Categorical model

Theoretical derivation of estimation

This model can describe any discrete variable, however, its usage is prone to its extremely high dimension and for larger tasks it is unfeasible. Its general conditional form $f(y_t|u_t, y_{t-1}, p) = p_{y_t|u_t,y_{t-1}}$ can be expressed through a table (here for $y, u \in \{1, 2\}$)

u_t	1	1	2	2
y_{t-1}	1	2	1	2
$y=1$	$p_{1\|11}$	$p_{1\|12}$	$p_{1\|21}$	$p_{1\|22}$
$y=2$	$p_{2\|11}$	$p_{2\|12}$	$p_{2\|21}$	$p_{2\|22}$

where $p_{i|jk}$ are conditional probabilities of $y_t = i$, $u_t = j$ and $y_{t-1} = k$. As y_t is random variable and the values of u_t and y_{t-1}, the probabilities are normalized in columns, i.e. it holds

$$p_{i|jk} \geq 0 \text{ and } \sum_i p_{i|jk} = 1$$

Posterior pdf has Dirichlet form

$$f(p|y(t), u(t)) = \prod_{i|jk} p_{i|jk}^{S_{i|jk;t}}$$

The statistic $S_{i|jk;t}$ is again of the form of table structurally identical with the model one. The entries are denoted by combinations of values of variables y_t, u_t, y_{t-1}. Each entry includes number of occurrences of the particular combination coming from data.

Update of the statistics

$$S_{y_t|u_t y_{t-1};t} = S_{y_t|u_t y_{t-1};t-1} + 1$$

i.e. we increment only the entry of the statistics which corresponds to the actual combination of the values of y_t, u_t, y_{t-1}.

Point estimates are obtained from the statistics by normalization its columns to the sum equal to one.

Generation For generation we can use the method of projection values of $U(0,1)$ through the inverse distribution function. The result is

$$y_t = \text{sum}\left(\text{cumsum}\left(p_{\bullet|u_t y_{t-1}}\right) \leq \text{randu}()\right) + 1$$

where $p_{\bullet|u_t y_{t-1}}$ is the column of the model table corresponding to the combination of values of u_t, y_{t-1}, cumsum() is cumulative sum and randu() is generator of $U(0,1)$.

Program and its description

Back to Main