Calculating candidate probabilities in Stochastic HGs with normal noise

The probability of candidate *a* being selected is the probability that it has higher harmony than all other candidates. Considering first a single pair of candidates *a* and *b*, candidate *a* has higher harmony than candidate *b* if $\epsilon_b - \epsilon_a < h_a - h_b$, so the probability that candidate *a* has higher harmony than candidate *b* is the probability that the random variable $\epsilon_b - \epsilon_a < h_a - h_b$. For candidate *a* to be optimal it must have higher harmony than all other candidates, so the probability of selecting candidate *a* is the probability that $\epsilon_b - \epsilon_a < h_a - h_b$ for all *b* not equal to *a*.

For a tableau with *N* candidates, there are *N*-1 of these $\epsilon_b - \epsilon_a$ noise difference variables. To calculate the probability of candidate *a* winning, we have to consider the joint distribution of the *N*-1 $\epsilon_b - \epsilon_a$ difference variables because they are not independent since they all include ϵ_a.

If the noise terms follow a normal distribution, as in normal MaxEnt or NHG, then the $\epsilon_b - \epsilon_a$ are also normally distributed, so their joint distribution is a multivariate normal distribution. A multivariate normal distribution specifies the distribution of a vector of random variables, and is specified in terms of a vector of the means of these variables and a covariance matrix, $\mathbf{\Sigma}$, where $\mathbf{\Sigma}_{ij}$ is the covariance between the *i*th and *j*th random variables*,* Cov(*i,j*). In both normal MaxEnt and NHG, the mean of the noise terms is 0, but the derivation of the covariance matrix differs between the two frameworks.

1. Normal MaxEnt

In normal MaxEnt grammar, the ϵ_i are independent normal random variables with mean 0 and variance σ^2. Covariance of $\epsilon_b - \epsilon_a$ and $\epsilon_c - \epsilon_a$ is given by (1) (e.g. Evans & Rosenthal 2009:153f.).

(1) $\mathrm{Cov}(\epsilon_b - \epsilon_a, \epsilon_c - \epsilon_a) = \mathrm{Cov}(\epsilon_b, \epsilon_c) - \mathrm{Cov}(\epsilon_a, \epsilon_b) - \mathrm{Cov}(\epsilon_a, \epsilon_c) + \mathrm{Cov}(\epsilon_a, \epsilon_a)$

The covariance of a variable with itself is its variance, so $\mathrm{Cov}(\epsilon_a, \epsilon_a) = \sigma^2$, and the covariance of independent variables is 0, so where *b* and *c* are distinct, (1) reduces to , $0 - 0 - 0 + \sigma^2 = \sigma^2$, and where *b* = *c*, it reduces to $\sigma^2 - 0 - 0 + \sigma^2 = 2\sigma^2$. So the covariance matrix, $\mathbf{\Sigma}$, of the distribution of noise differences is an $N - 1 \times N - 1$ matrix of the form shown in (2), with $2\sigma^2$ on the leading diagonal and σ^2 in all other positions.

Given the mean and covariance matrix of the multivariate normal distribution of $\epsilon_b - \epsilon_a$ variables, we can use numerical integration algorithms to calculate the probability of each of these variables simultaneously falling below the corresponding harmony difference $h_a - h_b$ (Genz 2009).

(2) $\begin{pmatrix} 2\sigma^2 & \cdots & \sigma^2 \\ \vdots & \ddots & \vdots \\ \sigma^2 & \cdots & 2\sigma^2 \end{pmatrix}$

For example, to calculate the probability of candidate (a) in tableau (3) (repeated from tableau (1) in the paper), we need to determine the probability that $\epsilon_b - \epsilon_a < h_a - h_b$ and $\epsilon_c - \epsilon_a < h_a - h_c$, where $h_a - h_b$ and $h_a - h_c$ both equal 1.

(3)

weights:	15	8	8		Normal MaxEnt	NHG
/input/	C_1	C_2	C_3	h_i	P_i	P_i
a	-1			-15	0.634	0.6
b		-2		-16	0.183	0.26
c		-1	-1	-16	0.183	0.14

If the variance of the noise terms, σ^2, is 1, then the covariance matrix of the joint distribution of $\epsilon_b - \epsilon_a$ and $\epsilon_c - \epsilon_a$ is as shown in (4), and using the function, `pmvnorm` from the R package `mvtnorm` (Genz et al 2020), we can calculate that the probability of candidate (a) winning (i.e. the probability of both difference variables being less than 1) is 0.634. Using the same method we can calculate that the probabilities of candidates (b) and (c) are both 0.183.

(4) $\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$

These probabilities do not match the MaxEnt probabilities given the same constraint weights because the variance of the standard logistic distribution that results from taking the difference between two standard Gumbel random variables is $\pi^2/3$ or about 3.3, which is higher than the variance of the difference between two standard normal random variables. So to match the MaxEnt probabilities, we need to increase the variance of the normal noise terms. A good match is obtained with $\sigma^2 = 1.54$.

2. NHG

The same basic line of reasoning applies to calculation of candidate probabilities in NHG, but calculation of the covariance matrix of the distribution of the difference variables $\epsilon_b - \epsilon_a$ is a bit more involved because the ϵ_i are not identically distributed, instead each is a linear function of independent normal random variables, n_k, each with variance σ^2. Specifically, the noise added to the harmony of candidate *a*, ϵ_a, is given by the expression in (5), where c_{ak} is the violation assigned to candidate *a* by constraint *k*, and n_k is the random variable that is added to the weight of constraint *k*.

