An underwater photograph of a coral reef. In the foreground, there are various types of coral, including some that are yellowish and others that are green. Several small fish are swimming in the water. The background is a deep blue, suggesting the open ocean. A semi-transparent dark blue rectangle is overlaid on the center of the image, containing the title text in white.

High Performance Computational Methods for Ecology and Evolution

Lauren Attfield

Aims

- Understand what deterministic and stochastic modelling are
- Be able to build stochastic models based on given parameters
- Appreciate the added complexity of drawing from random distributions rather than applying rates and averages
- Be able to run stochastic simulations on a cluster
- Be able to process and visualise stochastic simulations
- Appreciate where deterministic and stochastic modelling substantially differ

Population modelling

1. Population modelling overview
and stochastic vs deterministic
2. Demographic population models
and matrix population models
3. Stochastic adaptation

Population modelling

1. Population modelling overview and stochastic vs deterministic

2. Demographic population models
and matrix population models

3. Stochastic adaptation

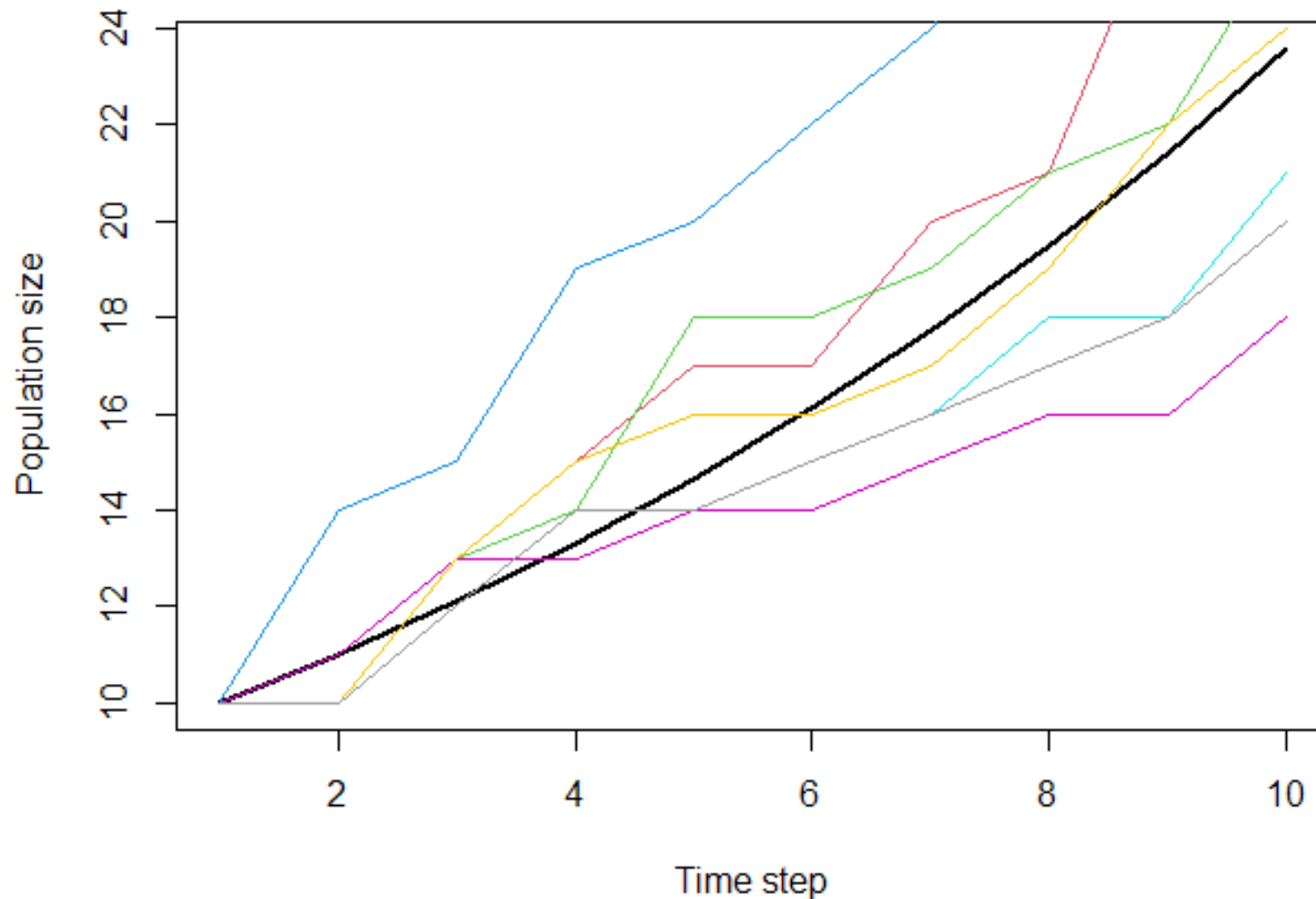
What is a population model?

- Models how a population (or populations) changes over time
- We might be interested in...
 - The effect of different parameters on the population dynamics
 - Identifying an appropriate control or conservation strategy
 - The likelihood of extinction or recovery
- We can make different assumptions:
 - Homogeneous – every individual is the same
 - Heterogeneous – individuals can be different
 - Compartmental – there are different compartments and every individual in the same compartment is the same
 - Individual-based – we explicitly model every single individual and all their traits

Stochastic vs deterministic modelling

- Deterministic = determined
 - Assumes that outcomes follow determined paths
 - Can be simulated once to observe modelled behaviour
- Stochastic = random
 - Assumes that outcomes are drawn from random distributions
 - Must be simulated many times to observe modelled behaviour, since every simulation can be different
 - Typically closer to reality as life is inherently random!

Example: Exponential population model



Deterministic:
(black line)

Every time step, the bacterial population multiplies by 1.1

$$p_{t+1} = 1.1p_t$$

Stochastic:
(coloured lines)

