Group 21: Cognitive Modeling - Lab 3

Olusanmi Hundogan - 6883273
Evangelia Giannikou - 6988229

April 20, 2021

Question 1

Height: mean = 180, st.d = 10

Show two graphs: (i) what the ***probability*** *of using threshold looks like on the scale 1-250 cm; (ii) the graph of the function σ on the same scale. State which degree point has the highest probability of being used as a threshold, and on which degree point it is most likely the speaker will use the adjective tall. If the two values differ or are the same, say in a few words why you think this is so.*

In Figure 1, the **probability** of using threshold on the scale 1-250 cm is shown. The highest degree point is when $P(\Theta; \lambda; c) = 0.083$, and $\theta = 184$ cm.

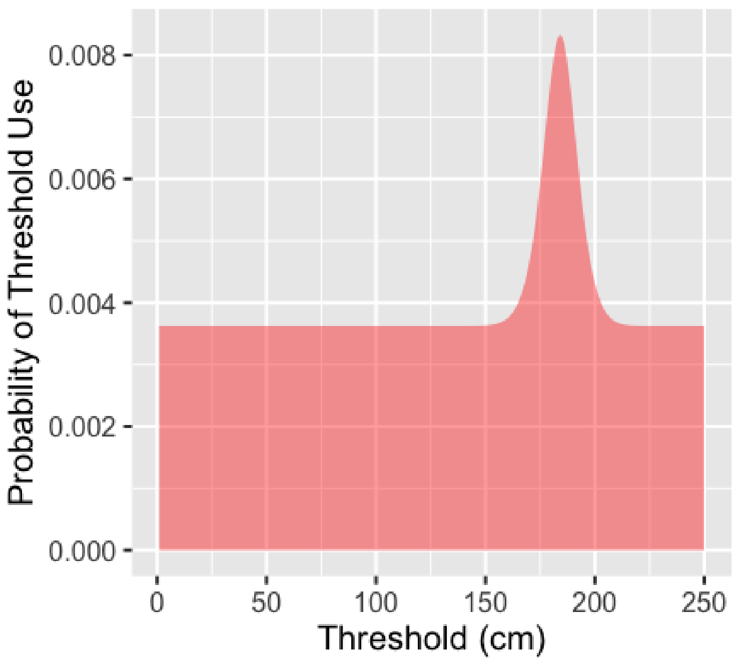

Figure 1: The graph shows the probability of threshold use on a scale of 1-250 cm.

Similarly, in Figure 2, it is shown the graph of the function σ on the same x-scale. The highest degree point is 1 and height = 250 cm. 250 yields the highest likelihood of adjective use. At this threshold, it is certain that if the speaker utters "tall" he will be referring to 250 and the listener will understand it as such. As the value decreases, the likelihood is diminishing as well.

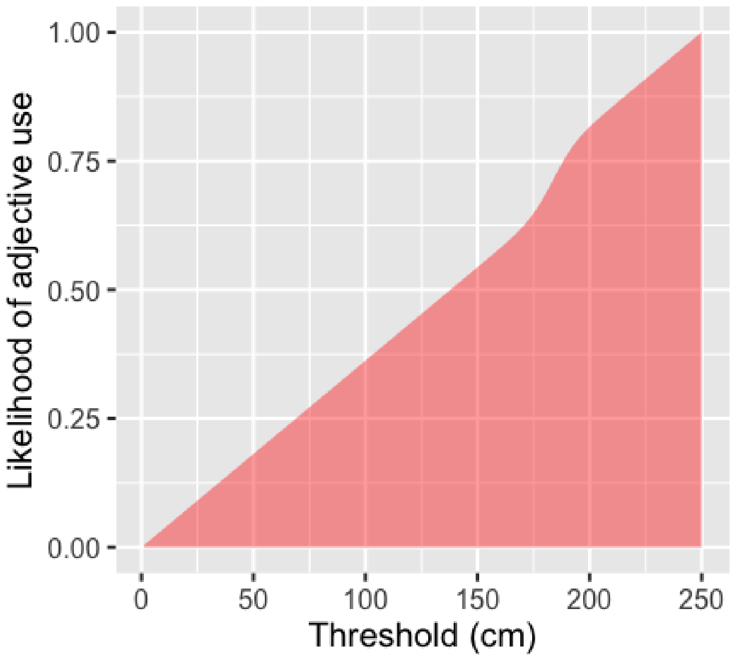

Figure 2: The graph shows the function σ on a scale 1-250 cm.

