Integrated Step Selection Analysis

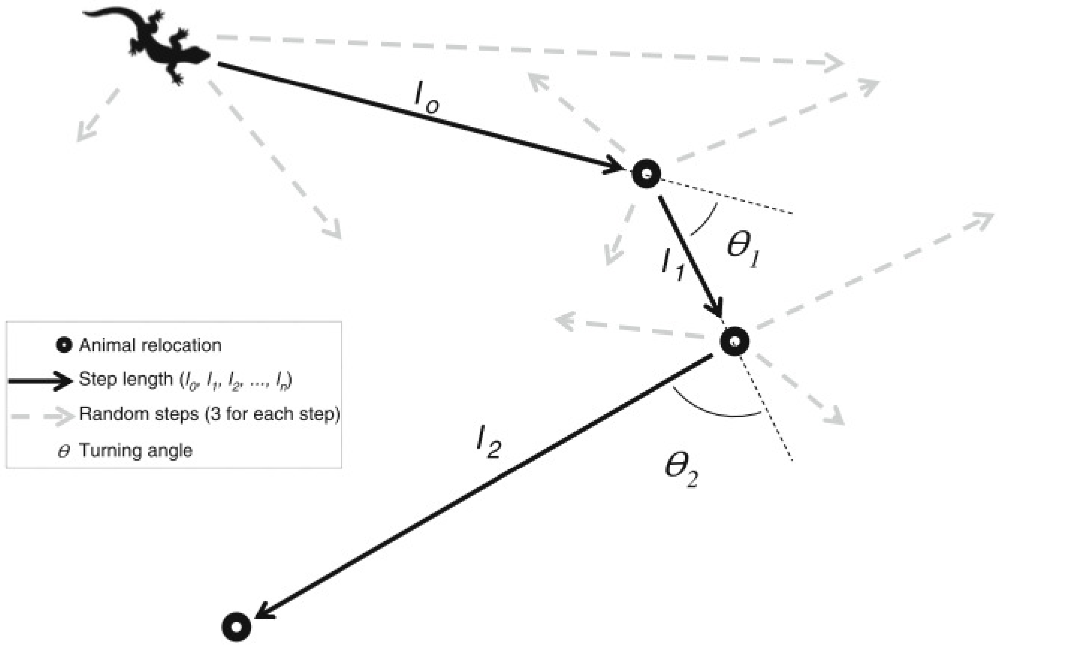


Figure 1 Example of movement pathway in SSFs. Example of how a movement pathway can be simplified into linear step lengths and turning angles occurring between successive locations in any type of animal tracked visually or using VHF or GPS devices. In this example, 3 random steps have been matched with actual steps walked by the lizard.

Swansea University
Prifysgol Abertawe

Luca Börger (@lucaborger)
modified after J. Potts (U. Sheffield, UK)

UNESP Rio Claro, 23 March 2018

College of Science
Coleg Gwyddoniaeth

Resource selection analysis

Step selection analysis (SSA) modifies RSA by relaxing the requirement that data points are independent, making the milder assumption that consecutive data points describe a Markov process. Thus, rather than attempting to model the utilisation distribution u(x), SSA models the probability distribution of the next measurement being at a position x given that the current measurement is at position y. Thus the model it attempts to parametrise has the following form

$$f(\mathbf{x}|\mathbf{y}) = K_{\mathrm{s}}^{-1} \underbrace{\phi(\mathbf{x},\mathbf{y})}_{\text{Availability function}} \underbrace{W(\mathbf{x},\mathbf{y})}_{\text{Resource (weighting) function}}$$

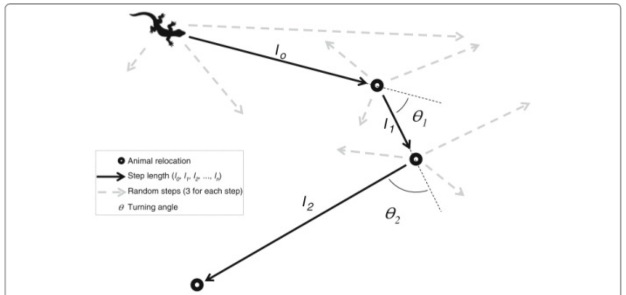


Figure 1 Example of movement pathway in SSFs. Example of how a movement pathway can be simplified into linear step lengths and turning angles occurring between successive locations in any type of animal tracked visually or using VHF or GPS devices. In this example, 3 random steps have been matched with actual steps walked by the lizard.