(5)

$$\epsilon_a = \sum_k c_{ak} n_k$$

So the difference between noise terms of two candidates, $\epsilon_b - \epsilon_a$, is given by (6).

(6)

$$\epsilon_b - \epsilon_a = \sum_k c_{bk} n_k - \sum_k c_{ak} n_k = \sum_k (c_{bk} - c_{ak}) n_k$$

The covariance of two linear functions of random variables is the sum of the covariances of the terms that are added together in each linear function, so the covariance of two noise difference variables $\epsilon_b - \epsilon_a$ and $\epsilon_c - \epsilon_a$ is:

(7)
$$\begin{aligned}&\mathrm{Cov}(\epsilon_b - \epsilon_a, \epsilon_c - \epsilon_a)\\ &= \mathrm{Cov}\left(\sum_k (c_{bk} - c_{ak})n_k, \sum_{k'} (c_{ck'} - c_{ak'})n_{k'}\right)\\ &= \sum_k \sum_{k'} (c_{bk} - c_{ak})(c_{ck'} - c_{ak'}) \,\mathrm{Cov}(n_k, n_{k'})\end{aligned}$$

Where k and k' range over the constraints. The covariance of two distinct noise variables, $\mathrm{Cov}(n_k, n_{k'})$, is 0 because the n_k are independent, while the covariance of a variable with itself, $\mathrm{Cov}(n_k, n_k)$, is the variance of n_k, i.e. σ^2. So all of the terms involving the covariance of distinct noise variables are 0, leaving (8).

(8)
$$\mathrm{Cov}(\epsilon_b - \epsilon_a, \epsilon_c - \epsilon_a) = \sigma^2 \sum_k (c_{bk} - c_{ak})(c_{ck} - c_{ak})$$

The sum in (8) can be rewritten as the dot product of two vectors as in (9), where $\mathbf{c_a}$ is the vector of violations of candidate *a*, so $\mathbf{c_a}$ in tableau (3) is (-1, 0, 0).

(9)
$$\mathrm{Cov}(\epsilon_b - \epsilon_a, \epsilon_c - \epsilon_a) = \sigma^2(\mathbf{c_b} - \mathbf{c_a}) \cdot (\mathbf{c_c} - \mathbf{c_a})$$

As a result, the covariance matrix for calculating the probability of candidate *a* can be calculated as in (10), where $\mathbf{C}$ is a matrix with one row for each candidate other than *a*, and row $\mathbf{C}_i$ is $\mathbf{c_i} - \mathbf{c_a}$, i.e. a matrix of differences in constraint violations between candidate *a* and the other candidates. Multiplying this matrix by its transpose is equivalent to calculating the dot product in (9) for each pair of vectors, and placing the covariance of the noise differences represented by rows *i* and *j* of $\mathbf{C}$ at $\mathbf{\Sigma}_{ij}$ in the covariance matrix.

(10)
$$\mathbf{\Sigma} = \sigma^2 \mathbf{C}\mathbf{C}^{\mathrm{T}}$$

Given this covariance matrix, the probability that each noise difference $\epsilon_b - \epsilon_a$ is smaller than the corresponding difference in harmonies, $h_a - h_b$ can be calculated exactly as with normal MaxEnt.

For example, to calculate the probability of candidate (a) in tableau (3), the matrix $\mathbf{C}$ is given in (11).

(11)
$$\mathbf{C} = \begin{pmatrix} -1 & 2 & 0 \\ -1 & 1 & 1 \end{pmatrix}$$

This yields the covariance matrix in (12). If $\sigma^2 = 1$, then the probability of selecting (a) comes out as 0.6.

(12) $\Sigma = \sigma^2 \begin{pmatrix} 5 & 3 \\ 3 & 3 \end{pmatrix}$

Note that candidates with the same harmony can have different probabilities in NHG, as demonstrated by candidates (b) and (c) in (3). This is because the covariance matrix of the relevant multivariate normal distribution depends on the matrix of violation differences, **C**, which can be different for each candidate. In normal MaxEnt, the covariance matrix is the same for every candidate in a tableau, so candidates with the same harmony are assigned the same probability.

R code implementing these methods for calculating candidate probabilities in normal MaxEnt and NHG is also included in the file 'schwa models.R' in the supplementary materials.

References:

Evans, Michael J. & Rosenthall, Jeffrey S. 2009. *Probability and Statistics: The Science of Uncertainty, 2nd Edition*. New York: W.H. Freeman. URL http://www.utstat.toronto.edu/mikevans/jeffrosenthal/

Genz, Alan & Bretz, Frank. 2009. *Computation of Multivariate Normal and t Probabilities.* Lecture Notes in Statistics. Heidelberg: Springer-Verlag.

Genz, Alan & Bretz, Frank & Miwa, Tetsuhisa & Mi, Xuefei & Leisch, Friedrich & Scheipl, Fabian & Hothorn, Torsten. 2020. *mvtnorm: Multivariate Normal and t Distributions*. R package version 1.1-1. https://CRAN.R-project.org/package=mvtnorm.