm{B}) = 0.75$

Law of Total Probability

Law of Total Probability

Let A_i be a sequence of events that partition the sample space. i.e. $S = \bigcup A_i$. Then, for any event B in the same sample space,

$$P(B) = \sum_{k=1}^{\infty} P(B|A_i)P(A_i)$$

Example

Suppose that two factories supply light bulbs to the market. Factory X's bulbs work for over 5000 hours in 99% of cases, whereas factory Y's bulbs work for over 5000 hours in 95% of cases. It is known that factory X supplies 60% of the total bulbs available and Y supplies 40% of the total bulbs available. What is the chance that a purchased bulb will work for longer than 5000 hours?

$$
\begin{aligned}
P(L > 5000) &= P(L > 5000|F_X)P(F_X) + P(L > 5000|B_Y)P(B_Y) \\
&= 0.99 \times 0.6 + 0.95 \times 0.4 = 97.4\%
\end{aligned}
$$

https://en.wikipedia.org/wiki/Law_of_total_probability