LSA 308 homework #3

July 17, 2007

For this assignment, you will need a "calculator" to query the probability density of the multivariate normal distribution. If you don't know how to do this immediately, I suggest downloading `R` from *http://www.r-project.org/* and installing it on your computer. Then run `install.packages("mvtnorm")` and then `library(mvtnorm)`. You can use the function `dmvnorm()` to query the multivariate normal density. Here's an example of how to use `dmvnorm()` to calculate the likelihood of a single sample point at $(0.5, 0.5)$ for a symmetric multivariate normal distribution centered around $(0, 1)$ with covariance matrix $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$:

```
library(mvtnorm)
> Mean <- c(0,1)
> Sigma <- matrix(c(1,0,0,1),2,2)
> dmvnorm(c(0.5,0.5),Mean,Sigma)
[1] 0.12395
```

`?dmvnorm` or `help(dmvnorm)` will show you more information on how to use this function.

1. Consider a two-letter language with only one-letter words, `a` and `b`. These can be represented in a 2D space at $(1, 0)$ and $(0, 1)$ respectively. Consider the following set of samples $s_{1\cdots10}$:

   (-0.99,-2.50)
   (-4.85,2.30)
   (-0.44,2.27)
   (-1.10,0.96)
   (-0.55,1.18)
   (-0.40,-1.00)
   (-2.53,1.63)
   (0.70,0.58)
   (-2.07,-0.81)
   (0.45,2.09)

   First, a bit of fun: eyeball the data (no calculator yet!) and guessstimate the posterior probabilities $P(\mathtt{a}|s_{1\cdots j})$ and $P(\mathtt{b}|s_{1\cdots j})$. Then, calculate the actual posterior probabilities $P(\mathtt{a}|s_{1\cdots j})$ for all $j = 2, 3, \cdots, 10$. In the Bayesian Reader as Norris defined it, it is not possible to calculate $P(\mathtt{a}|s_1)$. Why not?