SAR vs EAR

Dan McGlinn

1/31/2020

Motivating Question: Do the SAR and EAR under random placement contain unique information?

It has been proposed that the EAR is potentially of more interest to conservation biologists as they examine the effects of habitat loss.

For the purposes of MoB we want to make informative comparisons between the scaling of biodiversity in different treatments. We have been using the IBR which is essentially a random placement model (RPM) for the SAR. If the RPM SAR and EAR curves possess unique information that could be informative to an empiricist.

To examine if the RPM EAR and SAR do in fact have unique information we'll examine their form. Here it is important to remember that by definition both the SAR and EAR have no other information than the SAD. They are just different expressions of the marginal cumulative probability of that distribution. The question I have is if those different formulations are informative or simply redundant with one another.

Note the models are defined in terms of area but that can be substituted for number of individuals.

```r
exp_S_binom = function(A, A0, ni) {
  ## Expected number of species from Coleman (1981), Eq. 1.12
  ## See Harte 2011 Eq. 3.13 for an alternative formultaion using the SAD
  ## Arguments:
  ## A: sample area or number of individuals
  ## A0: the total area or the total number of individuals across all species
  ## ni: the species abundance distribution
  ## Returns:
  ## the average expected number of species under the binomial distr for
  ## a sample of area A out of A0 and species abundances ni
  S = sum(1 - (1 - (A/A0))^ni)
  # alternatively but identically
  #S = sum(1 - dbinom(0, ni, A/A0))
  return(S)
}


exp_E_binom = function(A, A0, ni) {
  ## Expected number of endemic species under random placement
  ## see argument definitions in exp_S_binom
  # note this forumulation is just very subtely different from
  # exp_S_binom the question changes from if any single individual of a
  # species occurs when calculating S to if all of the individuals of a
  # species occur when calculating E
  E = sum(1 - (1 - (A/A0)^ni))
  # alternatively but identically
  #E = sum(dbinom(ni, ni, A/A0))
  # the usage of dbinom does is more initutiative
```

```
    return(E)
}
```

OK let's compare the SAR and EAR for a simulated community with a log normal SAD that has 20 species in it

```
Spool = 20
sad = ceiling(rlnorm(Spool, 1, 1))

# Now compute the SAR and EAR

N = sum(sad)
n = seq(1, N, 1)
S = sapply(n, function(n) exp_S_binom(n, N, sad))
E = sapply(n, function(n) exp_E_binom(n, N, sad))
SfromE = Spool - sapply(n, function(n) exp_E_binom(N - n, N, sad))
EfromS = Spool - sapply(n, function(n) exp_S_binom(N - n, N, sad))

plot(n, S, xlab = 'Number of Individuals (n)', ylab = 'Number of Species',
     ylim = c(0, Spool), col='red')
points(n, E, col='blue')
lines(n, EfromS, col='red')
lines(n, SfromE, col='blue')
legend('bottomright', c('S(n)', 'E(n)', 'E(n) = Spool - S(N-n)', 'S(n) = Spool - E(N-n)'),
       col=c('red', 'blue', 'red' , 'blue'), pch = c(1, 1, NA, NA),
       lty = c(NA, NA,1, 1), bty='n', cex = 0.75)
```
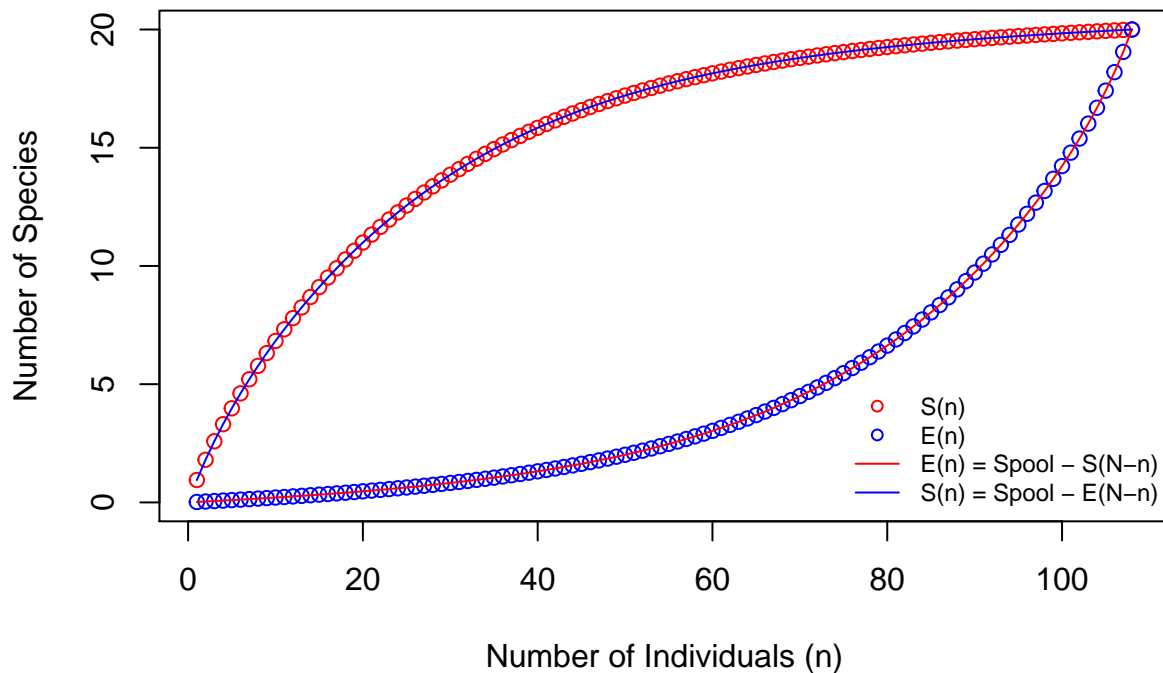

The above shows that the RPM SAR can be re-expressed as the RPM EAR and vice-versa. While that may be informative in certain conservation applications, for the purposes of comparing the spatial structure of biodiversity the two curves do not contain unique information. Therefore, only one of them should be examined when comparing treatments or testing the strength of a theory. If a theory accurately predicts the SAR then it by definition also can get the EAR it would seem.

So in regards to keeping MoB simple this is good news. This is somewhat bad news for developing strong tests of theory because it says really you only need to test one of these relationships not both of them. I'm not sure if others already know this. I guess maybe its obvious when you consider that the EAR is just species that you haven't sampled in area A because they all exist in area A0 - A.

The other way this relates to a broader MoB analysis of beta-diversity which is why I got interested in the EAR is that it tells us something about the information we gain when we consider unique species between sets of samples. The endemic species are unique species that are contained entirely in the sample of n individuals which means that they are the species that will turnover if you compare to a mutually exclusive set of n individuals (i.e., sampling without replacement) from the larger sample.

I haven't worked it all out but where I'm going is that multi-site beta-diversity metrics can be derived from just the IBR it would seem under an RPM. If we could show this clearly then it would be an important step forward in the unification of beta diversity metrics. Note this also has close ties with zeta diversity as well.