

Notes on Spark Project

Paulinha

Overview

Costs and benefits are inherent aspects of species interactions. Costs associated with mutualistic interactions, such as pollination and seed dispersal, involve production of nectar and fruit from a plant perspective and from an animal perspective there are costs associated with finding and acquiring nectar as well as digesting and processing fruits. The benefits are associated with the provision of resources to the animals and the reproduction to plants. Beyond the importance of the mutualistic interactions to the persistence of interacting species, these interactions also provide ecosystem services, such as production of crops and landscape connectivity. The costs and benefits associated with species interactions are not fixed and are subjected to changes due to environmental conditions, which scale up affecting the provision of ecosystem services.

We developed a framework to explore how costs and benefits associated with species interactions, and hence ecosystem services, determine the persistence of species and the robustness of service provision under environmental changes. Our modeling framework combines cost-benefit analysis defining a fitness functions that determines the outcome of a Boolean process. The Boolean process establishes which species are present in the community given the costs and benefit associated with interactions between species. Our model allows us understanding transition phases in mutualisms based on increasing costs of interactions, parameterized using empirical networks and sampled cost/benefit functions.

Given the broad nature of the modeling framework we took two complementary approaches. Within the first approach we are interested in defining how the processes incorporated in the model lead to network structures that resemble empirical ones. The second approach takes empirical networks as a starting point and explores the consequences of the processes incorporated in the model to the persistence of mutualisms and ecosystem services. These two approaches are explored concomitantly and attack (i) how processes lead to patterns and (ii) the consequences of processes in the patterns.

Interaction networks

We use our modeling framework as a benchmark to understand what distributions of costs and benefits associated with the interactions leads to networks whose structure resemble empirical networks. Starting from a pool of potential interacting species from two groups (i.e., a bipartite structure of plants and pollinators, for example), we assume all species can potentially interact, given a certain connectance and a random structure to define which species interact. *This is a first approximation, we can also use other models to define potential interactions that take into account realistic degree distributions* Once interactions are established, we define the costs and benefits associated with each interaction for each partner.

Cost and benefit functions

Costs and benefits can be described as functions with a giving underlying distribution.

Beta distribution

Continuous probability distribution defined on the interval $[0, 1]$ in terms of two positive parameters α and β . It is a suitable distribution for modeling the costs and benefits as they are defined only in the positive realm, and can take diverse shapes. Additionally, it provides a limit to the maximum benefit (cost) a species can have. The upper limit in this case is k (the degree of the species, assuming it receives the maximum benefit from all its partners and benefits are an additive function). The beta distribution has been used to model interactions in food web models before Stouffer et al. (2005).

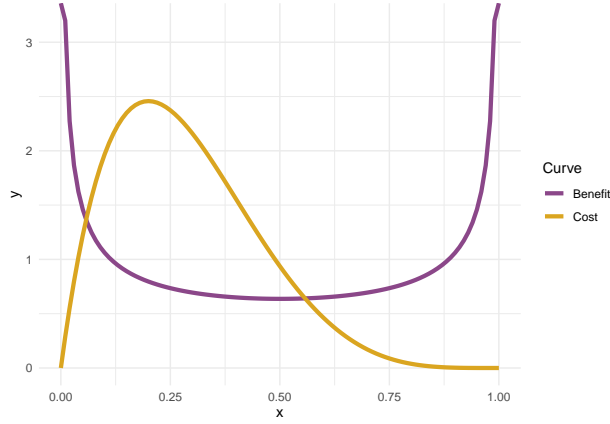


Figure 1: Costs and benefits drawn from a beta distribution

Boundary conditions for costs and benefits and the persistence/extinction of species

