CO902
Supporting Lecture Notes:
Bernouli, Time-invariant, 1st Order Markov Chain
Lecture 2 — 14 Jan 2012

Let $X_1, X_2, \ldots, X_N$ be dependent Bernoulli random variables, such that

$$P(X_i|X_{i-1}, X_{i-2}, \ldots, X_1) = P(X_i|X_{i-1}). \tag{1}$$

This is the 1st order Markov property, also know as a "memoryless" property, and we say $(X_1, X_2, \ldots, X_N)$ is a "Markov Chain". If we further assume that

$$\begin{aligned} P(X_i = 1|X_{i-1} = 0) &= p_{01} \\ P(X_i = 1|X_{i-1} = 1) &= p_{11} \end{aligned}$$

for all $i = 2, 3, \ldots, N$, then the Markov Chain is time-invariant, with p_{01} being the probabilty of a success following a failure and p_{11} being the probabilty of a success following a success. Specifying the initial success probability

$$P(X_1 = 1) \quad = \quad p_0$$

completes the specification of the process.

The (1st order) Markov property allows the joint likelihood to factor into a concise form:

$$\begin{aligned} P(X_1, X_2, \ldots, X_N) &= P(X_N|X_1, X_2, \ldots, X_{N-1}) \times \\ &\quad P(X_{N-1}|X_1, X_2, \ldots, X_{N-2}) \times \\ &\quad \vdots \\ &\quad P(X_2|X_1) \times \\ &\quad P(X_1) \\ &= P(X_N|X_{N-1}) \times \\ &\quad P(X_{N-1}|X_{N-2}) \times \\ &\quad \vdots \\ &\quad P(X_2|X_1) \times \\ &\quad P(X_1) \\ &= P(X_1) \prod_{i=1}^{N} P(X_i|X_{i-1}) \end{aligned}$$

At the risk of incurring confusion, please realize that the above notation with "big" X is just shorthand for

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_N = x_N) \quad = \quad P(X_1 = x_1) \prod_{i=1}^{N} P(X_i = x_i|X_{i-1} = x_{i-1}),$$

that is, this expression is the joint probability mass function that tells you the probability (mass) associated with one particular set of realized values $(x_1, x_2, \ldots, x_N)$.)

The log-likelihood can be written

$$\begin{aligned}\mathcal{L}(p_0, p_{01}, p_{11}) \;=\;& x_1 \log(p_0) + (1 - x_1)\log(1 - p_0) + \\ & \sum_{i\,:\,2 \le i \le N,\, x_{i-1}=0} (x_i \log(p_{01}) + (1 - x_i)\log(1 - p_{01})) + \\ & \sum_{i\,:\,2 \le i \le N,\, x_{i-1}=1} (x_i \log(p_{11}) + (1 - x_i)\log(1 - p_{11}))\end{aligned}$$

There are three components, one for the initial condition, one for the successes preceeded by failure, and one for successes preceeded by successes. (Note I have switched to lower case x, as is often done, reflecting how the likelihood is a function of the parameters *for* a particular value of the observed data.)

Taking derivatives with respect to the parameters, setting to zero and solving gives the Maximum Likelihood Estimators (MLE's). I leave the algrebra to your memory of the chalkboard (make sure you can do it!). The MLE's are

$$\begin{aligned}\hat{p}_0 &= x_1 \\ \hat{p}_{01} &= n_{01}/n_0 \\ \hat{p}_{11} &= n_{11}/n_1\end{aligned}$$

where

$$\begin{aligned}n_0 &= \sum_{i\,:\,2 \le i \le N,\, x_{i-1}=0} 1 && \#\text{ failures } 1, \ldots, N-1 \\ n_1 &= \sum_{i\,:\,2 \le i \le N,\, x_{i-1}=1} 1 && \#\text{ successes } 1, \ldots, N-1 \\ n_{01} &= \sum_{i\,:\,2 \le i \le N,\, x_{i-1}=0} x_i && \#\text{ successes following failures} \\ n_{11} &= \sum_{i\,:\,2 \le i \le N,\, x_{i-1}=1} x_i && \#\text{ successes following successes}\end{aligned}$$

These results will be handy for the lab exercise.

TEN / January 14, 2013