Every time step, there is a 10% chance of each individual bacterium reproducing

$$p_{t+1} = p_t + N_{t+1}$$

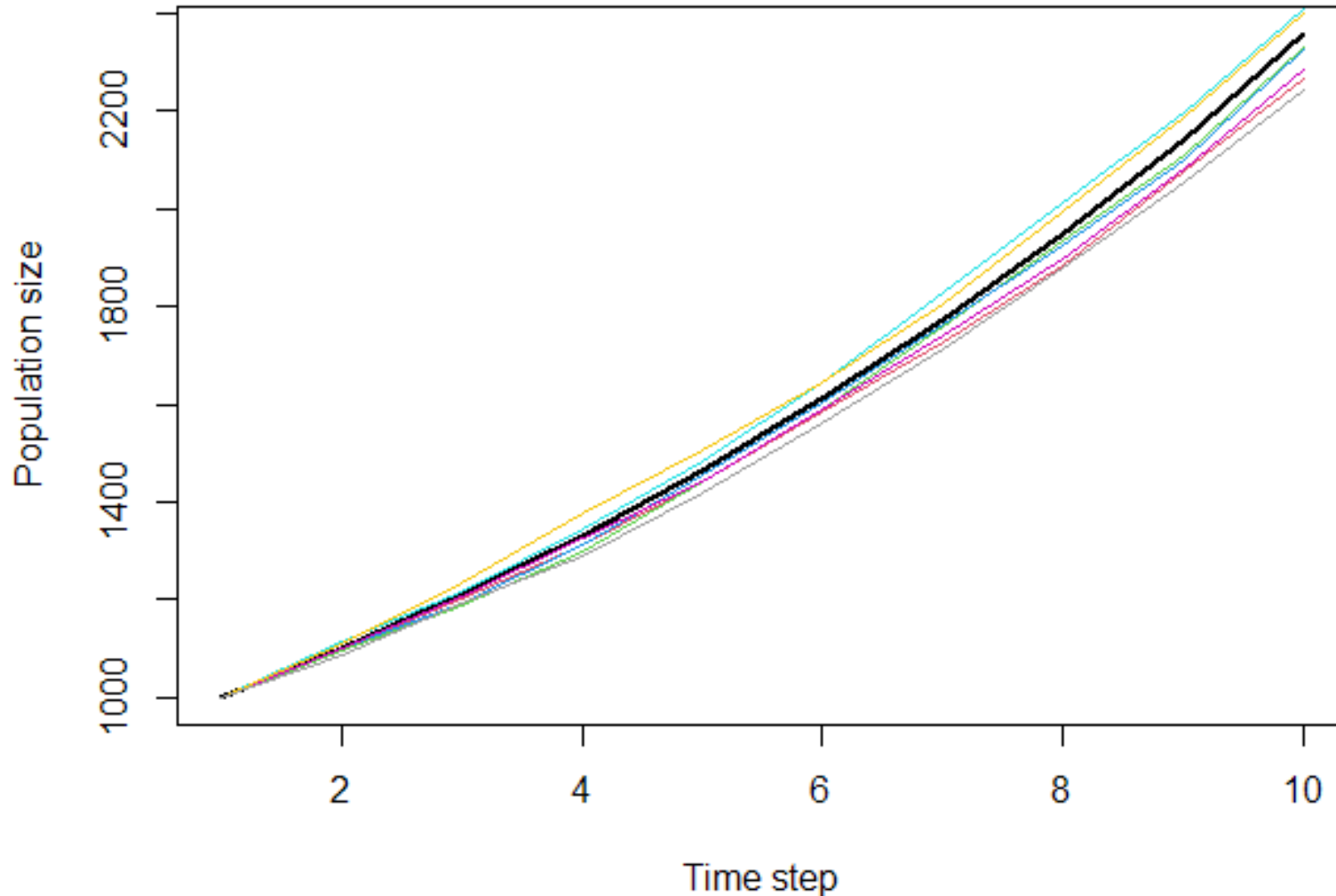
$$N_{t+1} \sim \text{Bin}(p_t, 0.1)$$

Example: Exponential population model

model_type	simulation_number	time_0	time_1	time_2	time_3	time_4	time_5	time_6	time_7	time_8	time_9
deterministic	<NA>	10	11	12.1	13.31	14.64	16.11	17.72	19.49	21.44	23.58
stochastic	1	10	12	13	14	17	19	20	22	24	26
stochastic	2	10	11	11	14	15	15	16	20	24	29
stochastic	3	10	11	12	15	16	19	20	22	25	27
stochastic	4	10	11	13	14	16	19	20	22	24	26
stochastic	5	10	11	13	13	15	17	18	19	20	22
stochastic	6	10	11	12	12	12	12	12	12	12	14
stochastic	7	10	10	11	12	13	14	16	18	18	19

- Note that the deterministic simulation is based on averages and therefore results in continuous – not discrete – outputs
- If we were modelling number of bacteria, this would be biologically unrealistic
- In this particular situation, the deterministic simulations reflects the average behaviour of the stochastic model

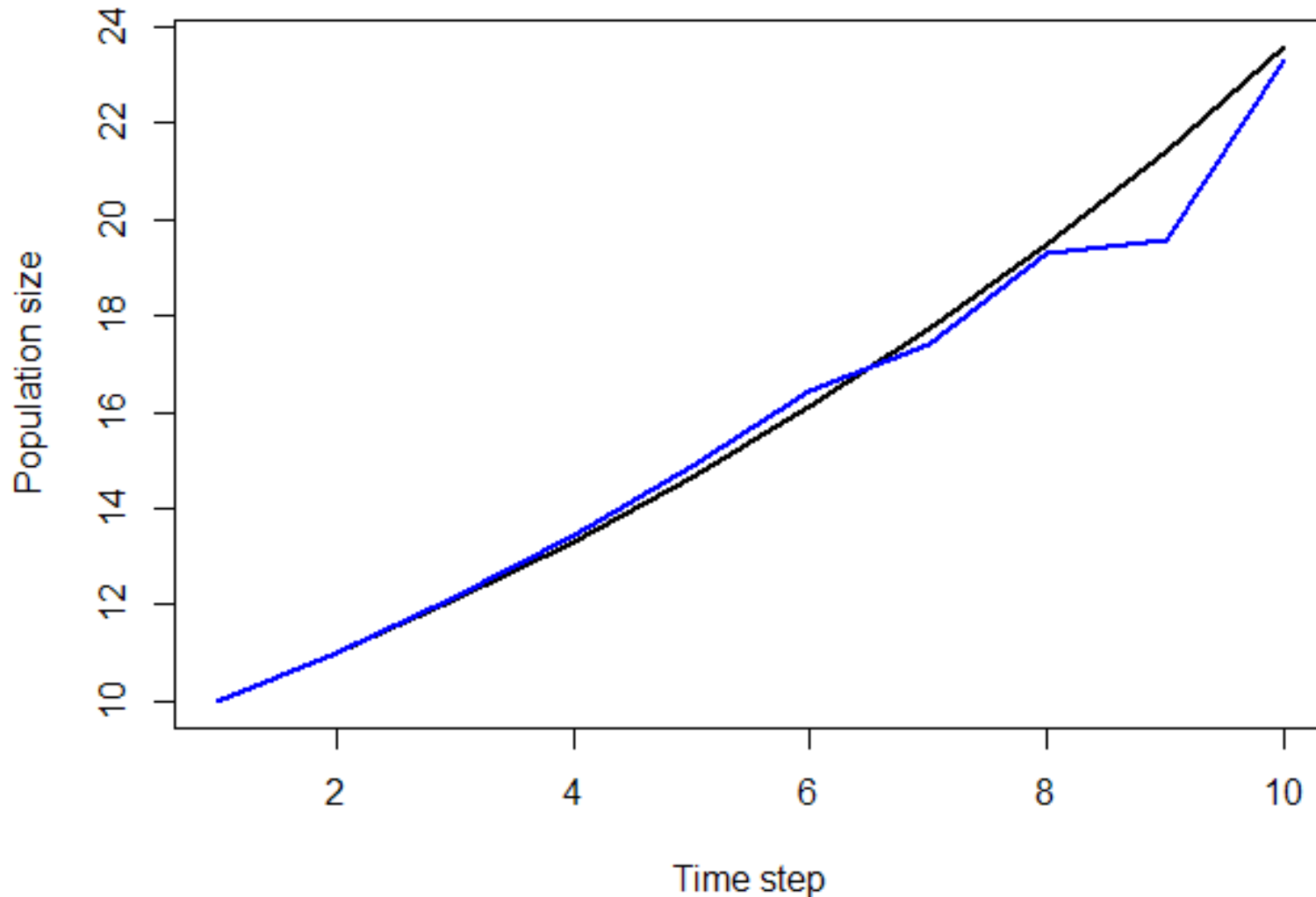
Example: Exponential population model



In some situations, deterministic and stochastic modelling may produce very similar outcomes – the deterministic model is a **good approximation** to the stochastic model.

E.g., in this example the initial population size was set at 1000 rather than the previous value of 10.

Example: Exponential population model



In some situations, the deterministic model reflects the **average behaviour** of the stochastic model.

E.g., the blue line in this plot shows the average population size at each time point across the 7 stochastic simulations (with initial population size of 10) (the black line is still the deterministic simulation)

Stochastic modelling

- Enables us to include realistic probability distributions (rather than assuming that everything follows the “average” outcome)
- Can enable us to handle discrete outcomes
 - E.g., whole number of organisms; not 16.11 bacteria
- Most important in situations where:
 - Stochastic events can make a substantial difference
 - Future changes are sensitive to the current value
- Should be run many times in order to get an overall sense of model behaviour – this is where cluster computing can be useful
- Can present a trade-off: biological realism vs computational effort

Population modelling

1. Population modelling overview
and stochastic vs deterministic

**2. Demographic population models
and matrix population models**

3. Stochastic adaptation

Demographic population models

- A demographic population model describes how a population changes in terms of, and in response to, a demographic structure in the population.
- The demographic structure would typically be a set of traits which affect life history processes (i.e., birth and death).
- A simple demographic trait to structure a population model is age
- Other traits could include body weight, sex, or sexual maturity
- Demographic population models can also incorporate evolutionary processes (“evolutionary demography”) which have been used to explore evolution of life history “strategies”

Matrix population models

- The term **matrix population model** usually refers to an **age-** or **stage-structured matrix population model**
- In this format, a population is divided into **discrete life stages** or **age classes**
- Each **stage/class** has a **set of parameters** to describe **demographic processes** (i.e., survival, maturation, and recruitment)
- The population is represented by a **state vector** containing the number of individuals in each stage/class
- The parameters are put into a **projection matrix** which can be applied to the state vector to project the population forwards in time
- Matrix population models are typically **deterministic**

