DS 6014 Final Project: Estimating the Posterior Probability of Refugee Asylum Rejection and its Latent Change Across Time

Youssef Abubaker (yaa2vb); Joseph Cho (jc4jx); Andrew Graves (ajg3eh); Gabe Yohe (gjy7kb)

12/8/2020

1 Problem Description

By the end of 2019, 79.5 million people had been displaced due to war, persecution, famine etc [1]. These 79.5 million people are refugees. Refugees escape suffering by applying for asylum in different countries to begin a new life far away from their homes [2]. Refugees are not located equally around the world. Certain countries are seen as more desirable locations of asylum than others, such as North America and Western Europe. However, the more desirable locations do not necessarily accept the most refugees. Certain countries accept refugees with higher probability than others. In other words, more desirable countries are also more selective. Based on the UNHCR data the most desirable locations are applied to by a vast majority of refugees regardless of origin.

Let us consider an example scenario. Refugee A originates from Syria and seeks asylum in Germany and Egypt. Germany is considered "desirable" based on application statistics, while Egypt is not as sought after. Refugee A receives asylum in Egypt but not in Germany. This could occur because Egypt and Syria share cultural ties while Germany and Syria have very poor relations, or possibly because Germany is just more selective. A plausible mechanism for this phenomenon is that political and cultural relationships between the country of asylum and the country of origin effects whether or not a refugee is granted asylum. If refugees from Syria are rejected at higher rates than refugees from China even though both are applying for asylum in the most desirable locations, then it could be useful for the UN and other agencies to have access to principled and calibrated estimates of these probabilities.

We can build a Bayesian model that incorporates the difference between probabilities of rejection for refugees based on their country of origin AND their country of asylum, so we can understand whether there is a systematic rejection of refugees by the countries deemed most desirable. We also can model the latent rate of change for rejection of refugees in both country of origin and asylum to analyze the trajectory in that country over all the available data in our source (years 2000-2019). This can provide insight into the population and political dynamics that have unfolded over the 21^{st} century with respect to asylum.

2 Mathematical linkage between problem and methods

In order to model the probabilities of rejections for refugees and the rate of change across time for these rejection rates, we built a Bayesian hierarchical model to manage the uncertainty for these probabilities with respect to both country of origin and country of asylum. Let i be a "from-to" country relationship observation for a given year, j the country of origin, k the country of asylum, and y the rejections rates weighted by the number of asylum applications for that "from-to" relationship mapped to the real line with a logit transformation. This model can be represented by the following equation

$$y_i = \alpha_{origin_j[i]} + \alpha_{asylum_k[i]} + \theta_{origin_j[i]}\text{year} + \theta_{asylum_k[i]}\text{year} + \epsilon_i$$

where α and θ represent random intercept and slope parameters respectively. We can enrich this mathematical expression with the relevant weakly informative prior distributions by representing this model graphically as follows

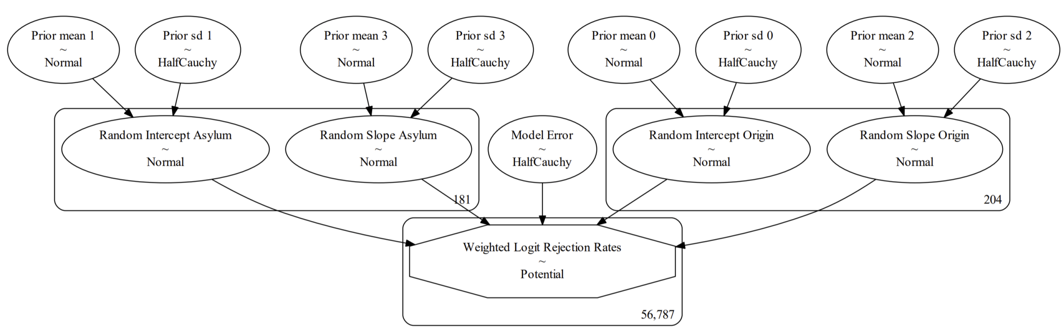

Figure 1: Graphical representation of the hierarchichal model

The graph in Figure 1 incorporates uncertainty around the probability of rejection and its fluctuations throughout time as a function of country of origin and country of asylum by modeling the *Potential* as a weighted Gaussian likelihood function.