K_{s} is a normalising constant ensuring that $\int_{\Omega} f(\mathbf{x}|\mathbf{y})d\mathbf{x} = 1$

Resource selection analysis

Notice that W now depends on x and y, rather than just x. By extension, it could also depend upon information between x and y (such as roads and rivers).

$$f(\mathbf{x}|\mathbf{y}) = K_{\mathrm{s}}^{-1}\phi(\mathbf{x},\mathbf{y})W(\mathbf{x},\mathbf{y})$$

SSA has an interesting history. The term 'Step Selection Analysis' was first coined by Fortin et al. (2005). However, the key concept of examining resource selection between consecutive locations was also present in a paper by Rhodes et al. (2005) (same year and journal) and can be traced back to Arthur et al. (1996). Although the take-up of SSA was rapid and wide-spread amongst empirical biologists (Thurfjell et al., 2014), it took a little while for various aspects of the theory to be worked out. These include differences in what SSA and RSA estimate (Barnett and Moorcroft, 2008), eliminating bias from conflation between the `availability' kernel and 'weighting' function W (Forester et al., 2009), as well as links to mechanistic space-use models (Moorcroft and Barnett, 2008; Potts et al., 2014a), collective motion models (Delgado et al., 2014; Potts et al., 2014b), and biased/correlated random walks (Duchesne et al., 2015).

RSF to SSF

Use-availability likelihood:

$$g_u = \frac{\exp(x\beta)g_a}{\int_{s\in A}\exp(x\beta)g_a}$$

How should we model availability with GPS data collected on a fine temporal scale?

Circular buffers

Use-availability likelihood:

$$g_u = \frac{\exp(x\beta)g_a}{\int_{s\in A}\exp(x\beta)g_a}$$

Arthur et al. (1996) defined g_a using circular buffers, with radius r around used locations.

Determined r by considering maximum distance moved between points.

Does this make sense?

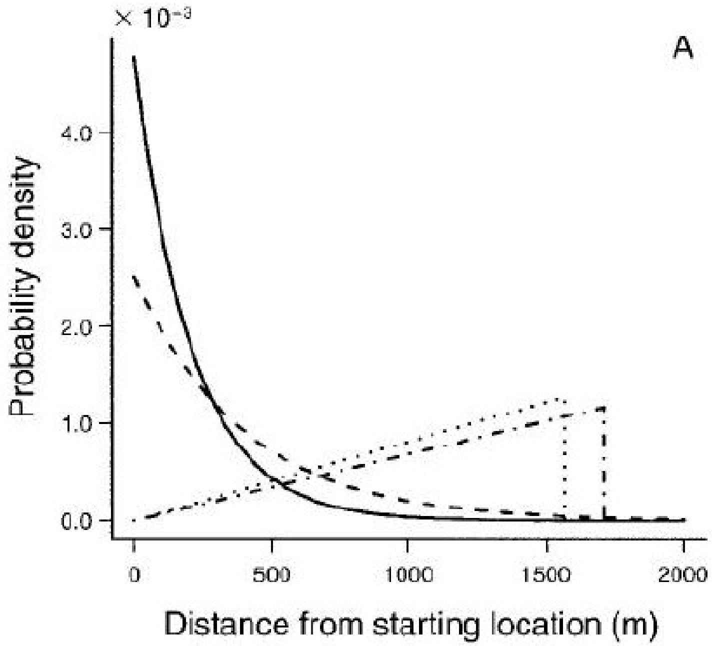
Movement Distributions

Arthur et al. (1996):

$$g_a = \frac{1}{2\pi r^2} \text{ for } d \leq r \text{ (0 otherwise)}$$

Rhodes et al. (2005):

$$g_a = \lambda \exp(-\lambda d)/(2\pi d)$$



Rhodes, J. R., McAlpine, C. A., Lunney, D. & Possingham, H. P. 2005 A spatially explicit habitat selection model incorporating home range behavior. Ecology 86, 1199-1205.

Movement to define availability

Rhodes et al. (2005): ‘think in terms of movement’

$$g_u = \frac{\exp(x\beta) g_a}{\int_{s \in A} \exp(x\beta) g_a}$$

$g_a = g_a(d, \theta)$ = resource-independent movement kernel/model that describes how an animal would move in homogeneous habitat.