Comparing the two degree points to evaluate the communication efficiency, it is evident that the two values deviate. Considering that $\theta = 250$ is not efficient because it does not apply to a large population of people, the speaker is most likely to use 184 cm as threshold. For other values of θ, someone described as "tall" would probably make it inefficient to use.

Question 2

IQ - normal distribution

Specify the normal distribution and generate figures for ES and σ function using this distribution and report which degree has the highest ES.

For specifying the IQ distribution, we use $IQ \sim \mathcal{N}(100, 15)$ since $mean = 100$, given from the description, and with 95% confidence interval $st.d. = 15$.

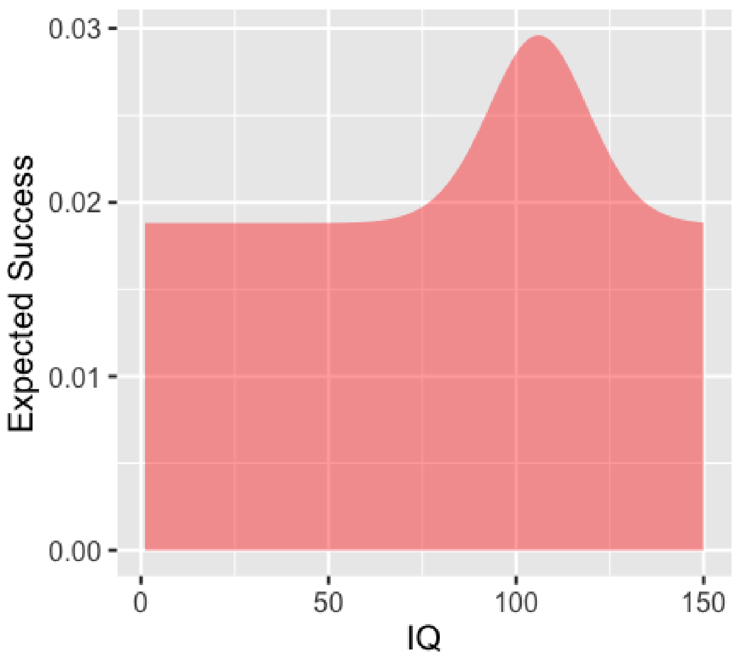

Figure 3: The graph shows the Expected Success function for IQ on a scale of 1-150.

The highest expected success $ES = 0.029$ in Figure 3 is at $IQ = 106$.

