Heidi Swanson, Martin Lysy, Michael Power, Ashley Stasko, Jim Johnson, and James Reist. 2014. A new probabilistic method for quantifying n-dimensional ecological niches and niche overlap. *Ecology*

APPENDICES

Appendix A. Computation of niche regions.

APPENDIX A: COMPUTATION OF NICHE REGIONS.

Suppose that $X = (X_1, \ldots, X_n)$ is an n-dimensional isotope distribution with joint pdf $f(x)$. We have defined the niche region to be

$$\mathrm{N_R} = \{x : f(x) > C\}, \tag{A.1}$$

where C is chosen such that $P(X \in \mathrm{N_R}) = 95\%$. This is in fact the smallest region occupying 95% of the probability space (e.g., Box and Tiao, 1973; Wei and Tanner, 1990; Chen and Shao, 1999). Figure A1 illustrates this for a one-dimensional isotope X.

For a symmetric distribution (left panel), $\mathrm{N_R}$ simply consists of the interval between the 2.5% and 97.5% quantiles of the distribution. For skewed distributions (middle panel), however, this is not the case. In fact, for bimodal distributions (right panel) the niche region can even be disjoint.

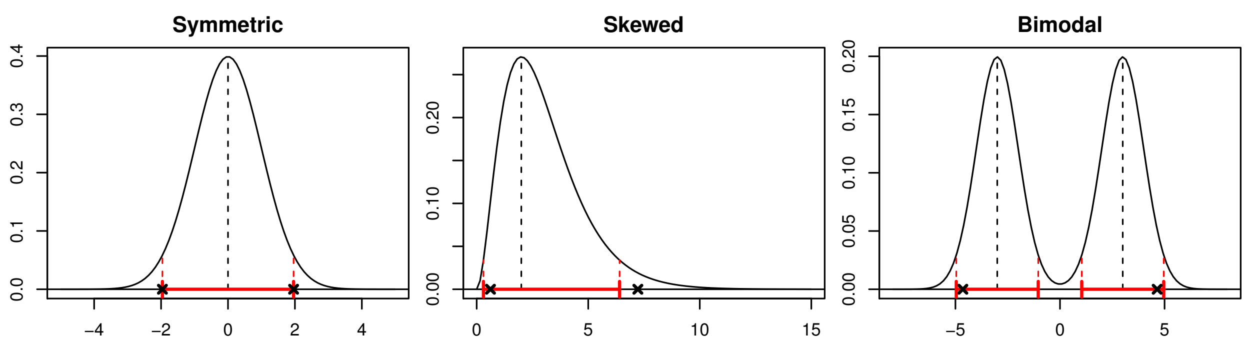

FIG. A1: Probabilistic 95% niche region for various univariate distributions. These are the intervals or collection of intervals highlighted in red. For symmetric distributions, these coincide with the 2.5% and 97.5% quantiles which are depicted with crosses. In each case the red interval(s) contain(s) the smallest region covering 95% of the probability distribution.

Of particular interest here is the case where $X = (X_1, \ldots, X_n) \sim \mathcal{N}(\mu, \Sigma)$ is an n-dimensional multivariate Normal isotope distribution. The joint pdf of X is

$$f(x) = (2\pi)^{-n/2}\,|\Sigma|^{-1/2}\exp\left\{-\tfrac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right\},$$

which is a constant in x for fixed $Q(x) = (x-\mu)'\Sigma^{-1}(x-\mu) = q$. Since increasing $Q(x)$ results in smaller values of $f(x)$, the niche region for Normal data is

$$\mathrm{N_R} = \left\{x : (x-\mu)'\Sigma^{-1}(x-\mu) < C\right\}. \tag{A.2}$$

Upon rearranging terms, $\mathrm{N_R}$ can be written as the quadratic inequality

$$\mathrm{N_R} = \left\{x = (x_1,\ldots,x_n) : \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_ix_j + \sum_{i=1}^{n} b_ix_i + d < C\right\}, \tag{A.3}$$

although the canonical form in A.2 is computationally simpler to use for the niche overlap calculation described in Appendix B.

In order to determine the value of C in A.2, we may utilize the fact that

$$(X-\mu)'\Sigma^{-1}(X-\mu) = Z'Z = \sum_{i=1}^{n} Z_i^2,$$

where $Z = (Z_1,\ldots,Z_n)' = \Sigma^{-1/2}(X-\mu) \sim \mathcal{N}(0,I)$. In other words, $Q(X)$ is the sum of the squares of n iid $\mathcal{N}(0,1)$ random variables, such that

$$Q(X) = \sum_{i=1}^{n} Z_i^2 \sim \chi^2_{(n)}.$$

Thus, for a 95% niche region we require that $P(Q(x) < C) = .95$, a value which can be determined by the `R` command `qchisq(p = 0.95, df = n)`.

References.

Barnard, J., McCulloch, R., and Meng, X.L. (2000). Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica*, 10(4): 1281–1312.

Box, G.E. and Tiao, G. (1973). *Bayesian Inference in Statistical Analysis*. Addison-Wesley, Reading, MA.

Chen, M.H. and Shao, Q.M. (1999). Monte Carlo estimation of Bayesian credible and HPD intervals. *Journal of Computational and Graphical Statistics*, 8(1): 69–92.

Gelman, A. and Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press, New York.

Huang, A. and Wand, M.P. (2013). Simple marginally noninformative prior distributions for covariance matrices. *Bayesian Analysis*, 8(2): 439–452.

Leonard, T. and Hsu, J.S.J. (1992). Bayesian inference for a covariance matrix. *The Annals of Statistics*, 20(4): 1669–1696.

Liechty, J.C., Liechty, M.W., and Müller, P. (2004). Bayesian correlation estimation. *Biometrika*, 91(1): 1–14.

Mardia, K.V., Kent, J.T., and Bibby, J.M. (1979). *Multivariate Analysis*. Probability and mathematical statistics. Academic Press, London.

OMalley, A.J. and Zaslavsky, A.M. (2008). Domain-level covariance analysis for multilevel survey data with structured nonresponse. *Journal of the American Statistical Association*, 103(484): 1405–1418.

Wei, G.C.G. and Tanner, M.A. (1990). Calculating the content and boundary of the highest posterior density region via data augmentation. *Biometrika*, 77(3): 649–652.