3 Bayesian methodology and optimization approach

Our data was obtained from the United Nations High Commissioner for Refugees (UNHCR) database. Using the "Data Finder" on this website, we queried the database to retrieve asylum claims aggregated by year, country of origin, and asylum. Our data includes decisions from 204 and 181 countries of origin and asylum respectively from the years 2000-2019. For this model, we used Automatic Differentiation Variational Inference (ADVI) as an optimization procedure for fitting the hierarchical model. ADVI approximates the latent parameter space with Monte Carlo integration and performs an unconstrained optimization through mapping the latent variables to the set of real numbers. For this variational method, we optimize the evidence lower-bound (ELBO; $\mathcal{F}(q(\mathbf{Z},\Theta))$) component of the total data likelihood ($\ln p(\mathcal{D},\mathbf{Z}|\Theta)$) given the following expression

$$\ln p(\mathcal{D},\mathbf{Z}|\Theta) = \mathcal{F}(q(\mathbf{Z},\Theta)) + \mathbb{KL}(q||p)$$

where (Note: sums are replaced with integrals in the case of continuous variables)

$$\mathcal{F}(q(\mathbf{Z},\Theta)) = \sum_z q(\mathbf{Z}) \ln \left\{ \frac{p(\mathcal{D},\mathbf{Z}|\Theta)}{q(\mathbf{Z})} \right\}$$

$$\mathbb{KL}(q||p) = -\sum_z q(\mathbf{Z}) \ln \left\{ \frac{p(\mathbf{Z}|\mathcal{D})}{q(\mathbf{Z})} \right\}$$

Optimization of the ELBO is a convenient and tractable method for maximizing the total data likelihood. Explicitly minimizing the Kullback-Liebler (KL; $\mathbb{KL}(q||p)$) divergence is typically not as straightforward as explicitly maximizing the ELBO. We use the theory undergirding variational inference, which shows that the KL divergence is minimized as the ELBO is maximized. From a practical perspective, ADVI is preferable over Markov-chain Monte Carlo (MCMC) sampling techniques *for this specific application*, simply because the number of observations and number of parameters in this model are quite large. MCMC methods, while their estimates of the posterior will likely converge to the true distribution given enough time, ADVI provides a good approximate solution in a reasonable amount of time. Figure 2 displays the convergence process of this model.

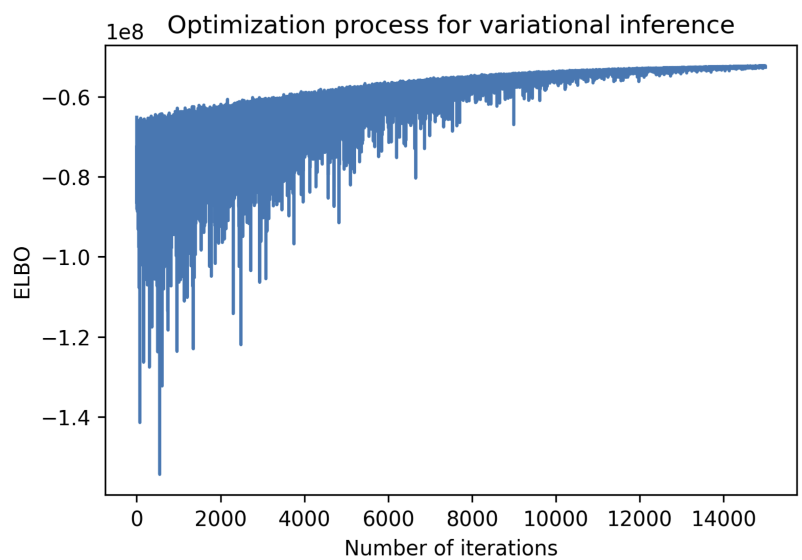

Figure 2: The hierarchichal model converged after 15,000 iterations

4 Results and conclusions

Overall, our results satisfy our goals for this analysis: to estimate probabilities of rejection rates for each country (both origin and asylum location) and its change through time. Since we employed the ADVI method for fitting the model, the credible intervals may be underestimated. Importantly though, the vast majority of our estimated probabilities correlate with conditional means from the raw data, Pearson's $r = 0.869$. Furthermore, the hierarchical model provides a framework for drawing conclusions about the rates of change in rejection and expected rejection percentage per country. Primarily, we were interested in observing the intercept differences across countries as these probabilities served as a medium for comparison. In addition, the random slopes were insightful for observing the attitudinal shifts of countries in our period of analysis. For instance, when modeling on the past ten years, we were able to identify a positive trend in rejection rate in Denmark. This aligned with Denmark's growing strictness on immigration, and our preliminary data analysis which shows the average percentage of rejection in Denmark trending upwards over time (see code for figure) [3]. We were able to estimate the rejection rates and their respective latent rates of change for each country of origin and asylum.