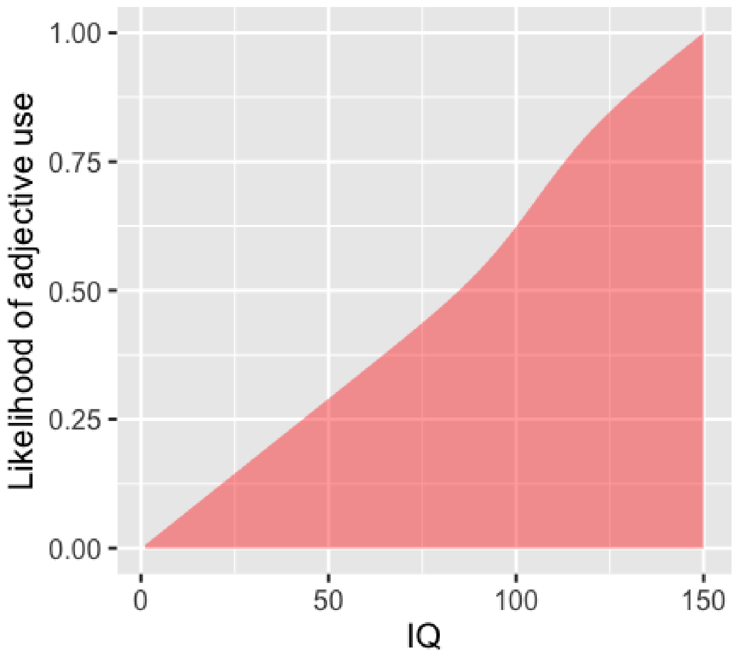

Figure 4: The graph shows the σ function for IQ on a scale of 1-150.

The highest likelihood of adjective use $\sigma(A|d;\lambda;c) = 1$ in Figure 4 is at $IQ = 150$.

Waiting times - gamma distribution

Specify the gamma distribution and generate figures for ES and σ. Report which degree has the highest ES.

For specifying the waiting times distribution, we use Waiting times $\sim \Gamma(2,1)$. These values were estimated by using $\frac{\mu^2}{\sigma^2}$ and $\frac{\sigma^2}{\mu}$ for the shape and scale parameters, respectively.

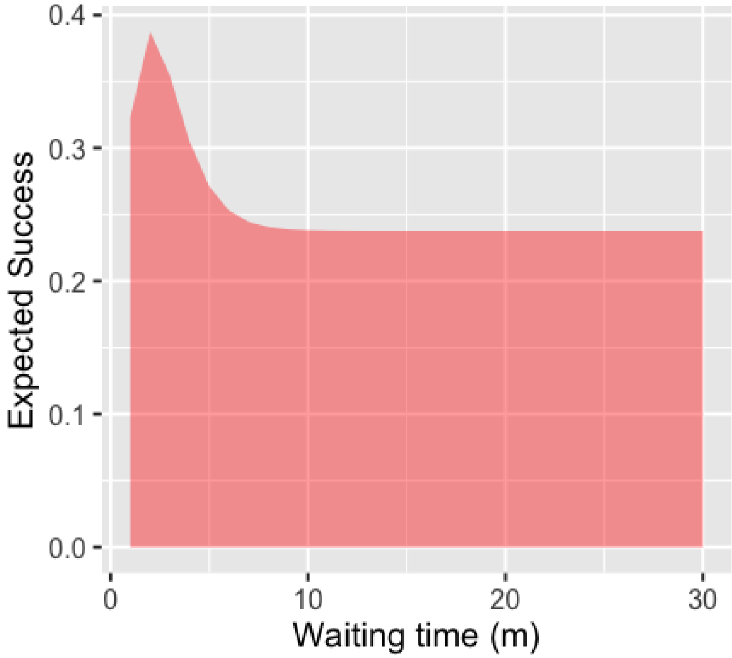

Figure 5: The graph shows the Expected Success function for waiting on a scale of 1-30.

The highest expected success $ES = 0.387$ in Figure 5 is at waiting time $= 2$ minutes.