Matrix population models

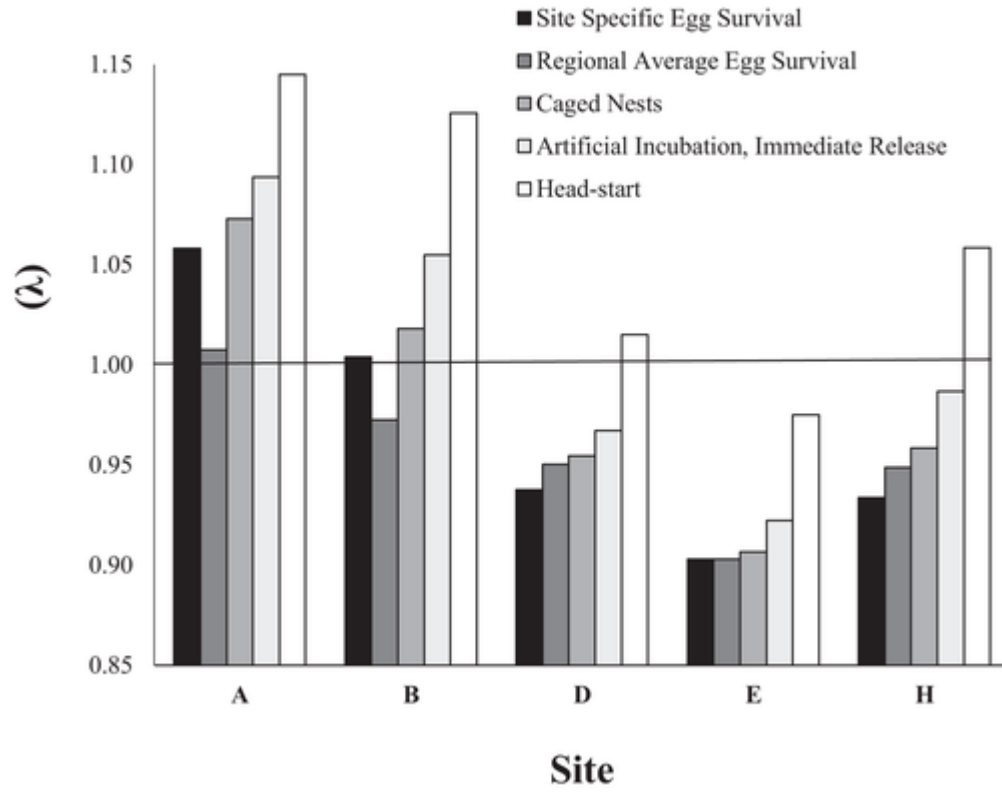
- Allows us to model population dynamics, which are hard to observe, based on individual-level processes, which are easier to observe
- The parameters can be inferred by field studies such as **capture-mark-recapture**
- Online database of matrices, so there is field-based evidence to support the modelling of many different populations



- Version 6.22.5.0
- 760 Taxonomic Species
- 643 Publications
- 8851 Matrix Population Models

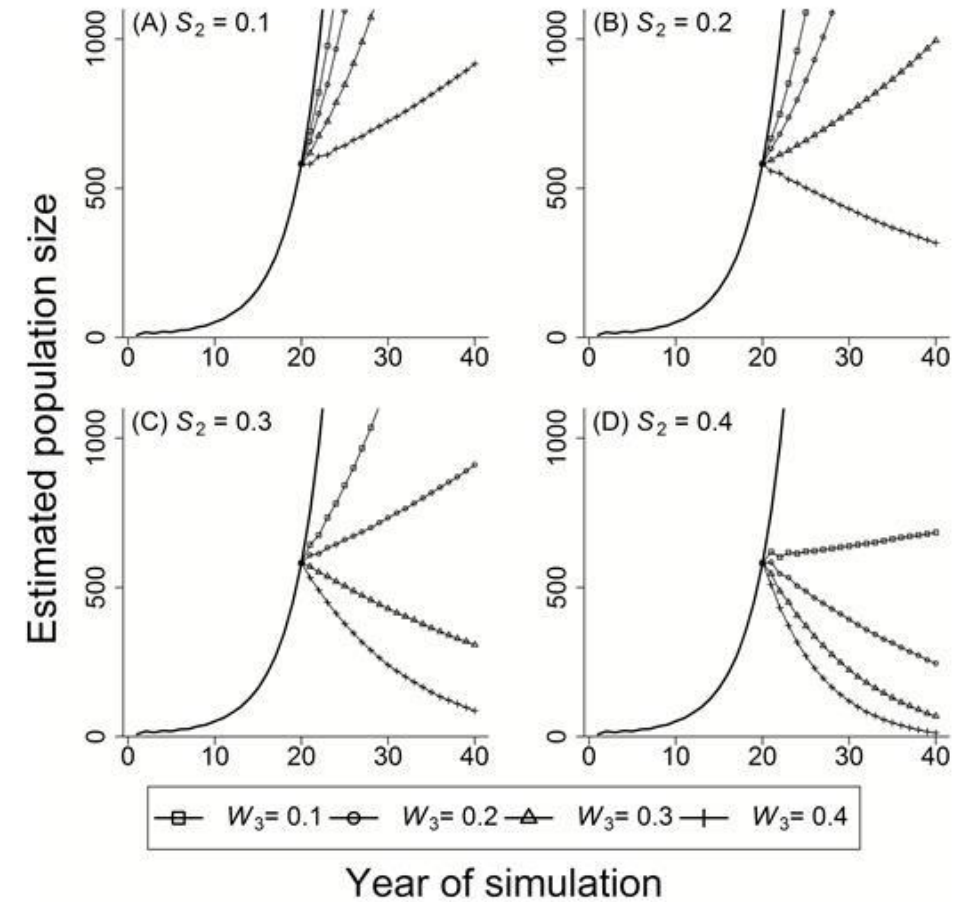


- Version 4.21.8.0
- 415 Taxonomic Species
- 395 Publications
- 3317 Matrix Population Models



Bog turtle conservation in North Carolina, USA.

Knoerr, M.D., Tutterow, A.M., Graeter, G.J., Pittman, S.E. and Barrett, K. (2022), Population models reveal the importance of early life-stages for population stability of an imperiled turtle species. *Anim. Conserv.*, 25: 53-64. <https://doi.org/10.1111/acv.12718>



Control of the invasive Chinese mitten crab in North America and Europe.

Zhang, Z., Yokota, M., Strüssmann, C.A. (2019), A periodic matrix population model to predict growth potential of the invasive Chinese mitten crab *Eriocheir sinensis* (H. Milne Edwards, 1853) (Decapoda: Brachyura: Varunidae), *J. Crustac. Biol.*, 39 (1):28-35. <https://doi.org/10.1093/jcbl/ruy090>



Matrix multiplication

$$\begin{array}{l}
 \text{row 1} \\
 \text{row 2} \\
 \text{row 3}
 \end{array}
 \begin{pmatrix}
 a_{11} & a_{12} & a_{13} \\
 a_{21} & a_{22} & a_{23} \\
 a_{31} & a_{32} & a_{33}
 \end{pmatrix}
 \begin{array}{l}
 \text{column 1} \quad \text{column 2} \\
 \begin{pmatrix}
 b_{11} & b_{12} \\
 b_{21} & b_{22} \\
 b_{31} & b_{32}
 \end{pmatrix}
 \end{array}
 =
 \begin{pmatrix}
 \boxed{\text{row 1} \cdot \text{column 1}} & \boxed{\text{row 1} \cdot \text{column 2}} \\
 \boxed{\text{row 2} \cdot \text{column 1}} & \boxed{\text{row 2} \cdot \text{column 2}} \\
 \boxed{\text{row 3} \cdot \text{column 1}} & \boxed{\text{row 3} \cdot \text{column 2}}
 \end{pmatrix}$$

$$\begin{pmatrix}
 5 & 9 & 3 \\
 \boxed{2} & \boxed{1} & \boxed{4} \\
 7 & 6 & 8
 \end{pmatrix}
 \begin{pmatrix}
 10 & \boxed{14} \\
 11 & \boxed{17} \\
 15 & \boxed{12}
 \end{pmatrix}
 =
 \begin{pmatrix}
 c_{11} & c_{12} \\
 \boxed{c_{21}} & \boxed{c_{22}} \\
 c_{31} & c_{23}
 \end{pmatrix}$$

$$c_{22} = (2 \times 14) + (1 \times 17) + (4 \times 12) = 24 + 17 + 48 = \underline{\underline{89}}$$