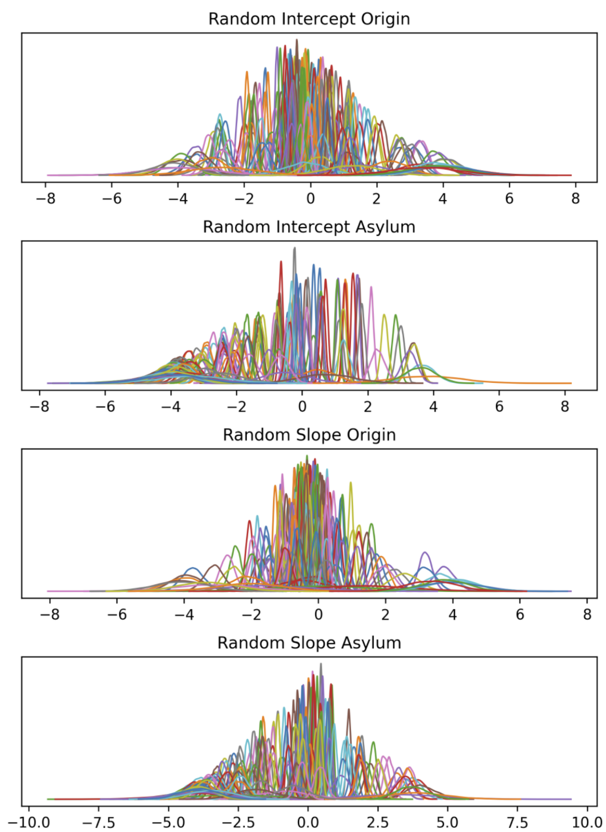

Figure 3: Posteriors for hierarchichal terms of interest. It is clear that there is heterogeneity for both the intercepts and slopes for country of origin and asylum, which suggests that the hierarchichal model is a good approach for managing the uncertainty across countries.

In the future, certain steps could be taken to smooth the fit of this modeling approach by grouping countries into geographic regions and/or binning the time variable by intervals of interest such as US presidential administrations. This would reduce the variability and thus produce more credible confidence intervals. However, there is a trade-off between the smoothness of fit in the model and the loss of granularity in inference. Last, our model provides a good framework for incorporating new information such as demographic or economic variables into our hierarchical model. Unfortunately, our specific data source did not reliably encode these types of features at the level of granularity and completeness required for our modeling procedure. Future research could utilize multiple data sources and join them together to ask additional questions with this type of data. We believe this model is a good starting point for structuring future research on this topic, and that individuals who have specific questions in mind related to this topic could substitute their grouping variable of interest into this analysis to make inference on asylum rejection rates for particular groups.