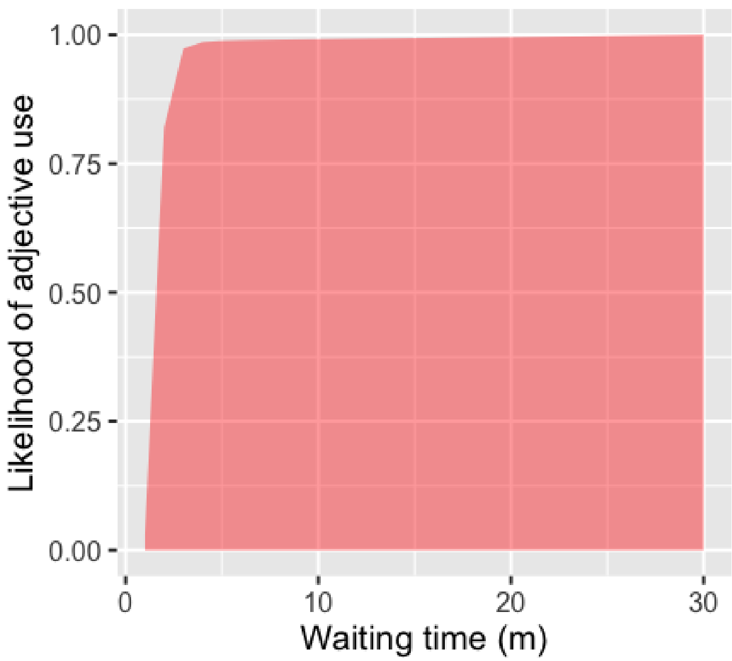

Figure 6: The graph shows the σ function for waiting on a scale of 1-30.

The highest likelihood of adjective use $\sigma(A|d;\lambda;c) = 1$ in Figure 6 is at waiting time $= 30$ minutes.

Question 3

What is Pearson's correlation coefficient, r, between model's predictions and actual observations. Check the value of r with respect to three different parameters of the model (9 comparisons in total):

- *lambda=40, coverage.parameter=0.1*
- *lambda=40, coverage.parameter=-0.1*
- *lambda=40, coverage.parameter=0*

What coverage parameter gives us the best and the worst linear correlation? On which distribution do we get the best results?

Afterward, pick the coverage parameter that worked best and select two distributions by flipping their prior belief distributions. Report what prior distribution you used and report r. Do we see that r is affected? Does the model suffer in having worse linear correlation with respect to the observed data? Say in a few words why this is the case.

As it is visible in Table 1, the best linear correlation is given when the coverage parameter is -0.1 in which the r value is on average $r = 0.97$. The worst correlation is given when coverage parameter is 0.1 in which the average correlation values is $r = 0.86$.

Similarly, when comparing the results between the distribution, the best correlation is that of the Left-skewed with an average of $r = 0.94$. The worst correlation is given for the Moved distribution in which the on average correlation is $r = 0.91$.

Model	r when $c = 0.1$	r when $c = -0.1$	r when $c = 0$
Gaussian	0.88	0.97	0.94
Left - skewed	0.85	0.98	0.98
Moved	0.84	0.97	0.93

Table 1: Pearson's correlation between model's prediction and actual observation for 3 different model distributions where $\lambda = 40$ and coverage parameter is 0.1, -0.1 and 0.

For the second part of the question, we picked $c = -0.1$ as the best coverage parameter and selected to flip the distribution of the Left-skewed and moved models. In particular, we used $\mathcal{N}(6, 2)$ for the Left-skewed model. The correlation we found was $r = 0.95$. For the moved model we used $\Gamma(4, 1.5)$. The correlation found was $r = 0.86$.

A sum of the results in regards to the previously observed data is found on Table 2. In the case of the Left-skewed model, the correlation worsens slightly with a Gaussian prior distribution set. This decrease was expected since Gaussian distribution would not work well with data that are concentrated on the right side of the distribution. Given that the data distribution was generated under a different distribution, it makes sense that a differing prior distribution will be biased towards that prior distribution. Hence, the model will have difficulties adjusting to the data, while keeping the prior in mind. Similar holds for the moved distribution as it's underlying prior is Gaussian and a Gamma distribution would not fit well.

Model	r with Normal distribution	r with Gamma distribution
Left - skewed	0.95*	0.98
Moved	0.97	0.86*

Table 2: Pearson's correlation between 2 models with $c = -0.1$. R values with * indicate the new correlation values with flipped distribution.

Question 4

Bayesian model - What values have been found for lambda and coverage.parameter? Report summary statistics on lambda and coverage.parameter (MAP, median, 2.5% to 97.5%). Discuss briefly the summaries (i.e., are values in out widespread or very narrow? If so, why do you think this is the case and is it good/bad?).

