Regression and exchangeability

Luca
<piero.mana@ntnu.no>

Draft of 25 December 2018 (first drafted 24 December 2018)

Exchangeability as a form of regression, regression as a form of exchangeability.

Note: Dear Reader & Peer, this manuscript is being peer-reviewed by you. *Thank you.*

1 Regression

'Regression', as usually intended, is the problem of inferring the value of a quantity x given that of another quantity t; the two quantities having values in two specific sets. It is also assumed that in two or more observations with the same value of t the values of x may be different.

In this note I use 'regression' in a restricted sense: two or more observations with the same value of t have also the same value of x. So the second quantity is given as a function of the first. This is the case in many classification problems. This function or map $m\colon t \mapsto x$ is unknown, however; we are therefore uncertain about x in every observation with a new value of t. Our goal is to infer the unknown function m, given observations of the pair of quantities for different values of t.

If we assume for the moment that t and x assume finite sets of values, then also the set of functions between them is finite. Hence the plausibility of observing x given t, if the function is unknown, is

$$\mathrm{p}(x|\,t, I) = \sum_m \mathrm{p}(x|\,t, m, I)\,\mathrm{p}(m|\,I), \tag{1}$$

with

$$\mathrm{p}(x|\,t, m, I) = \delta[x - m(t)], \tag{2}$$

which expresses the certain relation between t and x if the function relating them is known.

2 Exchangeability from regression

Consider a sequence of observations of the quantity x. We can use $t = 1, \dots, T$ to simply label these observations. In this case, each function $m\colon t \mapsto x$ simply corresponds to a specific sequence of values of x, and vice versa.

Denote by $\boldsymbol{f} \coloneqq (f_x)$ the relative frequencies with which the values of x appear in a particular sequence m. These frequencies are determined by the function: by counting,

$$f_x = \frac{1}{T} \sum_t \delta[x - m(t)] \qquad \text{for each } x. \tag{3}$$

The assumption of *exchangeability* is the judgement that sequences that differ by a permutation of their order are equally plausible. This means that sequences in which the values of x appear with the same frequencies are equally plausible. Now consider all functions $t \mapsto x$ yielding the same relative frequencies $\boldsymbol{f}$; there are $\binom{T}{T\boldsymbol{f}}$ such functions. Let's judge them to be equally plausible. Then we can write

$$\begin{aligned} \mathrm{p}(m \mid \boldsymbol{f}, I) &= \begin{cases} \binom{T}{T\boldsymbol{f}}^{-1} & \text{if } f_x = \frac{1}{T}\sum_t \delta[x - m(t)] \text{ for each } x, \\ 0 & \text{otherwise,} \end{cases} \\ &= \binom{T}{T\boldsymbol{f}}^{-1} \prod_x \delta\Bigl\{T f_x - \sum_t \delta[x - m(t)]\Bigr\}. \end{aligned} \tag{4}$$

Now use the law of total probability:

$$\mathrm{p}(m \mid I) = \sum_{\boldsymbol{f}} \mathrm{p}(m \mid \boldsymbol{f}, I)\, \mathrm{p}(\boldsymbol{f} \mid I), \tag{5}$$

in $\mathrm{p}(x|\,t, I)$, given by eq. (1):

$$\begin{aligned}
\mathrm{p}(x|\,t, I) &= \sum_m \mathrm{p}(x|\,t, m, I) \ \sum_f \mathrm{p}(m|\,f, I)\ \mathrm{p}(f|\,I), \\
&= \sum_f \sum_m \binom{T}{Tf}^{-1} \ \delta[x - m(t)] \ \prod_{x'} \delta\Bigl\{T f_{x'} - \sum_{t'} \delta[x' - m(t')]\Bigr\} \ \mathrm{p}(f|\,I), \\
&= \sum_f \binom{T}{Tf}^{-1} \ \binom{T-1}{Tf} \ \frac{(T f_x)!}{(T f_x - 1)!} \ \mathrm{p}(f|\,I), \\
&= \sum_f f_x \ \mathrm{p}(f|\,I).
\end{aligned} \tag{6}$$

We have recovered the formula for an exchangeable plausibility. Note in particular that the label of the observation, t, becomes irrelevant, as it should be by symmetry.

Thanks

. . . to Mari & Miri for continuous encouragement and affection, and to Buster Keaton and Saitama for filling life with awe and inspiration. To the developers and maintainers of LaTeX, Emacs, AUCTeX, Open Science Framework, Python, Inkscape, Sci-Hub for making a free and unfiltered scientific exchange possible.