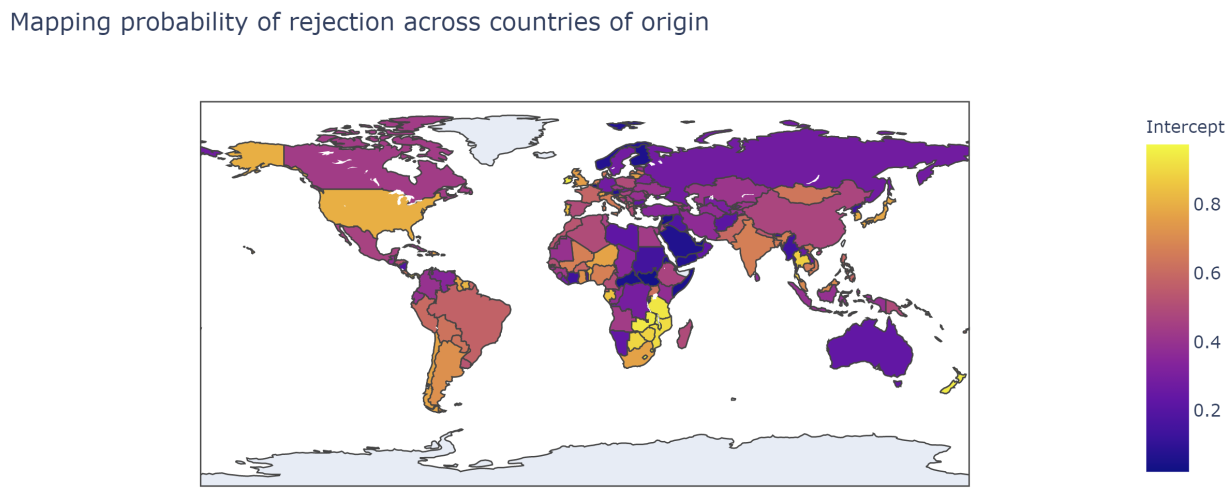

Figure 4: World map of posterior probabilities of rejection across countries of origin (see code for asylum). Refugees from the United States have an estimated rejection probability of 65.3% compared to China with 24.3%. Probability differences could indicate systematic or political factors in the asylum application process.

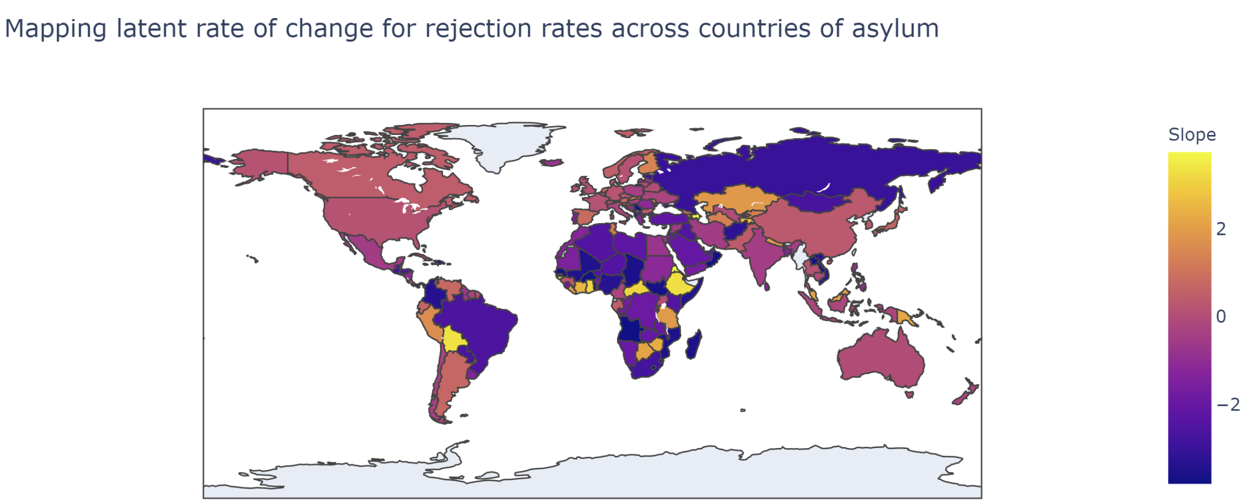

Figure 5: World map of latent rates of change across countries of asylum (see code for origin). We observe positive and negative rates suggesting there was heterogeneity in change across this period (2000-2019). Countries in Western Europe had positive growth in rejection rates, perhaps a response to the numerous terrorist attacks in that region during the latter half of our time period.

5 References

1. Nebehay, S., 2020. Record 79.5 Million Displaced At Year-End Despite COVID Slowing Exodus - UN. [online] U.S. Available at: https://www.reuters.com/article/refugee-day-un-report/record-79-5-million-displaced-at-year-end-despite-covid-slowing-exodus-un-idUSL8N2DU1FQ [Accessed 8 December 2020].

2. Refugees, U., 2020. Figures At A Glance. [online] UNHCR. Available at: https://www.unhcr.org/en-us/figures-at-a-glance.html [Accessed 8 December 2020].

3. Abend, L., 2020. An Island For 'Unwanted' Migrants Is Denmark's Latest Aggressive Anti-Immigrant Policy. [online] Time. Available at: https://time.com/5504331/denmark-migrants-lindholm-island/ [Accessed 8 December 2020].