For the Bayesian model, we used as a prior $\mathcal{N}(6, 2)$. As seen in Table 3, the values of 2.5% and 97.5% interval indicate the limits of c and λ. Hence, there is a 95% chance that $c \in [-0.332, -0.107]$ and $\lambda \in [18.483, 48.985]$. The MAP values are $c = -0.197$ and $\lambda = 33.364$. These values are very close to the ones we found in Question 3 ($\lambda = 40$, $c = -0.1$) for the highest correlation value given a Gaussian distribution. These estimates are reasonable in the context of the experiment as a negative coverage parameter hints that the gradable adjective is not generally applicable [1]. Intuitively, this holds for an adjective like 'big'.

In order to judge the estimates we have to look at the shape of the posterior distribution for each estimator, as seen in Figure 7. Both parameter densities are quite widespread. This tells us that the model's uncertainty about the parameter is quite large. Although the c parameter is slightly narrower. Ideally, we want narrow densities which represent high certainties about the estimates.

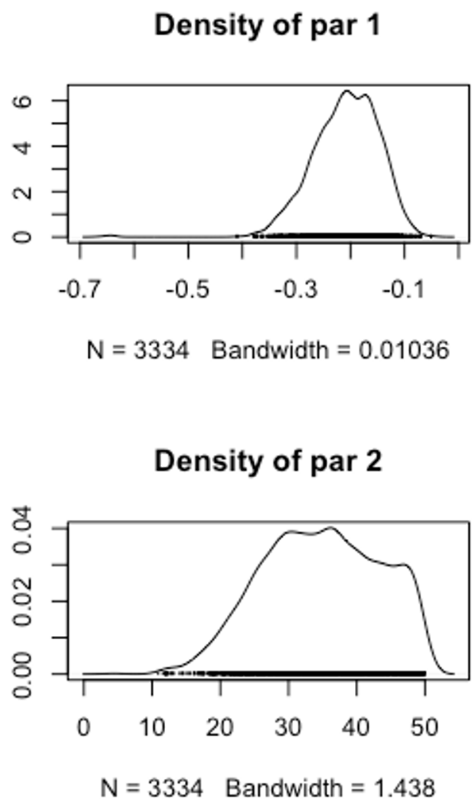

Figure 7: The density graph of par 1 (c) and par 2 (λ) where their values are widespread.

	MAP	median	2.5% interval	97.5% interval
c	-0.197	-0.204	-0.332	-0.107
λ	33.364	34.919	18.483	48.985

Table 3: A statistics summary of c and λ for adjective='big' on a Gaussian distribution.

Question 5

Expand your model to construct the posterior distribution of the two parameters using three adjectives (big, pointy, tall) in all three distributions. Explore the distribution of the parameters in out. *Plot the summary and report summary statistics on lambda and coverage.parameter.*

Similar to Question 4, we used as a prior $\mathcal{N}(6, 2)$ for all three adjectives 'big','pointy' and 'tall'. Again, there is a 95% chance that $c \in [-0.045, -0.016]$ and $\lambda \in [25.764, 39.160]$. The MAP values are $c = -0.031$ and $\lambda = 30.788$ which are further away from $c = -0.01$ and $\lambda = 40$. The increased deviation is reasonable as we now expand the model to include three adjectives. Beforehand, the model only focused on the adjective 'big'.

	MAP	median	2.5% interval	97.5% interval
c	-0.031	-0.031	-0.045	-0.016
λ	30.788	31.269	25.764	39.160

Table 4: A statistics summary of c and λ for adjectives = 'big', 'pointy', 'tall' for 3 distributions.

Bonus question

We specified the likelihood using dnorm with sd = 0.1*. Is this sensible or not? Check what dnorm is and state what problems there might be with this particular function and with sd* = 0.1 *for our data. Is there an alternative probability distribution that would make more sense?*

The dnorm function returns a probability/density for a given value, a mean and a st.d. In our case the mean is determined by a data point in the experimental data set. Hence, the simulation collects the sum of each degrees' percentage value for an adjective given their likelihood under their respective likelihood of adjective use.

