Stochastic Harmonic Grammars as Random Utility Models

Edward Flemming (flemming@mit.edu)
Massachusetts Institute of Technology

Introduction

• Stochastic grammars assign probabilities to outputs, making it possible to analyze variation and gradient acceptability in phonology.
• ‘Maximum Entropy’ Grammar is a form of stochastic grammar that is widely used in Harmonic Grammar (Legendre et al. 2006) rather than classical Optimality Theory.
• But MaxEnt grammar is not the only proposal for ‘stochasticizing’ Harmonic Grammar – an alternative is Noisy Harmonic Grammar (Boersma & Pater 2016).
• Identify a uniform framework for comparing and analyzing Stochastic Harmonic Grammars: Random Utility Models.
• Use it to draw out similarities and differences between MaxEnt and NHG.

Shape of the noise distribution

• Noise terms (εj) follow a Gumbel distribution in MaxEnt and a normal distribution in NHG (normal + normal ⇒ normal).
• This is not an important difference because independent Gumbel εj’s are essentially a tractable approximation to independent normal εj’s.
• The probability of a candidate having the highest harmony depends on the difference in harmony between it and competing candidates.

MaxEnt is simple

• In MaxEnt, probability is directly related to harmony: Pj = eβhj / Σk eβhk, where hj is the harmonic value of candidate j.
• The relative probabilities of two candidates are independent of the rest of the candidate set because εj’s are independent.
• In choice models this property is referred to as ‘Independence from Irrelevant Alternatives’ (Train 2009:45ff.).

NHG is complicated

• In NHG, εj’s are not independent so the relationship between harmony and probability is complex, and the relative probability of pairs of candidates can depend on other candidates.
• The same difference in harmony translates into different relative probabilities, depending on how many constraint violations are shared – ‘partial harmonic bounding’.

NHG and MaxEnt as Random Utility Models

• Although NHG and MaxEnt are superficially very different they can both be formulated as NHGs where the harmony of candidate i is hi + εi, where εi is a random variable (‘noise’).
• HG is made stochastic by adding noise to harmony.
• Referred to in economics as a Random Utility Model (Train 2009).

In NHG noise is added to the constraint weights, but the resulting harmony expression can be separated into hi + εi.

Independence of εi’s and harmonic bounding

• Important differences:
  – In MaxEnt, εi’s are independent and all drawn from the same distribution.
  – In NHG, εi’s are not independent – candidates that violate constraint k share a noise component nk – and are drawn from distributions with different variances.

Should harmonically bounded candidates be assigned P = 0?

• Assigning probability to bounded candidates is central to the MaxEnt analysis of local optimality (Hayes 2017), and to ‘markedness only’ analyses of gradient phonotactics (Hayes & Wilson 2008).
• The NHG mechanism for assigning zero probability to harmonically bounded candidates has additional effects – ‘partial harmonic bounding’.

Conclusions

• In spite of superficial differences, MaxEnt is actually a variant of NHG.
• Essentially the simplest form of NHG.
• In the absence of empirical evidence in favor of NHG, MaxEnt is to be preferred for its simplicity and tractability.