Matrix population models

- In a **deterministic** application of a matrix population model, we assume that population changes will follow the “average” (which are the parameters in the projection matrix)
- We can compute these changes by multiplying the **state vector** by the **projection matrix** using matrix multiplication

n life states : $n \times n$ projection matrix, $n \times 1$ state vector

$$\begin{pmatrix} 0.2 & 0 & 4.5 \\ 0.7 & 0.3 & 0 \\ 0 & 0.4 & 0.5 \end{pmatrix} \begin{pmatrix} 10 \\ 20 \\ 4 \end{pmatrix} = \begin{pmatrix} 0.2 \times 10 \text{ stay} + 4.5 \times 4 \text{ offspring} \\ 0.7 \times 10 \text{ matured} + 0.3 \times 20 \text{ stay} \\ 0.4 \times 20 \text{ matured} + 0.5 \times 4 \text{ stay} \end{pmatrix} = \begin{pmatrix} 2+8 \\ 7+6 \\ 8+2 \end{pmatrix}$$

projection matrix
current state vector
new state vector

Matrix population models

- In a **deterministic** application of a matrix population model, we assume that population changes will follow the “average” (which are the parameters in the projection matrix)
- We can compute these changes by multiplying the **state vector** by the **projection matrix** using matrix multiplication

projection matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

A projection matrix is a mapping from the current state to the new state vector.
e.g. column 2 corresponds to CURRENT life stage 2
row 3 corresponds to NEW life stage 3

Matrix population models

- In a **deterministic** application of a matrix population model, we assume that population changes will follow the “average” (which are the parameters in the projection matrix)
- We can compute these changes by multiplying the **state vector** by the **projection matrix** using matrix multiplication

projection matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

To find out what happens to individuals in stage 2 look at column 2.

e.g. 20 in life stage 2
⇒ $\begin{cases} \text{contributes } a_{12} \times 20 \text{ to new stage 1} \\ \text{contributes } a_{22} \times 20 \text{ to new stage 2} \\ \text{contributes } a_{32} \times 20 \text{ to new stage 3} \end{cases}$

Matrix population models

- In a **deterministic** application of a matrix population model, we assume that population changes will follow the “average” (which are the parameters in the projection matrix)
- We can compute these changes by multiplying the **state vector** by the **projection matrix** using matrix multiplication

projection
matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

To find out how many individuals will be in the new stage 3 look at row 3

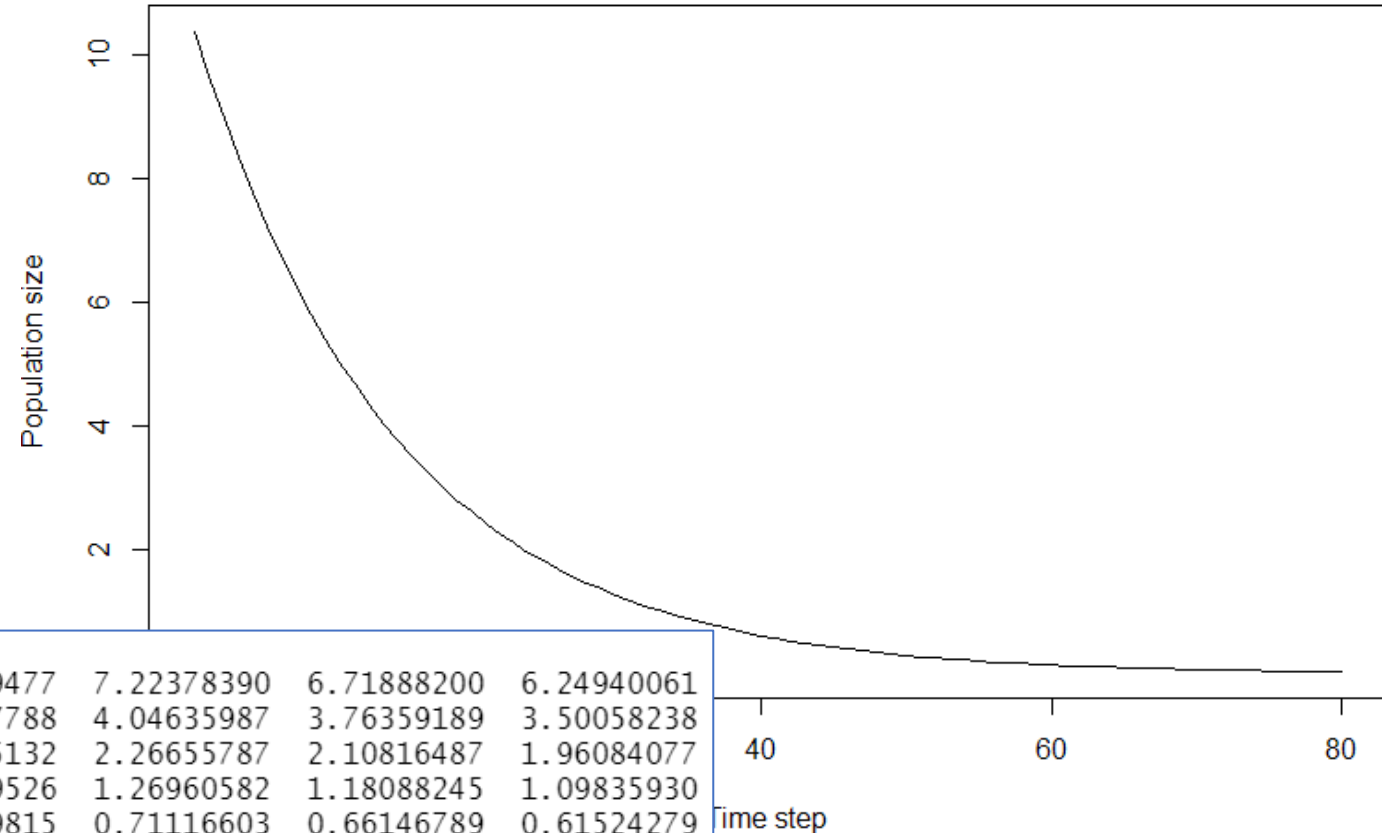
$$\Rightarrow \left\{ \begin{array}{l} a_{31} \times \text{current stage 1} \\ a_{32} \times \text{current stage 2} \\ a_{33} \times \text{current stage 3} \end{array} \right\} \text{ will contribute to new stage } \underline{\underline{3}}$$

Matrix population models

- The matrix formulation is helpful because there are some mathematical shortcuts we can use:
 - The population growth rate* is the **eigenvalue** of the matrix
 - The **eigenvector** represents the eventual proportions of individuals in each stage
- Limitations:
 - random outcomes such as random extinction cannot occur
 - technically the population state can never reach zero (it can only get closer and closer to zero) so we would have to define a threshold for extinction if we were modelling this

Matrix population models – no extinction?