What are the boundary conditions for the expectation of having all species present (*benefits* > *costs*) and for having all species extinct of the system (i.e., when *benefits* < *costs*)? In other words, we want to know what is the probability of having the benefits surpass the costs, resulting in all species present in the network, and alternatively, what is the probability of having all species removed from the network if the costs of interaction surpass their benefits. The solution to this problem, considering costs and benefits distributed following independent beta distributions, relies on Appell F1 Hypergeometric function and was derived by Pham-Gia, Turkkan, and Eng (1993). Since the Boolean function determining whether a species will be present is given by a positive fitness, this can be achieved by a species having greater benefits than costs for its interactions (i.e. $\theta_b > \theta_c$), with $\theta_b = X_1 \sim \text{Beta}(\alpha_1, \beta_1)$ and $\theta_c = X_2 \sim \text{Beta}(\alpha_2, \beta_2)$. We want to determine the distribution of the difference $\theta_d = \theta_b - \theta_c$. This difference is piece wise and given by ((Pham-Gia, Turkkan, and Eng 1993)):

Considering $A = \text{Beta}(\alpha_1, \beta_1) \cdot \text{Beta}(\alpha_2, \beta_2)$ and \cdot a product.

For $-1 \leq \theta_d < 0$, meaning costs are greater than benefits, we have:

$$p(\theta_d) = \text{Beta}(\alpha_2, \beta_1) \theta_d^{\beta_1 + \beta_2 - 1} (1 - \theta_d)^{\alpha_2 + \beta_1 - 1} \cdot F_1(\beta_1, \alpha_1 + \beta_1 + \alpha_2 + \beta_2 - 2, 1 - \alpha_1; \beta_1 + \alpha_2; 1 - \theta_d, 1 - \theta_d^2) / A$$

For $0 < \theta_d \leq 1$, with benefits greater than costs, we have:

$$p(\theta_d) = \text{Beta}(\alpha_1, \beta_2) \theta_d^{\beta_1 + \beta_2 - 1} (1 - \theta_d)^{\alpha_1 + \beta_2 - 1} \cdot F_1(\beta_2, 1 - \alpha_2, \alpha_1 + \beta_1 + \alpha_2 + \beta_2 - 2; \alpha_1 + \beta_2; 1 - \theta_d^2, 1 + \theta_d) / A$$

And finally, when the difference between the distribution of benefits and costs is zero $\theta_d = 0$ with $\alpha_1 + \alpha_2 > 1$ and $\beta_1 + \beta_2 > 1$ we have:

$$p(\theta_d) = \frac{\text{Beta}(\alpha_1 + \alpha_2 - 1, \beta_1 + \beta_2 - 1)}{A}$$

F_1 is the Appell F_1 Hypergeometric Function with the following form:

$$F_1(a, b_1, b_2; c; x, y) = \sum_{m+n=0}^{\infty} \frac{(a)_{m+n} (b_1)_m (b_2)_n}{(c)_{m+n} m! n!} x^m y^n$$

with $(q)_m$ representing the Pochhammer symbol, indicating the rising factor. Substituting the parameters, for the negative case ($-1 \leq \theta_d < 0$) the parameters are $a = \beta_1$, $b_1 = \alpha_1 + \beta_1 + \alpha_2 + \beta_2 - 2$, $b_2 = 1 - \alpha_1$; $c = \beta_1 + \alpha_2$; $x = 1 - \theta_2$, $y = 1 - \theta_d^2$. For the negative case ($0 < \theta_d \leq 1$) the parameters are $a = \beta_2$, $b_1 = 1 - \alpha_2$, $b_2 = \alpha_1 + \beta_1 + \alpha_2 + \beta_2 - 2$; $c = \alpha_1 + \beta_2$; $x = 1 - \theta_2^2$, $y = 1 + \theta_d$.

The Ecosystem-Services Matrix

The matrix describing the ecosystem services is composed of services on the rows and species contributing to the services as its columns. This view of multiple species contributing to an ecosystem service recognizes that services are built through species acting as providers of the service per se and species playing a supporting role in the service (Keyes et al. 2021). The structure of the ecosystem service matrix can be explored from a theoretical point of view as well as using specific case studies. For instance, we can connect the structure of the empirical network of interactions with the structure of the ecosystem services matrix, where species that play a central role in the matrix are associated with certain services whereas species in the periphery of the network are associated with other services. In this example we further explore how these services differ in their robustness to species' loss or to environmental changes.

