Appendix 2

The purpose of this appendix is to explicitly consider how our inability to measure some aspects of environmental variation affects our interpretation of a species niche. To do this I will consider the effects of two resources \((x_1, x_2)\), when it is only possible to measure the effects of variable \(x_2\). We can then ask what a particular value of \(x_2\) tells us about whether an environment will be suitable (i.e. what is the probability that an environment is a part of the niche, conditioned on \(x_2\)). We can also ask how much \(x_1\) is necessary for organisms to persist in a particular environment (i.e. what would a small scale study infer about the required level of \(x_2\)).

To determine the probability that an environment is suitable we must further define the set of environments that are available in nature. I will assume that each environmental variable is independent and uniformly distributed, and for convenience is scaled to range from zero to one. This represents a simple distribution of environments that is easy to plot.

Case I: Hutchinsonian resources:

Hutchinson originally described organisms that require intermediate levels of each resource, i.e.

\[
P(N) = \begin{cases} 
1 & \text{if } x_{1,\text{min}} < x_1 < x_{1,\text{max}} \text{ and } x_{2,\text{min}} < x_2 < x_{2,\text{max}} \\
0 & \text{otherwise} 
\end{cases} 
\]  

See Fig. A1 for a diagram of this type of resource requirements. If organisms can persist in environments with the maximum possible levels of both resources \((x_{1,\text{max}} x_{2,\text{max}} = 1)\) this formulation also represents essential resources.

We can use Eq. 1 to define the probability that an environment will be suitable given \(x_1\) as the proportion of events that will allow a species to persist, at a particular value of \(x_2\). If \(x_2 \leq x_{2,\text{min}}\) or \(x_2 \geq x_{2,\text{max}}\) organisms will never persist, the probability that these environments will be a part of the niche is thus zero. For values \(x_{2,\text{min}} < x_2 < x_{2,\text{max}}\) the probability that an environment will be suitable is the probability \(x_{1,\text{min}} < x_1 < x_{1,\text{max}}\). Over the interval \([0,1]\) this is equal to:

\[
x_{1,\text{max}} - x_{1,\text{min}}
\]

So the probability that a species is present, given \(x_2\) is:

\[
P(N|x_2) = \begin{cases} 
x_{1,\text{max}} - x_{1,\text{min}} & \text{if } x_{2,\text{min}} < x_2 < x_{2,\text{max}} \\
0 & \text{otherwise} 
\end{cases}
\]

The outcome of a causal study will vary dramatically depending on the value of \(x_1\). If scientists happen to conduct a study in an environment where \(x_{1,\text{min}} \leq x_1 \leq x_{1,\text{max}}\) values of \(x_2\) between \(x_{2,\text{min}}\) and \(x_{2,\text{max}}\) will make an environment suitable, while all other values will make an environment unsuitable. The predicted probability that an organism will survive then be:

\[
\hat{P}(N|x_2) = \begin{cases} 
1 & \text{if } x_{2,\text{min}} < x_2 < x_{2,\text{max}} \\
0 & \text{otherwise} 
\end{cases}
\]

If the study occurs in any other environment, no value of \(x_2\) will allow the organism to persist and so \(P(N) = 0\). Assuming that such a result would not be published, a mechanistic model will accurately estimate the probability that an environment is unsuitable, but overestimate the probability that an organism is suitable (Fig. 2).

If an experimental study would be published when no organisms survived, the expected probability that an environment will be suitable given \(x_{2,\text{min}} \leq x_2 \leq x_{2,\text{max}}\) is equal to:
\[ P(N \mid x_2) = P(N \mid x_{1,\text{min}} < x_1 < x_{1,\text{max}}) P(x_{1,\text{min}} < x_1 < x_{1,\text{max}}) + P(N \mid (x_{1,\text{min}} < x_1 < x_{1,\text{max}})) P((x_{1,\text{min}} < x_1 < x_{1,\text{max}})) \]
\[ = 1 \left( x_{1,\text{max}} - x_{1,\text{min}} \right) + 0 \left( 1 - \left( x_{1,\text{max}} - x_{1,\text{min}} \right) \right) \]
\[ = x_{1,\text{max}} - x_{1,\text{min}} \]

In this case the probability that an environment will be suitable, across many studies will be the same as the true probability that an environment is suitable.

**Case II substitutable resources:**

If two resources are substitutable, the probability that an environment will be part of the niche is given by:

\[
P(N) = \begin{cases} 
1 & \text{if } x_2 \geq mx_o + b \\
0 & \text{otherwise} 
\end{cases}
\]

See Fig. S1A for a diagram of this type of resource requirements. To determine the probability that an environment is suitable given a particular value of \( x_1 \) first note that the value \( x_1 \) on the line is given \( mx_o + b \). When \( m \) is negative, and this value is less than zero all environments are suitable. Assuming that \( m \) is negative and this value is greater than one, then no environments are suitable. The probability that an environment is suitable when the line is between zero and one is:

\[
P(N \mid x_2) = \begin{cases} 
1 & \text{if } x_1 \leq 0 \\
\frac{x_1 - b}{m} & \text{if } 0 \leq x_1 \leq 1 \\
0 & \text{if } 1 \leq x_1 
\end{cases}
\]

When we examine a particular environment, \( x_1 \) is fixed to a particular value \( x_o \). Organisms will then persist when \( x_2 \geq mx_o + b \). A mechanistic understanding of a species niche would lead to the following rule:

\[
\hat{P}(N \mid x_2) = \begin{cases} 
1 & \text{if } x_2 \geq mx_o + b \\
0 & \text{otherwise} 
\end{cases}
\]

An experimental determination of the effect of one variable can be radically different from the relationship between that variable and whether a species will persist in nature. Notably, it may be impossible or undesirable to measure the effects of variable \( x_2 \) in the absence of variable \( x_1 \) (i.e. when \( x_1 = 0 \)).
Figure A1. (A) An illustration of a species niche defined by Hutchisonian resources. An environment is only suitable if the amount of resource present is above a minimum threshold and above a maximum threshold for each resource. (B) An illustration of a species niche defined by substitutable resources, where a species can persist when there is a sufficient quantity of both resources, i.e. when $x_2 \geq m_0 + b$. 