The projection matrix used here is making the population smaller and smaller, but the population size actually never reaches zero.



```
> population_size_time_series
[1] 10.37686161  9.65243561  8.97725984  8.34983123  7.76659477  7.22378390  6.71888200  6.24940061
[9]  5.81269201  5.40646068  5.02864527  4.67723943  4.35037788  4.04635987  3.76359189  3.50058238
[17]  3.25595163  3.02841747  2.81678401  2.61993959  2.43685132  2.26655787  2.10816487  1.96084077
[25]  1.82381210  1.69635935  1.57781332  1.46755160  1.36499526  1.26960582  1.18088245  1.09835930
[33]  1.02160308  0.95021079  0.88380757  0.82204479  0.76459815  0.71116603  0.66146789  0.61524279
[41]  0.57224801  0.53225783  0.49506226  0.46046601  0.42828744  0.39835760  0.37051933  0.34462647
[49]  0.32054307  0.29814268  0.27730769  0.25792871  0.23990397  0.22313886  0.20754533  0.19304152
[57]  0.17955127  0.16700376  0.15533310  0.14447802  0.13438152  0.12499059  0.11625592  0.10813165
[65]  0.10057513  0.09354667  0.08700939  0.08092895  0.07527342  0.07001312  0.06512042  0.06056963
[73]  0.05633687  0.05239990  0.04873806  0.04533211  0.04216419  0.03921764  0.03647701  0.03392790
```

Population modelling

1. Population modelling overview
and stochastic vs deterministic
2. Demographic population models
and matrix population models

3. Stochastic adaptation

Stochastic version

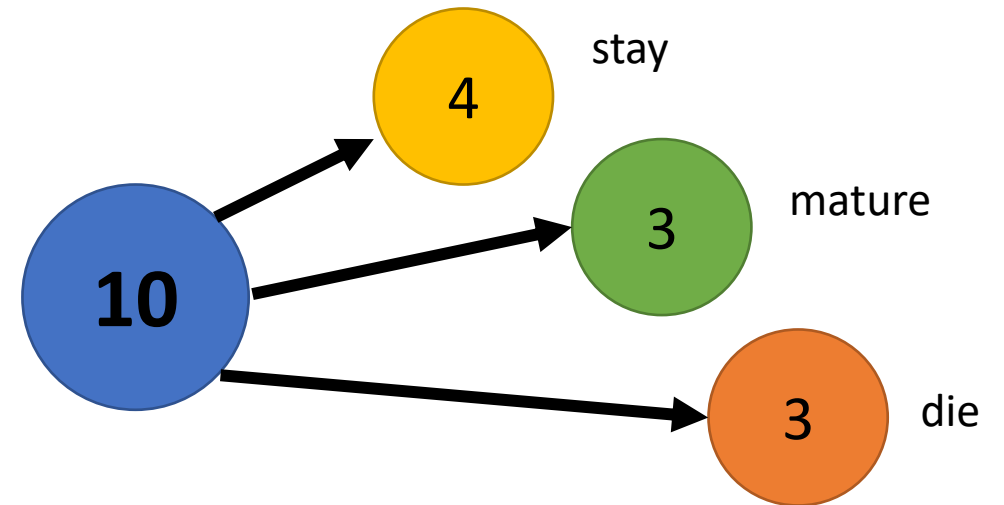
- As discussed already, deterministic models do not always reflect reality
- We will carry out the same kind of age-structured population model except this time we will not assume determinism – processes will be stochastic
- This will allow us to compare the deterministic and stochastic equivalents
- How can we convert biological parameters inferred from field studies into a stochastic formulation?
- **Think about every process involved/assumed in the model and what probability distribution it is drawn from**

Making a deterministic matrix population model stochastic: Survival and maturation

Suppose there are 10 individuals in stage 2. The projection matrix tells us that at the next time step **we expect** a proportion 0.4 to stay in stage 2, and **we expect** a proportion 0.3 to have matured into stage 3. (We expect the remaining proportion 0.3 to have died.)

$$\begin{pmatrix} 0.1 & 0 & 3.5 \\ 0.5 & 0.4 & 0 \\ 0 & 0.3 & 0.6 \end{pmatrix}$$

Expectation:

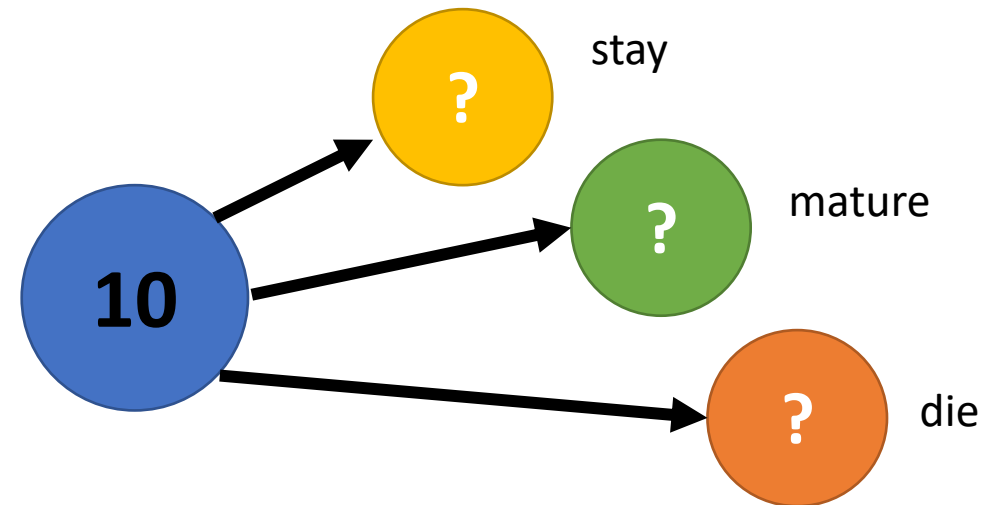


Making a deterministic matrix population model stochastic: Survival and maturation

Suppose there are 10 individuals in stage 2. The projection matrix tells us that at the next time step **we expect** a proportion 0.4 to still be in stage 2, and **we expect** a proportion 0.3 to have matured into stage 3. (We expect the remaining proportion 0.3 to have died.)

In a stochastic formulation, we consider the **probability** of each of these three events (staying in the same life stage, maturing into the next life stage, and death) for every individual.

$$\begin{pmatrix} 0.1 & 0 & 3.5 \\ 0.5 & 0.4 & 0 \\ 0 & 0.3 & 0.6 \end{pmatrix}$$



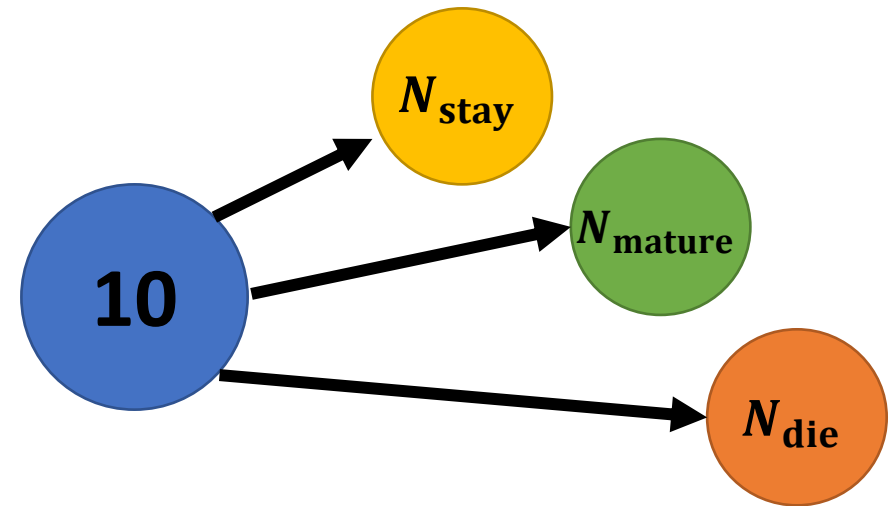
Making a deterministic matrix population model stochastic: Survival and maturation

For each individual in stage 2,

$$P(\text{stay in stage 2}) = 0.4$$

$$P(\text{mature into stage 3}) = 0.3$$

$$P(\text{die}) = 0.3$$



So if there were 10 individuals in stage 2, then we could draw the outcomes as

$$(N_{\text{stay}}, N_{\text{mature}}, N_{\text{die}}) = \text{Multinomial}(n = 10, p = (0.4, 0.3, 0.3))$$

However, in base R there is not a multinomial random draw function. Can we split this up into two steps?