- d distance between (x_{t-1}, y_{t-1}) and (x_t, y_t)
- θ are unknown movement parameters

Step Selection Functions

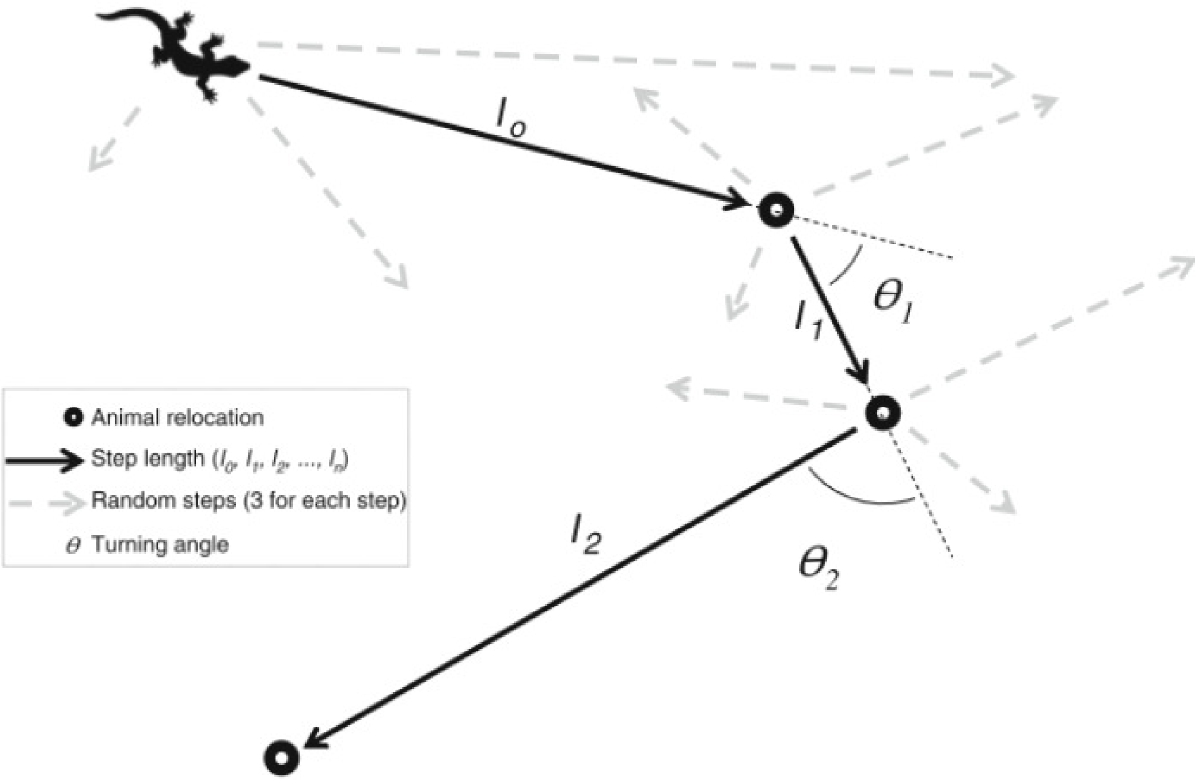


Figure 1 Example of movement pathway in SSFs. Example of how a movement pathway can be simplified into linear step lengths and turning angles occurring between successive locations in any type of animal tracked visually or using VHF or GPS devices. In this example, 3 random steps have been matched with actual steps walked by the lizard.

Using movement to define g_a

If we knew g_a, we could take a random sample of n_a points from this distribution and evaluate:

$$\prod_{i=1}^{n} \prod_{t=1}^{T} \frac{exp(x_{it}^{u}\beta)}{exp(x_{it}^{u}\beta) + \sum_{j=1}^{n_a} exp(x_{jt}^{a}\beta)}$$

- x_{it}^{u} = the t^{th} used point taken on the i^{th} animal.
- x_{jt}^{a} is the j^{th} available point associated with the t^{th} used point taken on the i^{th} animal.

This is the same as the likelihood of a conditional logistic regression model

We can fit using `clogit` function in R `survival` library:

```
clogit(y~ x+ strata(strataID), data=)
```

