Introduction to Bayesian Statistics: Practicum 2

As we saw earlier, the 2014 Ebola Outbreak in West Africa had differing mortality rates based on the country of origin (Guinea was 67%, Sierra Leone was 28%, and Liberia was 45%). Let’s use these to construct a distribution representing our overall prior belief of mortality from Ebola in West Africa to be used for future outbreaks in this region.

1. What is a reasonable point estimate of the West African mortality rate from Ebola and how would you describe the uncertainty of that point estimate?
2. Which of the common distributions (of those presented in this workshop) would be appropriate for the point estimate identified in problem 1? Argue in favor of one of them giving specifics for why it would be better than others.
3. Use the roulette method in MATCH to identify a distribution and its parameters that could be used for this prior that reflects a high level of certainty in the point estimate. Repeat this for a prior reflecting a low level of certainty in the point estimate.

The observed difference in mortality rates is likely due to multiple factors such as quality of health care, virulence of the Ebola species, and characteristics of the population that may make them more robust or susceptible to the effects of Ebola infection. One researcher with expertise in infectious disease in West Africa suspects that part of the difference may be due to differing case definitions among the various health care systems of the three countries. Specifically, this researcher believes the lower mortality rate in Sierra Leone is due, in part, to a more liberal case definition which identifies more cases making the infection look more wide spread but less lethal than in other countries. After carefully reviewing the case definitions used for the three countries, the expert feels certain that the mortality rate in Sierra Leone would have been closer to 40%

1. How does this modify the point estimate and uncertainty previously computed in problem 1 above?
2. Use the roulette method in MATCH to identify a prior distribution and its parameters that this researcher might use in a future outbreak (again for both high and low certainty in the point estimate).

As an interesting exercise that will prepare the way for things to come later in this workshop, we’ll point out now that sometimes the uncertainty of a measure (i.e. the variance of a measure) will need to be modeled (this is almost always the case with continuous data). To do so, we will need to specify a prior distribution for the variance. Using the characteristics of the variance, which of the common distributions we have discussed could be appropriate for modeling such a quantity? (Hint: There are a few that would work here.). Based on what you have learned about these distributions, what advantages and disadvantages do foresee in each?

**Note**: The term precision, often denoted as τ, is the inverse of the variance. For example if we denote variance by $ σ^{2}$then the precision is $τ = 1/σ^{2}$. In the United Kingdom (UK), this is a very popular way to talk about uncertainty and will be important later because some of the software we use (developed in the UK) will require us to specify the precision rather than the variance. The thing to remember here is that a large value of variance, or uncertainty, means a small precision value and a small value of variance, or little uncertainty, means a large precision value.