The standard deviation determines how much samples "disperses" from the mean. A low value of st.d., indicates that the data points tend to be close to the mean. A st.d 0.1 seems reasonable as the likelihoods of adjective use will remain within the bounds of the 0 and 1. Increasing the standard deviation increases the range of "highly" probable deviations. Reducing the st.d might not capture the unexplained variations within the data, that are necessary for a "faithful" simulation. Determining the correct standard distribution is therefore difficult. As a result, choosing the normal distribution may not be the best approach to model the simulation. Two other distributions may be more fitting. In the next two sections, we will describe both briefly.

With most of the true data being Left-skewed it could make more sense to model the generated data as random samples from a Left-skewed distribution. In a Left-skewed distribution, small data values are gathered on the left, whereas higher values are concentrated to the right. This pattern seemingly also emerges from the likelihood of the adjective use σ. This can be seen in Figure 8. Results from Question 3 reinforce this notion as the Left-skewed gamma prior yields the best correlation values. Hence, we can model the simulation with the gamma distribution. This distribution is parametrized as follows: $\Gamma(x|\alpha, \beta)$. Here, the shape α and scale β parameters can be derived from knowing the mean and variance of the experimental data. However, this modelling approach purely relies on structural properties of the data. As we know the experimental conduct[1] we can construct a much more pronounced and realistic simulation.

As the percentage values where computed by taking the rate of how many participants made a binary judgement for a degree, the binomial distribution naturally comes into mind. The binomial distribution models repeated Bernoulli experiments (stochastic binary events) and belongs to the family of discrete distributions. It is parametrized by three factors $\mathcal{B}(k|p,n)$. We can choose p to be the likelihood of adjective use σ and n as total number of items for a particular degree d_i. With both values, we can conduct the simulation by collecting the sum of each degrees' percentage value, again. This approach follows the parameter estimation approach of Qing and Franke.

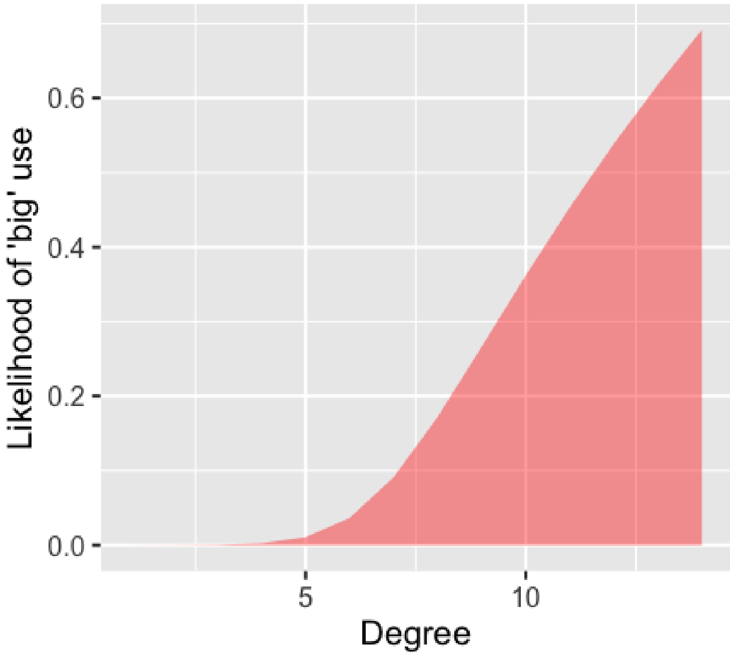

Figure 8: The graph shows the σ function for adjective 'big' on a Left-skewed distribution.

References

[1] Ciyang Qing and Michael Franke. "Meaning and Use of Gradable Adjectives: Formal Modeling Meets Empirical Data". In: *CogSci*. Proceedings of the Annual Meeting of the Cognitive Science Society. Vol. 36. 36. 2014. ISBN: 1069-7977.