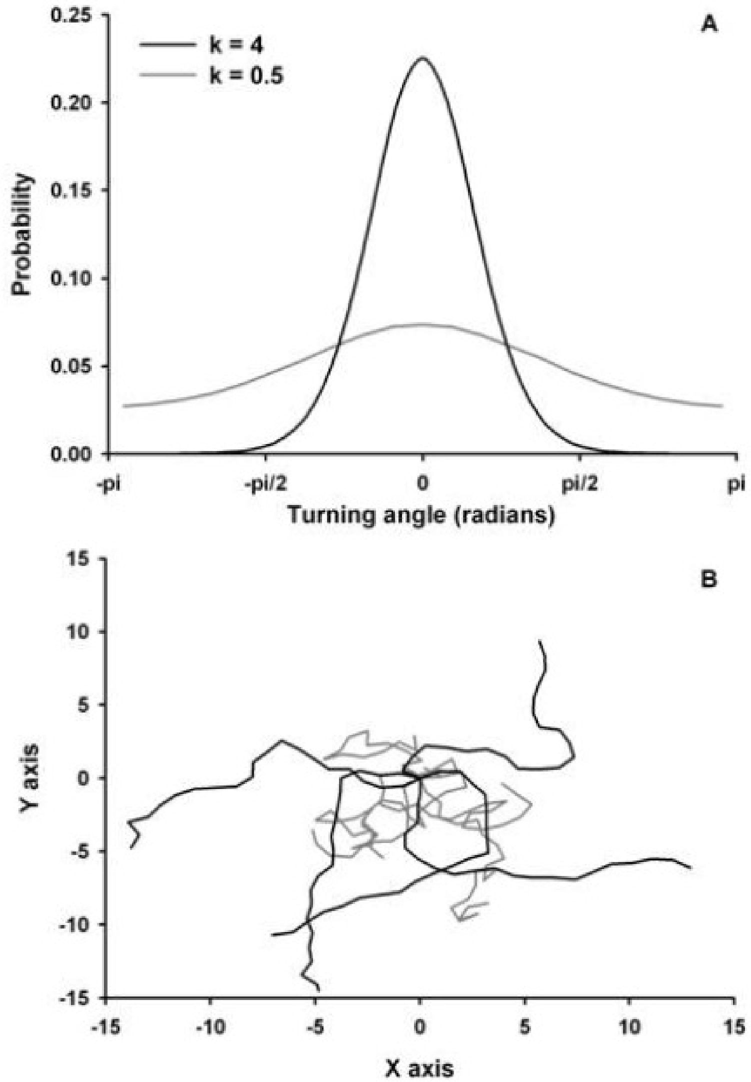
PROBLEM

We don't know g_a!

Options:

- Resample observed step lengths and turn angles
- Use step-lengths and turn angles to parameterize statistical distributions
  - Step lengths: gamma distribution
  - Turn angles: von Mises distribution (or uniform on $-\pi$ to π)

von Mises distribution



Estimating movement parameters

$$g_u = \frac{\exp(x\beta)g_a}{\int_{s\in A}\exp(x\beta)g_a}$$

Problem:

- g_a is meant to reflect movement in the absence of resource selection.
- We observe movements that reflect the combination of g_a AND $w(x, \beta)$

Can give biased estimates of movement parameters

Forester, J.D., Im, H.K. & Rathouz, P.J. (2009). Accounting for animal movement in estimation of resource selection functions: Sampling and data analysis. Ecology, 90, 3554–3565.

Estimating movement parameters

Solution:

- Include step length, ln(step length), and cos(turn angles) to modify (re-estimate) movement parameters

Opportunity:

- Can include interactions with predictors to model how movement is influenced by habitat

Avgar, T., Potts, J.R., Lewis, M.A. & Boyce, M.S. (2016). Integrated step selection analysis: Bridging the gap between resource selection and animal movement. Methods Ecol. Evol., 7, 619–630.

Integrated Step Selection Analysis (Avgar, Potts et al. (2016) Methods Ecology Evolution)

Presents an SSA framework that is both as general as possible and theoretically well-grounded. The resulting formalism is termed ‘integrated step selection analysis (iSSA).

BTW, Jonathan considers it his best paper – hence certainly worth reading it!

An advantage of iSSA is that it gives the user an explicit way of dealing with the nefarious concept of `availability'. It does this by a two-step fitting procedure. First, we define the function Φ(x|y) to be a very general type of distribution called the gamma distribution, which can be fitted to the data to give an initial estimate of availability. This has the following form:

$$\phi(\mathbf{x}|\mathbf{y}) = \frac{l^{b_1-1}\exp(-l/b_2)}{\Gamma(b_1){b_2}^{b_1}}$$

where:

$$l = |\mathbf{x} - \mathbf{y}|$$

is the distance between x and y. The symbols b1 and b2 are typically referred to as the shape and scale of the distribution, respectively.

