Regression and exchangeability

Luca
<piero.mana@ntnu.no>

Draft of 26 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 $S\colon t \mapsto x$ 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 S, 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 a sequence $\boldsymbol{x} \coloneqq (x_1, \dots)$ given the sequence $\boldsymbol{t} \coloneqq (t_1, \dots)$, if the function is unknown, is

$$\mathrm{p}(\boldsymbol{x} \mid \boldsymbol{t}, I) = \sum_S \mathrm{p}(S \mid I) \prod_i \mathrm{p}(x_i \mid t_i, S, I), \tag{1}$$

with

$$\mathrm{p}(x_i \mid t_i, S, I) = \delta[x_i - S(t_i)], \tag{2}$$

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

This document is designed for screen reading and two-up printing on A4 or Letter paper

2 Exchangeability from regression

Consider the special case where $\boldsymbol{t} \coloneqq (t_1, \dots, t_M)$ simply label the M observations $\boldsymbol{x}$. In this case, each function $S\colon t \mapsto x$ 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 S of length M. The sequence determines the frequencies: by counting,

$$MF_x = \sum_t \delta[x - S(t)] \qquad \text{for each } x. \tag{3}$$

The assumption of *exchangeability* says that the particular order of any subset of observations $(x_1, \dots, x_m)$ is irrelevant for their prediction. This is equivalent to saying that only the frequencies of the values of x matter for our inferences, and that sequences that differ by a permutation of their orders are equally plausible. Consider all functions $t \mapsto x$ yielding the same relative frequencies $\boldsymbol{F}$; there are $\binom{M}{M\boldsymbol{F}}$ such functions. If we judge them to be equally plausible we can write

$$\begin{aligned}
\mathrm{p}(S \mid \boldsymbol{F}, I) &= \begin{cases} \binom{M}{M\boldsymbol{F}}^{-1} & \text{if } MF_x = \sum_t \delta[x - S(t)] \text{ for each } x, \\ 0 & \text{otherwise,} \end{cases} \\
&= \binom{M}{M\boldsymbol{F}}^{-1} \prod_x \delta\Bigl\{MF_x - \sum_t \delta[x - S(t)]\Bigr\},
\end{aligned} \tag{4}$$

which can be used together with the law of total plausibility,

$$\mathrm{p}(S \mid I) = \sum_{\boldsymbol{F}} \mathrm{p}(S \mid \boldsymbol{F}, I)\, \mathrm{p}(\boldsymbol{F} \mid I). \tag{5}$$

Let's see how the particular initial plausibility (5) for the function S leads to the usual formulae for exchangeability.

Consider sequences $\boldsymbol{t} \coloneqq (t_1, \dots, t_m)$ and $\boldsymbol{x} \coloneqq (x_1, \dots, x_m)$ and suppose that in the latter the values of x appear with relative frequencies $\boldsymbol{f} \coloneqq (f_x)$. Using the initial plausibility (5) in the formula for $\mathrm{p}(\boldsymbol{x} \mid \boldsymbol{t}, I)$, eq. (1), we find

$$
\begin{aligned}
\mathrm{p}(\boldsymbol{x} \vert \boldsymbol{t}, I) &= \sum_{S} \mathrm{p}(\boldsymbol{x} \vert \boldsymbol{t}, S, I) \ \sum_{\boldsymbol{F}} \mathrm{p}(S \vert \boldsymbol{F}, I)\, \mathrm{p}(\boldsymbol{F} \vert I), \\
&= \sum_{\boldsymbol{F}} \mathrm{p}(\boldsymbol{F} \vert I) \ \sum_{S} \binom{M}{M\boldsymbol{F}}^{-1} \times \\
&\qquad \Bigl\{\prod_{i} \delta[x_i - S(t_i)]\Bigr\} \ \prod_{x'} \delta\Bigl\{T f_{x'} - \sum_{t'} \delta[x' - S(t')]\Bigr\} \\
&= \sum_{\boldsymbol{F}} \binom{M}{M\boldsymbol{F}}^{-1} \binom{M-m}{M\boldsymbol{F} - m\boldsymbol{f}}\, \mathrm{p}(\boldsymbol{F} \vert I), \\
&= \sum_{\boldsymbol{F}} \Biggl[\prod_{x} \binom{MF_x}{m f_x}\Biggr]\, \mathrm{p}(\boldsymbol{F} \vert I).
\end{aligned}
\tag{6}
$$

In passing from the first to the second equality we have exchanged the sums, replaced eqs (2) and (4), and rearranged the terms. To pass from the second to the third equality, consider the sum over S for fixed $\boldsymbol{F}$: the second product of deltas restricts the sum to functions S yielding frequencies $M\boldsymbol{F}$; the first product of deltas further restricts the sum to functions having frequencies $m\boldsymbol{f}$ in a subsequence of length m. There are as many such functions as the number of ways to order the remaining $M - m$ observations of x, its values appearing $M\boldsymbol{F} - m\boldsymbol{f}$ times. This is just the multinomial coefficient $\binom{M-m}{M\boldsymbol{F}-m\boldsymbol{f}}$. Thus in the third line we find a combination of hypergeometric distributions, which is the formula for finite exchangeability and, as M increases indefinitely, for infinite exchangeability,

$$
\mathrm{p}(\boldsymbol{x} \vert \boldsymbol{t}, I) \to \int \mathrm{d}\boldsymbol{F} \, \Bigl(\prod_{x} {F_x}^{m f_x}\Bigr)\, \mathrm{p}(\boldsymbol{F} \vert I). \tag{7}
$$

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.