Making a deterministic matrix population model stochastic: Survival and maturation & **multinomial random draw**

For each individual in stage 2,

$$P(\text{stay in stage 2}) = 0.4$$

$$P(\text{mature into stage 3}) = 0.3$$

$$P(\text{die}) = 0.3$$

stay 0.4	mature 0.3
	die 0.3

First we can draw the number of individuals assigned to the first of the events:

$$N_{\text{stay}} \sim \text{Bin}(n = 10, p = 0.4)$$

→ Suppose the number drawn was $N_{\text{stay}} = 5$

Then we can draw the number of individuals assigned to the second event:

$$N_{\text{mature}} \sim \text{Bin}\left(n = 10 - 5, p = \frac{0.3}{1 - 0.4}\right) = \text{Bin}(n = 5, p = 0.5)$$

→ Suppose the number drawn was $N_{\text{mature}} = 2$

And finally the remaining unassigned individuals must be assigned to the third event: $N_{\text{die}} = 5 - 2 = 3$

Making a deterministic matrix population model stochastic: Survival and maturation – applying to all life stages

- When looking at **life stage 1**, we need to find out N_{stay} and N_{mature} for this stage. These will correspond to stage 1 and stage 2 respectively. Suppose that for life stage 1, $N_{\text{stay}} = 5$ and $N_{\text{mature}} = 27$.

$$\begin{pmatrix} 0.1 & 0 & 3.5 \\ 0.5 & 0.4 & 0 \\ 0 & 0.3 & 0.6 \end{pmatrix}$$


Making a deterministic matrix population model stochastic: Survival and maturation – applying to all life stages

- When looking at **life stage 1**, we need to find out N_{stay} and N_{mature} for this stage. These will correspond to stage 1 and stage 2 respectively. Suppose that for life stage 1, $N_{\text{stay}} = 5$ and $N_{\text{mature}} = 27$.

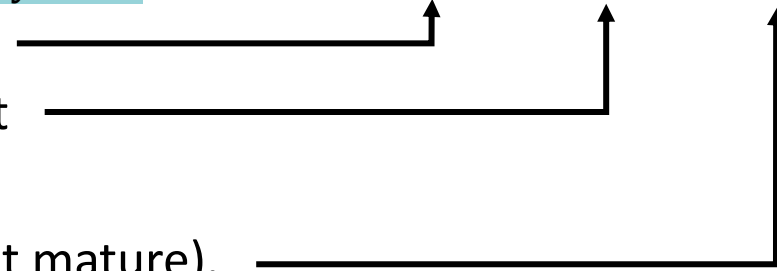
- We need to do the same for **life stage 2**. Suppose that $N_{\text{stay}} = 10$ and $N_{\text{mature}} = 8$.

$$\begin{pmatrix} 0.1 & 0 & 3.5 \\ 0.5 & 0.4 & 0 \\ 0 & 0.3 & 0.6 \end{pmatrix}$$

Making a deterministic matrix population model stochastic: Survival and maturation – applying to all life stages

- When looking at **life stage 1**, we need to find out N_{stay} and N_{mature} for this stage. These will correspond to stage 1 and stage 2 respectively. Suppose that for life stage 1, $N_{\text{stay}} = 5$ and $N_{\text{mature}} = 27$.
- We need to do the same for **life stage 2**. Suppose that $N_{\text{stay}} = 10$ and $N_{\text{mature}} = 8$.
- For **life stage 3**, individuals can only survive or die (not mature). Suppose that $N_{\text{stay}} = 14$

$$\begin{pmatrix} 0.1 & 0 & 3.5 \\ 0.5 & 0.4 & 0 \\ 0 & 0.3 & 0.6 \end{pmatrix}$$



Making a deterministic matrix population model stochastic: Survival and maturation – applying to all life stages

- When looking at **life stage 1**, we need to find out N_{stay} and N_{mature} for this stage. These will correspond to stage 1 and stage 2 respectively. Suppose that for life stage 1, $N_{\text{stay}} = 5$ and $N_{\text{mature}} = 27$.
- We need to do the same for **life stage 2**. Suppose that $N_{\text{stay}} = 10$ and $N_{\text{mature}} = 8$.
- For **life stage 3**, individuals can only survive or die (not mature). Suppose that $N_{\text{stay}} = 14$

$$\begin{pmatrix} 0.1 & 0 & 3.5 \\ 0.5 & 0.4 & 0 \\ 0 & 0.3 & 0.6 \end{pmatrix}$$

Then to obtain the population state at the next time step, we need to add these together in the correct entries of the life state vector:

$$\begin{pmatrix} 5 \\ 27 + 10 \\ 8 + 14 \end{pmatrix}$$

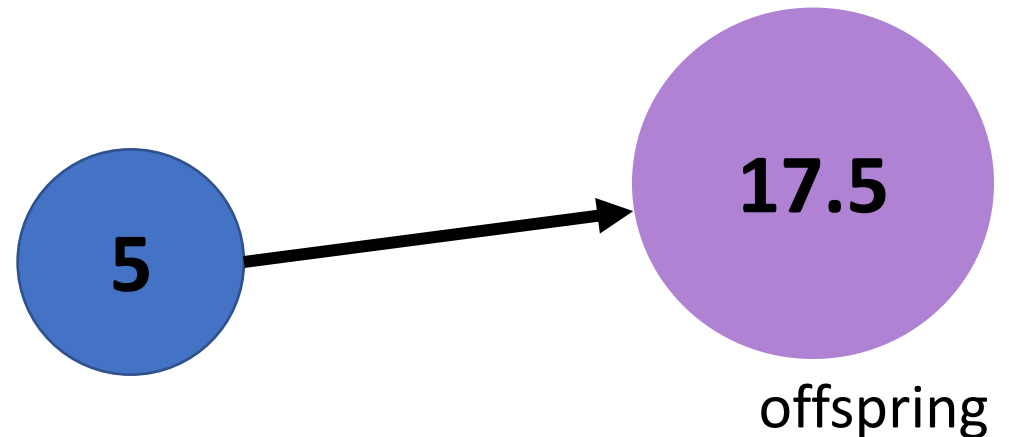
But this doesn't include recruitment – we ignored the 3.5 entry of the vector since this did not reflect survival or maturation.

Making a deterministic matrix population model stochastic: Recruitment

Suppose there are 5 individuals in stage 3. The projection matrix tells us that at the next time step **we expect** each individual to have recruiting 3.5 new offspring, which will be in stage 1.

$$\begin{pmatrix} 0.1 & 0 & 3.5 \\ 0.5 & 0.4 & 0 \\ 0 & 0.3 & 0.6 \end{pmatrix}$$

Expectation:

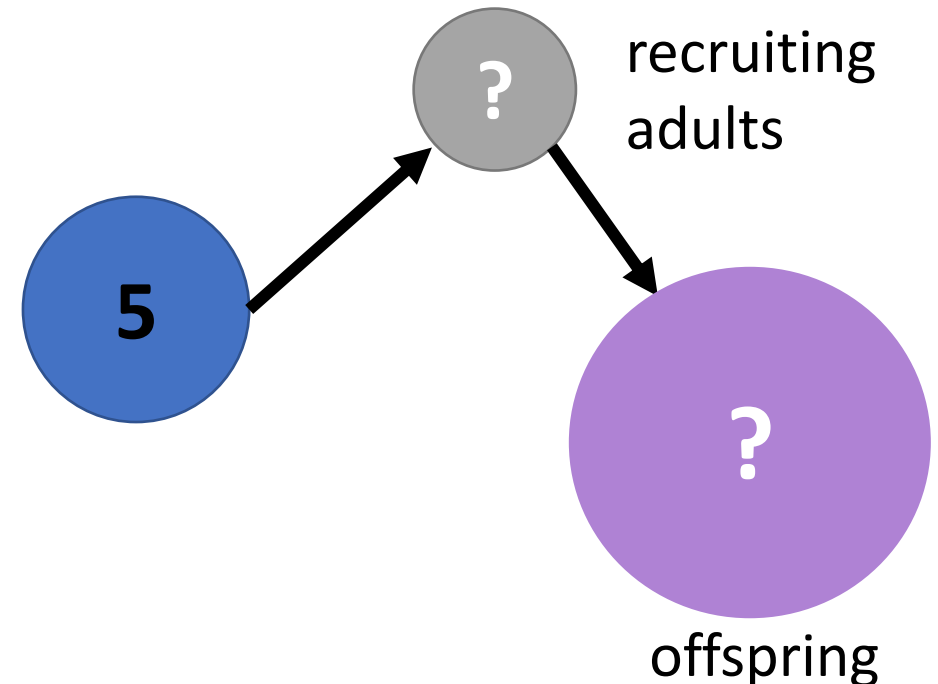


Making a deterministic matrix population model stochastic: Recruitment

Suppose there are 5 individuals in stage 2. The projection matrix tells us that at the next time step **we expect** each individual to have recruiting 3.5 new offspring.

$$\begin{pmatrix} 0.1 & 0 & 3.5 \\ 0.5 & 0.4 & 0 \\ 0 & 0.3 & 0.6 \end{pmatrix}$$

In a stochastic formulation, we consider the **probability** of each individual recruiting (either they recruit or they don't) and also the **probability** of the litter/clutch size taking different values.