Integrated Step Selection Analysis
(Avgar, Potts et al. (2016) Methods Ecology Evolution)

Then we incorporate covariates into the weighting function, W, to correct this choice of Φ. This procedure allows us to select our model of availability from a wide range of distributions, including exponential, half-normal, stretched-exponential, and log-normal, letting the data decide which distribution fits best. For example, suppose we use conditional logistic regression to fit to data the following form for W(x; y), sampling from the Gamma distribution (see previous slide) for our control values:

$$W(\mathbf{x},\mathbf{y}) = \exp[b_3 l^{b_4} + b_5 \ln(l) + b_6 l + \beta_0 + \beta_1 Z_1(\mathbf{x},\mathbf{y}) + \cdots + \beta_n Z_n(\mathbf{x},\mathbf{y})]$$

where Z1; …;Zn are environmental covariates dependent upon both x and y (as previously explained). Then, by the theory explained, we have fitted the following movement kernel to the data:

$$f(\mathbf{x}|\mathbf{y}) = K_{\mathrm{s}}^{-1} \underbrace{\exp[(b_1 - 1)\ln(l) - l/b_2 + b_3 l^{b_4} + b_5 \ln(l) + b_6 l]}_{\text{Resource-independent movement}} \times$$

$$\underbrace{\exp[\beta_0 + \beta_1 Z_1(\mathbf{x},\mathbf{y}) + \cdots + \beta_n Z_n(\mathbf{x},\mathbf{y})]}_{\text{(Step) selection function}}.$$

Integrated Step Selection Analysis (Avgar, Potts et al. (2016) Methods Ecology Evolution)

For those keen among you it is an instructive exercise to derive the previous equation and show how the resulting function for resource-independent movement contains exponential, half-normal, stretched-exponential, and log-normal distributions as special cases.

The iSSA procedure also allows us to vary our definition of availability across time and space, if necessary. For example, in times and areas of heavy snow, the ability of a roaming ungulate to move may be more limited than when snow is minimal or absent. In this case, we might exchange b6 (which denotes an exponential correction to the step-length distribution) for b6 - S(x, y)b7, where S(x, y) is the mean snow depth between x and y. This gives the following weighting function:

$$W(\mathbf{x}, \mathbf{y}) = \exp[b_3 l^{b_4} + b_5 \ln(l) + b_6 - S(\mathbf{x}, \mathbf{y}) b_7 + \beta_0 + \beta_1 Z_1(\mathbf{x}, \mathbf{y}) + \cdots + \beta_n Z_n(\mathbf{x}, \mathbf{y})]$$

Note. If you are concerned about the large amount of parameters to t, but are confident that the step length distribution is well-fitted by an exponential distribution, you can set b1 = 1 and b3 = b4 = b5 = 0. As well as being conceptually simpler, this can ease over-fitting. However, given the sizes of many modern datasets, an extra four parameters is hardly something to get overly-worried about.

Integrated Step Selection Analysis

Another feature of iSSA is that it can be used to incorporate bias and correlation in movement. For example, if you hypothesise that an animal at position y will have a bias towards a fixed point x_f then set Zi(x y) = cos(θ - θ_f) for some i, where θ is the bearing from y to x and θ_f is the bearing from y to x_f. Likewise, testing for correlation involves defining Zj = Zj(x y; θ_0) = cos(θ - θ_0) for some j, where θ_0 is the bearing on which the animal arrived at point y. Notice that, in this case, Zj depends not only on x and y but also on θ_0. Therefore the function f from Equation (16) depends on 0 and so should be written f(x|y; θ_0) and not f(x|y).

As well as giving a method for testing hypotheses about animal movement, iSSA can be used to parametrises a model of animal movement, given by the SLF (see two previous slides). Therefore you can simulate this model forwards in time to infer space use patterns, either using stochastic simulations (Potts et al., 2014b; Avgar et al., 2016) or by constructing the following integro-difference equation for space-use [sometimes called the Master Equation, see van Kampen (1981)]

$$u(\mathbf{x}, t+\tau) = \int_\Omega f(\mathbf{x}|\mathbf{y})u(\mathbf{y}, t)\mathrm{d}\mathbf{y}$$

where t is time and τ is the time-period between consecutive steps. This equation links iSSA back to the original RSA question linking resources and space use (see first equation in the RSA presentation).

Integrated Step Selection Analysis

$$u(\mathbf{x}, t+\tau) = \int_{\Omega} f(\mathbf{x}|\mathbf{y})u(\mathbf{y}, t)\mathrm{d}\mathbf{y}$$

Such techniques have been used to analyse the processes required for home range emergence in bison (Merkle et al., 2017b) as well as predicting spread of the disease brucellosis in elk (Merkle et al., 2017a). Notice that the equation above cannot be used directly in the case where f is dependent upon θ_0 (i.e. when there is correlation in the direction between consecutive steps). However, this technical difficulty can be overcome with a simple trick, which we will not go into here, but is explained in Potts et al. (2014a).

iSSA example analysis in R

We start with simulated data. Despite the underlying theory being a little formidable at first sight, you will hopefully observe that the practical use of iSSA is actually rather simple.