Follow-up questions

How is the robustness of food webs related to the ecosystem services they provide? (Keyes et al. 2021) explored the consequences of secondary extinctions to the maintenance of services in a food web system. They found that food web and service robustness are highly correlated but service robustness is dependent on trophic level. Additionally, they also found that species providing services **do not** necessarily play a critical role in stabilizing food webs and that species that play supporting roles in service are the critical ones to robustness (of food webs and services). How dependent are these results to the underlying system providing the service? In their case, services are provided through a food web, should we expect different results for pollination or seed dispersal systems? In other words *how does the nature of the underlying ecological system influences the robustness of ecosystem services they provide?*

Other potential distributions

The choice of the beta distribution is supported by biological arguments but it is still an arbitrary choice. Other possible distributions of costs and benefits include:

Lognormal distribution As costs and benefits should, in principle, have the same sign, it is more appropriate to use a distribution defined for positive values. The derivation for the lognormal distribution is not as straight forward as the one for the normal distribution, given we are interested in the difference between two i.i.d. random variables drawn from a lognormal distribution. One solution to find the probability that all species are present is through numerical integration assuming both values are drawn from $\ln N(0, 1)$.

Alternatively, (Parham 2023) proposes a solution describing the Difference-of-Log-Normals (DLN), characterizing its PDF, CDF, moments and parameter estimators.

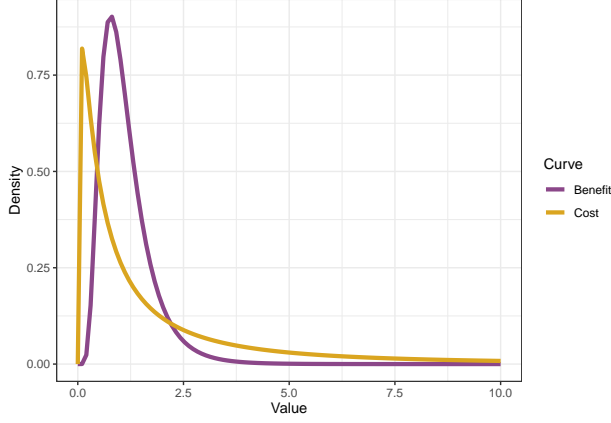


Figure 2: Costs and benefits drawn from a lognormal distribution

Normal distribution Assuming normally and independently distributed costs $C \sim N(\mu_c, \sigma_c^2)$ and benefits $B \sim N(\mu_b, \sigma_b^2)$, we want first to know $P(B > C)$.

$$\begin{aligned} P(B > C) &= P(B - C > 0) \\ &= 1 - P(B - C \leq 0) \end{aligned}$$

If B and C are independent and normally distributed so is their difference, with the expectation of their mean as $\mathbb{E}(B - C) = \mathbb{E}(B) - \mathbb{E}(C) = \mu_b - \mu_c = \mu$ and of their variance as $Var(B - C) = \sigma_b^2 + \sigma_c^2 = \sigma$. Now we subtract the mean and divide by the variance so that our difference is normally distributed with $\mu = 0$ and $\sigma^2 = 1$) such that $\frac{B-C-\mu}{\sigma} \sim N(0, 1)$. And so:

$$P(B > C) = 1 - \Phi\left(\frac{-\mu}{\sigma}\right) \quad (1)$$

where Φ is the cumulative distribution function of $N(0, 1)$.