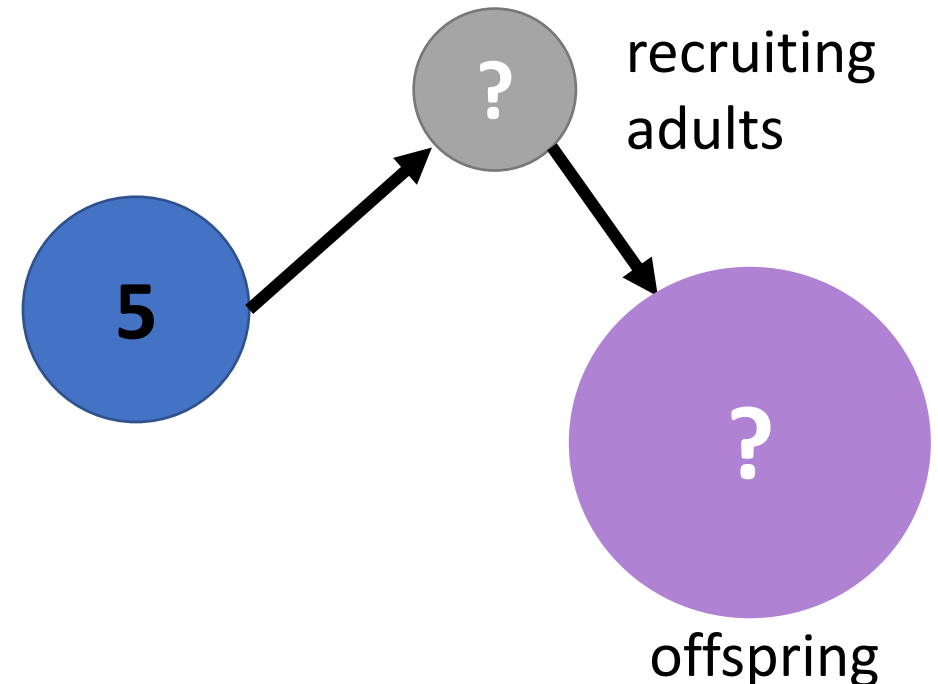


Making a deterministic matrix population model stochastic: Recruitment

We have a problem here because the **probability of a given adult recruiting** and the **probability of each clutch/litter size taking different values** have been conflated into one value, 3.5, which is the “recruitment rate”.

We need to **separately** know the **recruitment probability** and the **probability distribution for clutch/litter size**.

$$\begin{pmatrix} 0.1 & 0 & 3.5 \\ 0.5 & 0.4 & 0 \\ 0 & 0.3 & 0.6 \end{pmatrix}$$



Making a deterministic matrix population model stochastic: Recruitment

Suppose from field studies we also knew the probability distribution for clutch/litter sizes looked like this:

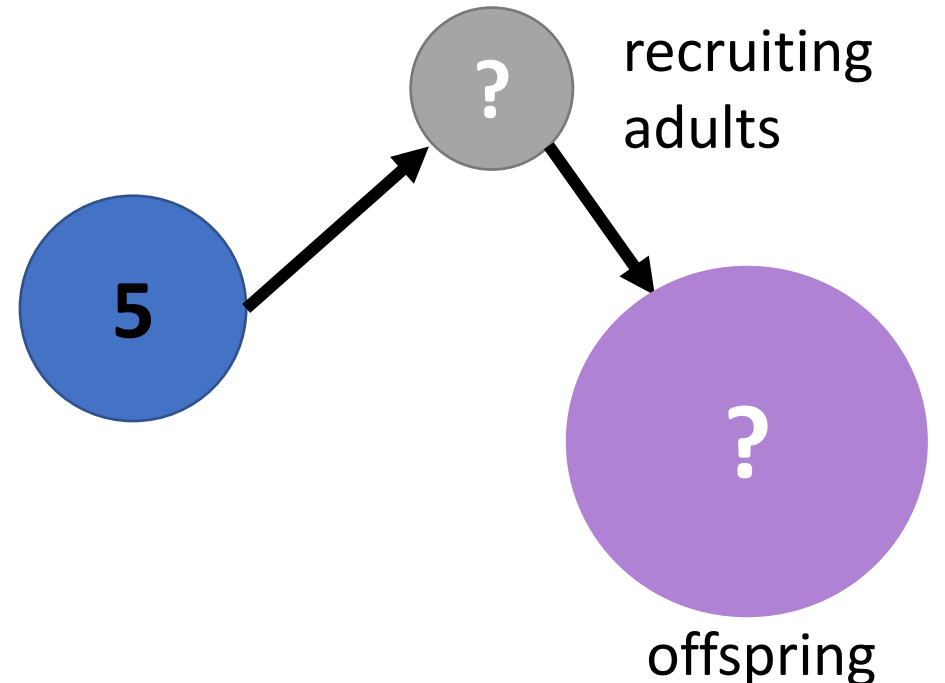
$$\begin{pmatrix} 0.1 & 0 & 3.5 \\ 0.5 & 0.4 & 0 \\ 0 & 0.3 & 0.6 \end{pmatrix}$$

Then we could compute the **expected clutch/litter size** using the probability distribution as

$$E(\text{clutch size}) = \sum_n nP(\text{clutch size} = n)$$

And we can then compute the individual **recruitment probability** as

$$P(\text{recruit}) = \frac{3.5}{E(\text{clutch size})}$$



Making a deterministic matrix population model stochastic: Recruitment

Suppose that the expected clutch size was

$$E(\text{clutch size}) = 7 \quad \text{so} \quad P(\text{recruit}) = 0.5$$

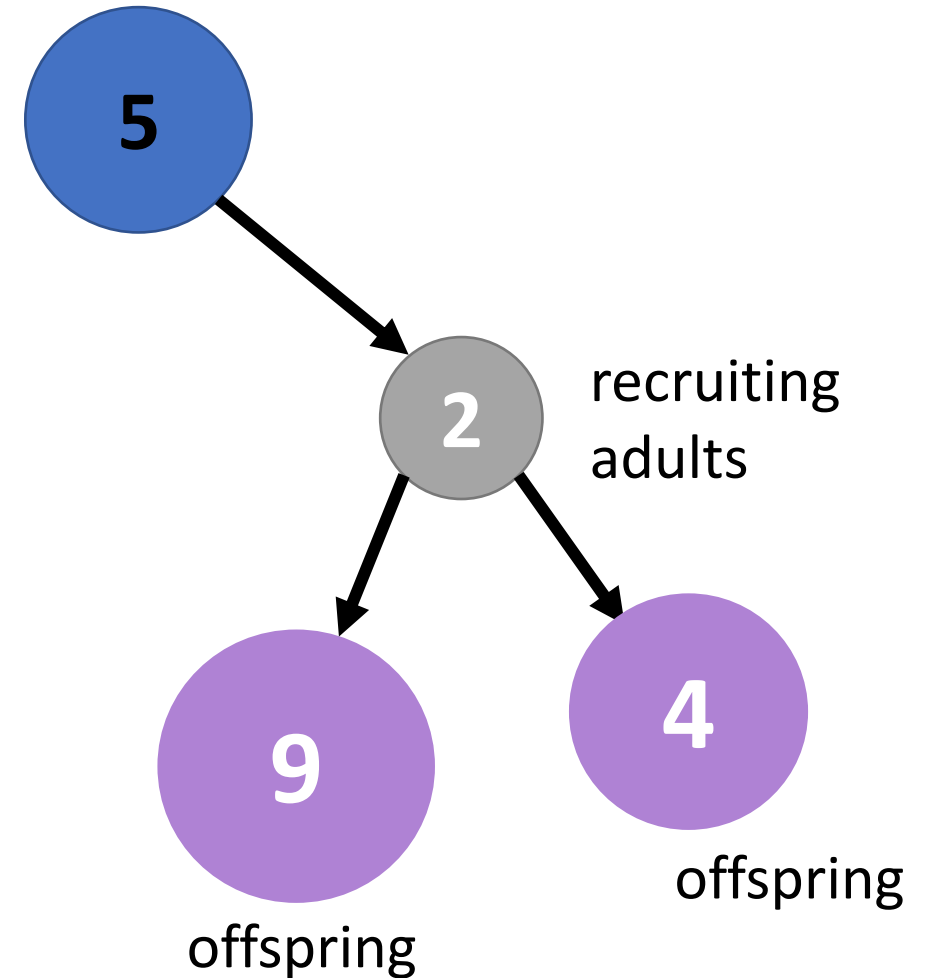
Using these probabilities, we can now draw the number of individuals which recruit:

$$N_{\text{recruit}} = \text{Bin}(n = 5, p = 0.5)$$

→ Suppose the number drawn was $N_{\text{recruit}} = 2$

For each clutch, we could then draw the clutch size based on our clutch size distribution.