We use the same simulated resource landscape as in Section 1.2. This time, the data is generated by simulating a path starting at the central point of the landscape. Each point, xj for j = 2; … ; n, is simulated by drawing from the following distribution, which assumes knowledge of the previous location x_{j-1}:

$$f(\mathbf{x}_j|\mathbf{x}_{j-1}) = \frac{\exp(-\lambda|\mathbf{x}_j - \mathbf{x}_{j-1}|)\exp(\beta r(\mathbf{x}_j))}{\int_\Omega \exp(-\lambda|\mathbf{x} - \mathbf{x}_{j-1}|)\exp(\beta r(\mathbf{x}_j))\mathrm{d}\mathbf{x}}$$

This is a very simple example of the equation where the weighting function $W(x_j ; x_{j-1})$ = exp(βr(x)), is only dependent upon the end of the step (rather than any information along the step), and the resource independent movement, $\exp(-\lambda|x_j – x_{j-1})$ is proportional to an exponential distribution [the gamma distribution (Equation 13) with shape parameter equal to 1].

iSSA example analysis in R

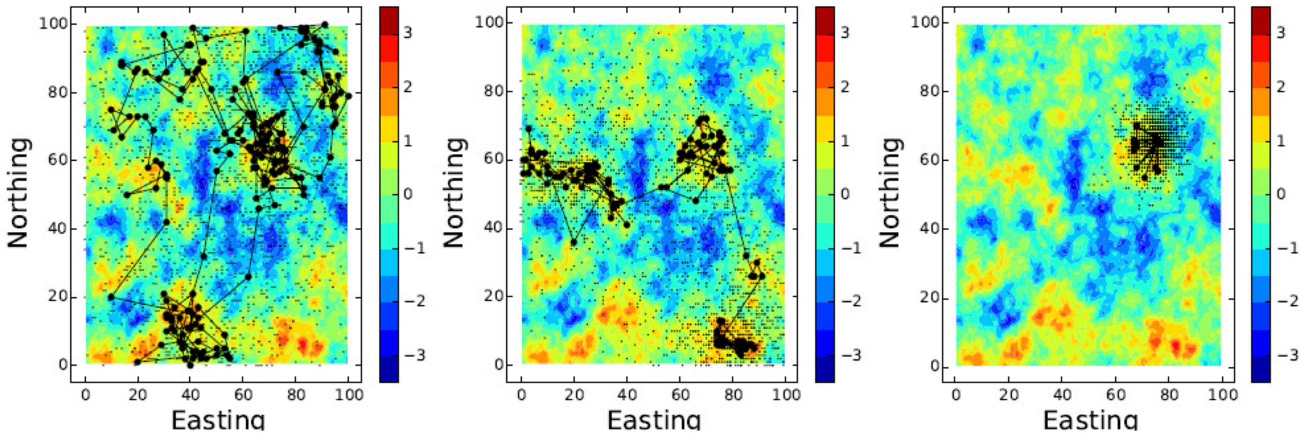


Fig. 3. Simulated data for step selection analysis. The colour contours for each panel show resource quality for a simulated landscape, giving a function $r(\mathbf{x})$ of the resource quality at each point, $\mathbf{x}$. The large dots in each panel show simulated animal positions, obtained by simulating a path whereby each location depends on the previous location, and is sampled from the distribution given by Equation (19). In the left panel, $\beta = 1$; in the central panel, $\beta = 3$; in the right panel, $\beta = 5$. For all panels, $\lambda = 0.2$. The small dots on each panel give the control positions (see Main Text for details).

Fig. 3 shows plots of three paths, given by three different betas. As in Section 1.2, we use the data we have on each of these paths to infer the values of beta. We assume we only know about the n = 200 datapoints for each panel, and not the actual beta -values (i.e. we take the point-of-view of the scientist rather than God).