Now let's consider an example: if species benefits have the distribution $B \sim N(1.2, 2)$ and costs are distributed as $C \sim N(2, 1.5)$, the distribution describing these values is:

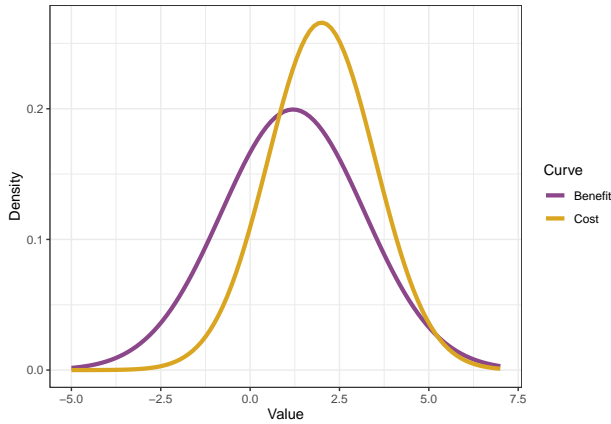


Figure 3: Example of cost-benefit functions following a normal distribution

Using the derived expression, the resulting probability of having all species present in this system, i.e., the $P(B > C)$, meaning benefits are greater than costs is: 0.3344646. Alternatively, the probability that all

species goes extinct in this example is: 0.6655354.

In more general terms, we can explore how the difference between benefits and costs influences the probability of all species coexisting. Assuming a standard deviation of $\sigma^2 = 1$ we explore the effect of the mean difference between costs and benefits in species' probability of persistence, deriving the probability according to equation 1.

```
## Ignoring unknown labels:  
## * colour : "Curve"
```

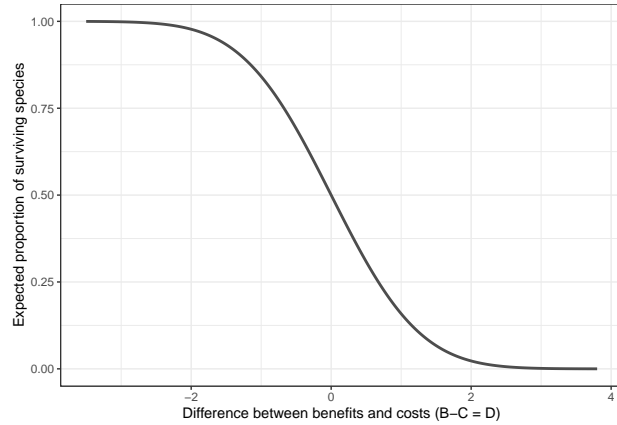


Figure 4: Expected proportion of species' persistence given the difference in benefits and costs

References

- Cattin, Marie-France, Louis-Félix Bersier, Carolin Banašek-Richter, Richard Baltensperger, and Jean-Pierre Gabriel. 2004. “Phylogenetic Constraints and Adaptation Explain Food-Web Structure.” *Nature* 427 (6977): 835–39.
- Dunne, Jennifer A, Richard J Williams, Neo D Martinez, Rachel A Wood, and Douglas H Erwin. 2008. “Compilation and Network Analyses of Cambrian Food Webs.” *PLoS Biology* 6 (4): e102.
- Keyes, Aislyn A, John P McLaughlin, Allison K Barner, and Laura E Dee. 2021. “An Ecological Network Approach to Predict Ecosystem Service Vulnerability to Species Losses.” *Nature Communications* 12 (1): 1586.
- Parham, Robert. 2023. “The Difference-of-Log-Normals Distribution: Properties, Estimation, and Growth.” *arXiv Preprint arXiv:2302.02486*.
- Pham-Gia, Thu, Noyan Turkkan, and P Eng. 1993. “Bayesian Analysis of the Difference of Two Proportions.” *Communications in Statistics-Theory and Methods* 22 (6): 1755–71.
- Stouffer, Daniel B, José Camacho, Roger Guimera, CA Ng, and LA Nunes Amaral. 2005. “Quantitative Patterns in the Structure of Model and Empirical Food Webs.” *Ecology* 86 (5): 1301–11.
- Williams, Richard J, and Neo D Martinez. 2000. “Simple Rules Yield Complex Food Webs.” *Nature* 404 (6774): 180–83.