Supposing the numbers drawn were 9 and 4 for each clutch respectively, then the total number of offspring would be 13.



Making a deterministic matrix population model stochastic: Recruitment & drawing from a distribution

But how can we obtain the numbers of offspring from a custom probability distribution?

Suppose that the probability distribution was simple and looked like:

$$P(\text{clutch size} = 1) = 0.2$$

$$P(\text{clutch size} = 2) = 0.5$$

$$P(\text{clutch size} = 3) = 0.3$$

One common way to draw a number from a given distribution is to **first** draw a number from the continuous uniform distribution on the interval (0,1):

$$u \sim \text{Unif}((0,1))$$

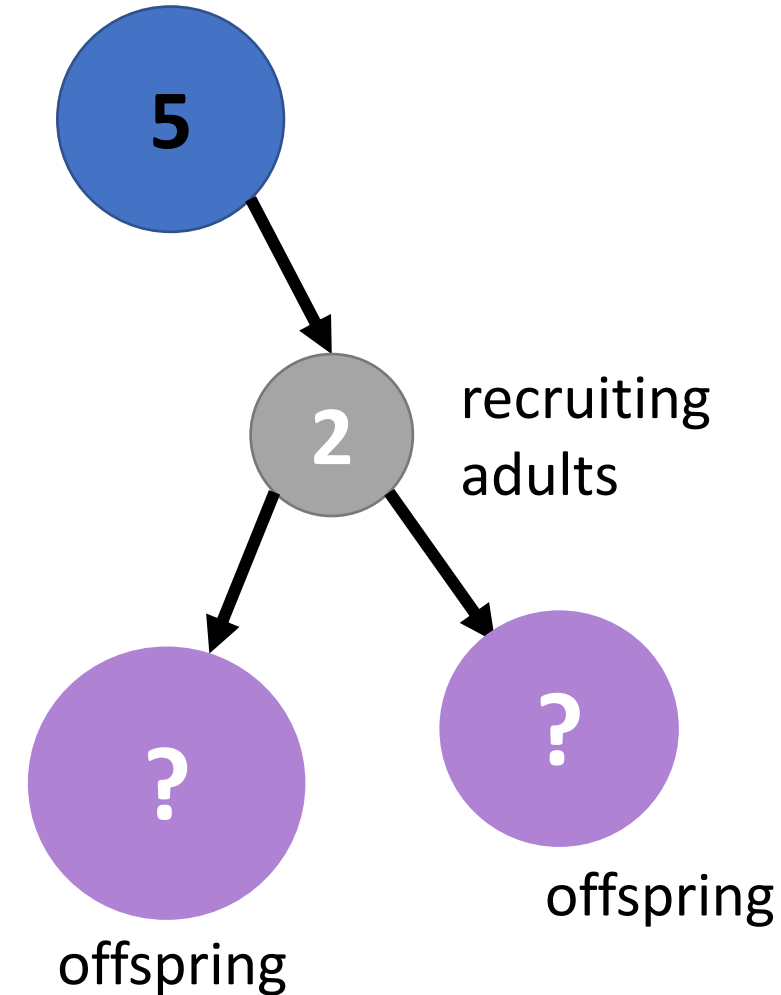
We can then **map this uniform distribution onto our custom distribution** as follows:

If $0 \leq u < 0.2$ then clutch size = 1

If $0.2 \leq u < 0.7$ then clutch size = 2

If $0.7 \leq u \leq 1$ then clutch size = 3

Because each interval width is equal to the corresponding probability in our distribution, we get the correct event probabilities.



Making a deterministic matrix population model stochastic: Applying processes in turn – **common mistakes**

One of the challenges with applying a stochastic simulation is that different processes have to be applied one-by-one, which seems counterintuitive since in a deterministic model everything happens at the same time.

It is important **not to make these mistakes** with our stochastic application of a matrix population model:

1. Apply survival and maturation and find out what the **new population** looks like after these processes
2. Apply recruitment and find out how many offspring this **new population** recruits, then add them into the population

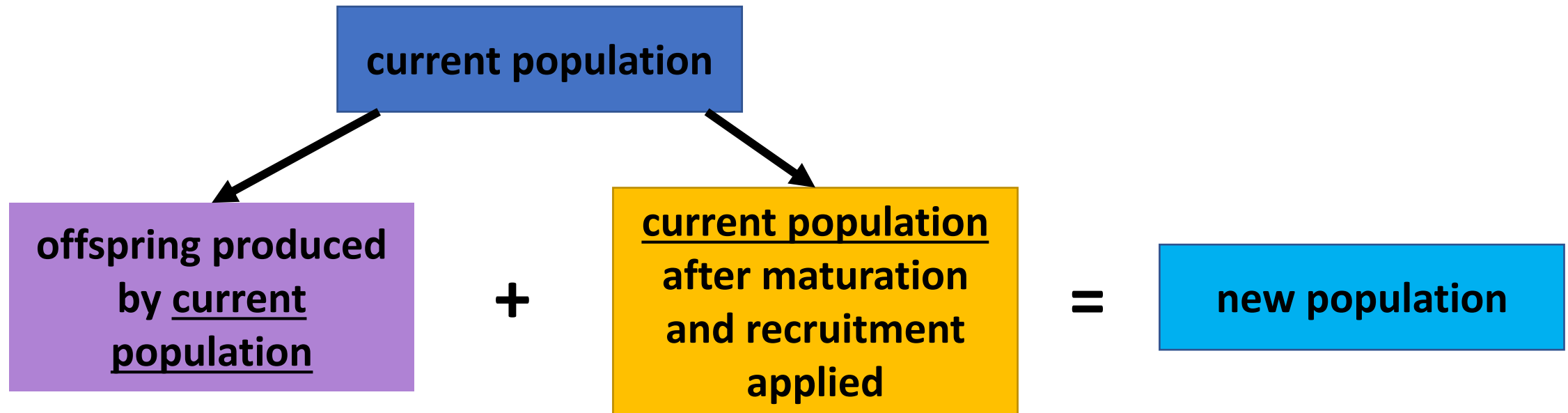
Or alternatively

1. Apply recruitment and find out how many offspring are produced, then **add them into the population**
2. Apply survival and maturation to the new population **including the new offspring**

We have to be careful to avoid these mistakes as they effectively jump ahead and make the processes occur more quickly than they should.

Making a deterministic matrix population model stochastic: Applying processes in turn

One of the challenges with applying a stochastic simulation is that different processes have to be applied one-by-one, which seems counterintuitive since in a deterministic model everything happens at the same time.



The resulting model – a stochastic demographic population model

Stochastic because processes are drawn from random distributions

Demographic because the population is structured by life stage (a demographic trait)

Pros:

- In general, it is a better reflection of reality than the deterministic model (as long as we run it many times to observe outcomes)
- We can observe random extinction and the population can reach zero, so we can compute extinction probability

Cons:

- It needs to be run many times to see a representative sample of outcomes
- It was more awkward to code than the deterministic matrix population model
- It required additional data to simulate correctly (clutch size distribution)