iSSA example analysis in R

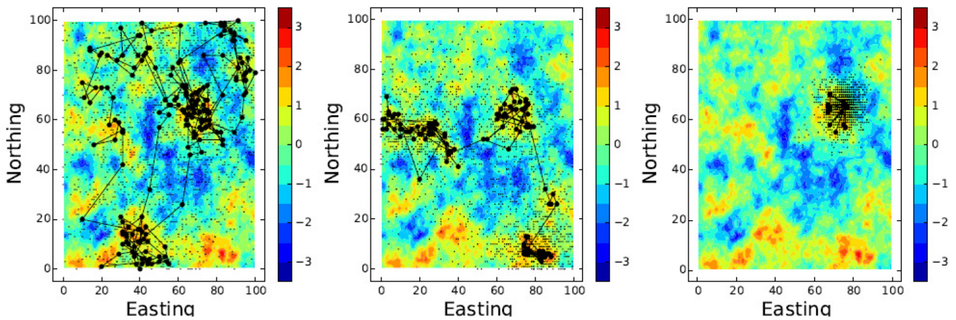


Fig. 3. Simulated data for step selection analysis. The colour contours for each panel show resource quality for a simulated landscape, giving a function $r(\mathbf{x})$ of the resource quality at each point, $\mathbf{x}$. The large dots in each panel show simulated animal positions, obtained by simulating a path whereby each location depends on the previous location, and is sampled from the distribution given by Equation (19). In the left panel, $\beta = 1$; in the central panel, $\beta = 3$; in the right panel, $\beta = 5$. For all panels, $\lambda = 0.2$. The small dots on each panel give the control positions (see Main Text for details).

Notice that we also do not have direct information on the value of lambda. The rubric of iSSA suggests that we make an initial estimate of this, by fitting the empirical step-lengths to an exponential distribution with parameter b (this corresponds to 1/b2 in this equation:

$$\phi(\mathbf{x}|\mathbf{y}) = \frac{l^{b_1-1}\exp(-l/b_2)}{\Gamma(b_1){b_2}^{b_1}}$$

For an exponential distribution, fitting is easy, since the value of b that maximises the likelihood is the reciprocal of the empirical mean step-length, <l>. In other words, b = 1/<l>.

iSSA example analysis in R

In the case where beta = 1, we have b ~ 0.0569. Notice that the parameter for the empirical step length distribution (b ~ 0.0569) is quite different from the resource independent movement kernel (lambda = 0.2). For each data-point, xj-1, for j ε {2,…,n}, we sample 10 locations from the empirical step-length distribution:

$$f_{\mathrm{e}}(\mathbf{x}|\mathbf{x}_{j-1}) = b^{-1}\exp(-b|\mathbf{x}_j - \mathbf{x}_{j-1}|)$$

to give 10 Control values to compare with each of the Cases x2, … , xn. Notice that the first location, x1, is not a Case for the purposes of iSSA. (→ d you understand why?)

We are now in a position to perform logistic regression to infer the real values of beta and lambda. The file sim_ssf_rf1_beta1_1.csv contains all the information we need. The first five columns are identical to the RSF example (sim_rsf_rf1_beta1_1.csv). The final column contains the step length for the step ending at the location given by columns 3 and 4.

iSSA example analysis in R

```
# read in the file
  ssfex1 = read.csv("sim_ssf_rf1_beta1_1.csv")

# Then the command for logistic regression is:

  beta1sim = clogit(ssfex1$Observed ~ ssfex1$resource + ssfex1$StepLength +
             strata(ssfex1$strata), data = ssfex1)

# This is very similar to the analogous command used in Section 1.2. However, this
# time, we have also incorporated the step length as a covariate. As in Equation (15),
# the strength of this covariate is b6. For simplicity, we set b3 = b5 = 0.
# To see the output of the logistic regression, we type beta1sim and observe the
# following (see next slide)
```

iSSA example analysis in R

```
Call:
clogit(ssfex1$Observed ~ ssfex1$resource + ssfex1$StepLength +,
strata(ssfex1$strata), data = ssfex1)

                      coef      exp(coef)  se(coef)  z       p
 ssfex1$resource      0.9371    2.5526     0.1215    7.71    1.2e-14
 ssfex1$StepLength   -0.1428    0.8669     0.0135    -10.58  <2e-16

Likelihood ratio test=397 on 2 df, p=0
n= 2189, number of events= 199
```

As in Section 1.2, we can read off the estimated β-value as $\beta = 0.9371 \pm 0.1215$, which is a decent estimate of the real value of $\beta = 1$. Comparing Equations (16) and (19), we see that the estimate for λ is $b - b_6$ (recall that $b_3 = b_5 = 0$, $b = 1/b_2$, and $l = |\mathbf{x}_n - \mathbf{x}_{n-1}|$). The value $b_6 = -0.1428$ is given in the second row of the table under the `coef` heading. Therefore our estimate for λ, using conditional logistic regression, is $\lambda = b - b_6 \approx 0.0569 + 0.1428 \pm 0.0135 = 0.1997 \pm 0.0135$. As you can see, this is a pretty good agreement with the true value of $\lambda = 0.2$.

