Bayesian Statistics

Exercises for lecture 05

1 The logistic function

4 pt(s)

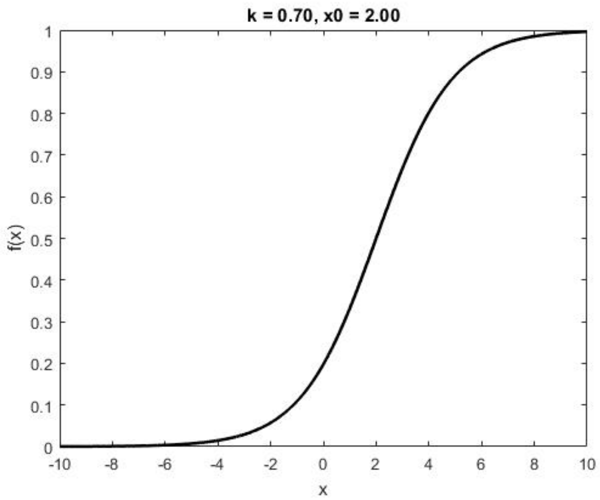

Figure 1: The logistic function.

During the lecture, we saw how an unbounded range of continuous numbers can be transformed to the interval $[0, 1]$ using the cumulative probability density function of the Gaussian distribution. Many other such transformations are possible. For the model of this exercise, you will need to use the *logistic function*, which is defined as

$$f(x; k, x0) = \frac{1}{1 + \exp[-k(x - x_0)]} \quad , \tag{1}$$

with

- x_0 the x-value where $f(x) = 1/2$.
- k the *growth rate* or *steepness* of the curve. See Fig. 1 for an example.

For the Eurovision Song Festival, it is hypothesized that whether a country votes for another country (let's assume each country i can either vote or not vote on each other country j, i.e. $x_{ij} \in \{0, 1\}$) does not actually depend on performance, but on the Euclidean distance d_{ij} between the countries. Importantly, countries *close* to each other, vote for each other, while countries *far* from each other, do not vote for each other.

We want to learn the parameters k and x_0 of the logistic function that describes the relationship between distance and voting probability. We observe the voting results x_{ij} ($i = 1..N, j = 1..N$, you may ignore self-votes) and the xy-coordinates $(c_{i,1}, c_{i,2})$ for each country i.

1. Implement the logistic function and plot a few curves with different k and x_0, for $x \in [-10, 10]$. What range of values are sensible for k and x_0, and what does that say about probability distributions describing them?

2. Write down a generative model and the corresponding graphical model that can learn k, x_0 based on the observed votes and xy-coordinates of countries. You will have to use the logistic function to transform distance into a probability, and you will have to compute distance from the coordinates.

2 Coins from different factories with different certainties

5 pt(s)

In this exercise, you'll implement in R some model comparison code for an example with a conjugate prior.

Recall that the beta distribution can be parameterized in different ways. The first, beta$(\theta|a, b)$, has the interpretation of a pseudo-heads and b pseudo-tails. The second, beta$(\theta|\omega, \kappa)$, has the interpretation of ω as the mode of the distribution (i.e. the tendency; θ will typically be near ω) and κ as the certainty in our belief (i.e. a high κ indicates that all θ's are close to ω, while for a low κ, θ might be off). The two can be rewritten as each other through these identities:

$$
\begin{aligned}
a &= \omega(\kappa - 2) + 1 \\
b &= (1 - \omega)(\kappa - 2) + 1 \ .
\end{aligned}
$$

Suppose we have a coin, which we flip 12 times, observing 8 heads and 4 tails. The coin can come from either of two factories (models, m_1 and m_2). We have no prior preference for either factory, so the prior on factories is uniform ($p(m_1) = p(m_2) = 0.5$). The factory from which the coin comes, determines our prior beliefs in the probability θ of observing heads (i.e., each factory has its own prior on θ).

- Factory 1 has $\omega_1 = 0.25$.
- Factory 2 has $\omega_2 = 0.75$.

1. Assume $\kappa = 8$. What values for θ do the models predict? To answer this, plot $p(\theta|z, N, m_1)$ and $p(\theta|z, N, m_2)$. Given these data, what model do you expect to make most sense (and why)?
2. What are the posterior *model* probabilities of each of the two factories, $p(m_1|z, N)$ and $p(m_2|z, N)$? Look at pages 270 and 271 of the book for a step-by-step explanation of the procedure you need. You need to implement this in R. Show your code as well as the Bayes factor for model 1 vs model 2, and the two posterior model probabilities.
3. What are the two model probabilities if you change κ to 170?
4. Why are your results for the model probabilities so different if you change κ, even though the mode of the beta distribution ω stays the same?

3 Fred & Bart

1 pt(s)

Fred the frequentist and Bart the Bayesianist together conduct an experiment, but evaluate its outcome with different methods. Both compare a null model m_0 and an alternative model m_1. Fred finds a p-value of $p = 0.64$, Bart finds a Bayes factor of $BF_{10} = 0.25$. What can either of them conclude about the null model m_0?