$$f(\mathbf{x}|\mathbf{y}) = K_s^{-1} \underbrace{\exp[(b_1 - 1)\ln(l) - l/b_2 + b_3 l^{b_4} + b_5 \ln(l) + b_6 l]}_{\text{Resource-independent movement}} \times \underbrace{\exp[\beta_0 + \beta_1 Z_1(\mathbf{x},\mathbf{y}) + \cdots + \beta_n Z_n(\mathbf{x},\mathbf{y})]}_{\text{(Step) selection function}}. \qquad (16)$$

$$f(\mathbf{x}_j|\mathbf{x}_{j-1}) = \frac{\exp(-\lambda|\mathbf{x}_j - \mathbf{x}_{j-1}|)\exp(\beta r(\mathbf{x}_j))}{\int_\Omega \exp(-\lambda|\mathbf{x} - \mathbf{x}_{j-1}|)\exp(\beta r(\mathbf{x}_j))d\mathbf{x}}. \qquad (19)$$

iSSA example analysis in R

Now repeat the analysis for the beta = 3 and beta = 5 cases. The relevant simulated data is found in the files sim_ssf_rf1_beta3_1.csv and sim_ssf_rf1_beta5_1.csv respectively.

Further analysis of simulated data

As mentioned in the lecture, we can use iSSA to test for correlation between bearings of successive steps. We'll do that here. First, we need to add columns to the input data from sim_ssf_rf1_beta3_1.csv, denoting the bearing of the animal and the cosine of the turning angle. I've done this for you in sim_ssf_rf1_beta1_bearings.csv. Notice that bearings are in radians.

It's a good exercise to construct this le yourself, either using Excel or R. But it's not a high-priority task, so probably it's best to do it later. Here's some tips for this task:
You may find the atan2() command helpful (available in both Excel and R)
If the animal does not move from one x to the next then the notion of " bearing" does not make sense. You will have to delete these rows. You will also have to delete the " control“ rows that correspond to the deleted cases.
For case n, the cosine of the turning angle is $\cos(\theta_n - \theta_{n-1})$ where θ_n is the bearing in case n.

```
ssfbear<-read.csv("sim ssf rf1 beta1 bearings.csv ")

# perform logistic regression incorporating a test of correlation from one bearing to
# the next using the following command:

ssfbear1<-clogit(Observed ~ resource + StepLength +CosTA + strata(strata),
             data = ssfbear)

# This command fits the following model for the movement kernel:
```

$$f(\mathbf{x}_j|\mathbf{x}_{j-1}, \mathbf{x}_{j-2}) \propto \exp(-\lambda|\mathbf{x}_j - \mathbf{x}_{j-1}|) \exp(\beta_1 r(\mathbf{x}_j) + \beta_2 \cos(\theta_{j-1})),$$

where θ_{j-1} is the turning angle at $\mathbf{x}_{j-1}$.

```
# The file sim_ssf_rf1_beta1_corr.csv contains positional data from a simulation
# where there is both correlation between successive bearings and a tendency to
# choose areas where the resources are better. The movement kernel for this
# simulation is given by Equation (1). The control values were sampled from the
# following movement kernel:
```

$$f_\mathrm{e}(\mathbf{x}|\mathbf{x}_{j-1}) = b^{-1}\exp(-b|\mathbf{x}_j - \mathbf{x}_{j-1}|)$$

where b = 0.0628. Use iSSA to infer the values of lambda, beta1, and beta2. What are the underlying resource-independent step-length and turning-angle distributions?

Analysis of real data

We will use the Elk data found in the file elk_positions.csv. However, we need to construct different control values. For each case (i.e. location x), sample 10 step lengths from the exponential distribution using the equation below, with b = 1/<l> and <l> is the average step length.

$$f_{\mathrm{e}}(\mathbf{x}|\mathbf{x}_{j-1}) = b^{-1}\mathrm{exp}(-b|\mathbf{x}_j - \mathbf{x}_{j-1}|)$$

Show that b = 240 (you can use Excel or R for this).
Sampling 32310 draws from the exponential distribution (corresponding to 3231 steps) is done in R as follows:

```
sl <- rexp(32310, rate=240)
```

Using this, compute the sampled positions, construct a file suitable for importing into clogit() and test for the effect of altitude on step-to-step movement. The tools for this task are all modifications of what you have already done in this worksheet and the previous one.

If you've completed everything above, try to use your understanding to test for correlations between successive bearings of the elk. This will take some thought and time, but all the ingredients are just small modifications of what you have already done.

The END!