Posteriors for sufficiency hypotheses and maximum-entropy

P.G.L. Porta Mana
<piero.mana@ntnu.no>

Draft of 23 February 2019 (first drafted 16 February 2019)

***

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

We have a population of N neurons whose activities we imagine to have time-binned into T bins and binarized. Denote their total population activity at time bin t by $S_t \in \{0, 1, \dots, N\}$, their total activity at an unspecified bin by S, and the time sequence of total activities by $\boldsymbol{S} \coloneqq (S_{t_1}, S_{t_2}, \dots, S_{t_T})$.

We have recorded the activities of a sample of n neurons from the population above. Denote the total activity of the sample at time bin t by $s_t \in \{0, 1, \dots, n\}$, at an unspecified bin by s, and the time sequence by $\boldsymbol{s} \coloneqq (s_{t_1}, s_{t_2}, \dots, s_{t_T})$.

We don't know how these sampled neurons were chosen from the full population. This fact leads, for each time bin, to the following degree of belief about the activity of the sample if we knew the activity of the full population (Porta Mana et al. 2015 § 2.3; 2018 § 2):

$$\mathrm{p}(s \mid S, I) = \binom{n}{s}\binom{N-n}{S-s}\binom{N}{S}^{-1} \eqqcolon H_{sS}, \tag{1}$$

namely, a hypergeometryc distribution.

Here and in the following I denotes the proposition stating our background information.

Now suppose that we knew the total activities $\boldsymbol{S}$ of the *full* population at some T time bins $\{t\}$, and we wanted to infer the total activities $\boldsymbol{S}'$ at T' *different* time bins $\{t'\}$:

$$\mathrm{p}(\boldsymbol{S}' \mid \boldsymbol{S}, I). \tag{2}$$

We want to consider the hypotheses that *only a specific set of statistics* about our data $\boldsymbol{S}$ are *relevant* for our inference about $\boldsymbol{S}'$; that is, they are a sufficient statistics. Any aspect of the data not contained in those statistics would be irrelevant for our inference. This inferential property could be the result of biological properties of the population.

Let's assume that there are R such statistics (besides T, which is always part of a set of sufficient statistics). Each statistic is the sum over time of a specific function of the total activity S. We can arrange these functions in an R-by-$(N+1)$ matrix $\boldsymbol{C} \coloneqq (C_{rS})$, where C_{rS} is the value of the function for the rth statistic when the total activity is S. The R sufficient statistics for the data $\boldsymbol{S}$ would thus be

$$\overline{C}_r \coloneqq \frac{1}{T}\sum_t C_{rS_t}, \quad r \in \{1, \dots, R\}. \tag{3}$$

Our goal is to quantify our uncertainty about these hypotheses of sufficient statistics, given the activity data from a sample of neurons. It's important to note that the hypotheses we must consider are not discrete or of a yes-or-no type: they form a continuum. This is because we have a continuum of degrees of relevance. Consider for example two statistics $\overline{C}_1$ and $\overline{C}_2$ from the bins $\{t\}$. Our degrees of belief about the activities at bins $\{t'\}$ are

$$\mathrm{p}(\boldsymbol{S}' \mid \overline{C}_1, \overline{C}_2, I). \tag{4}$$

It may happen that lack of knowledge about $\overline{C}_2$ doesn't change our degree of belief:

$$\mathrm{p}(\boldsymbol{S}' \mid \overline{C}_1, I) = \mathrm{p}(\boldsymbol{S}' \mid \overline{C}_1, \overline{C}_2, I), \tag{5}$$

in which case $\overline{C}_2$ is irrelevant. It may also happen that our degree of belief is changed but in a negligible way, for all values of $\boldsymbol{S}'$ and $\overline{C}_1$:

$$\mathrm{p}(\boldsymbol{S}' \mid \overline{C}_1, I) \approx \mathrm{p}(\boldsymbol{S}' \mid \overline{C}_1, \overline{C}_2, I), \tag{6}$$

so that $\overline{C}_2$ could be dropped in practice. We can imagine larger and larger changes to the point where dropping $\overline{C}_2$ would lead to drastically different degrees of belief. The question of the relevance of $\overline{C}_2$ is therefore not dichotomous. We will thus deal with a continuum of hypotheses, each representing a degree of relevance of some statistics. We shall shortly see how to mathematically represent this continuum of hypotheses.

How does a hypothesis about a sufficient statistic affect our degrees of belief? The answer comes from the Koopman-Pitman theorem (Koopman

1936; Pitman 1936; see also Darmois 1935; Barankin et al. 1963; Denny 1967; Hipp 1974; Lauritzen 1974; 1984; 1988; for the discrete version: Fraser 1963; Andersen 1970), which says that the degree of belief (2) has a very specific mathematical expression if only some statistics of $\boldsymbol{S}$ relevant. The main statement of the theorem is this: if R sufficient statistics are given by functions C_{rS}, then for any number of time bins T

$$\mathrm{p}(\boldsymbol{S} \mid I) = \int \mathrm{d}\boldsymbol{\lambda}\; \mathrm{p}(\boldsymbol{\lambda} \mid I) \prod_t \frac{g_{S_t}}{Z(\boldsymbol{\lambda})} \exp\bigl(\textstyle\sum_r \lambda_r C_{rS_t}\bigr) \tag{7a}$$

with

$$\boldsymbol{\lambda} \coloneqq (\lambda_1, \dots, \lambda_R) \in \mathbf{R}^R, \qquad Z(\boldsymbol{\lambda}) \coloneqq \sum_S g_S\; \exp\bigl(\textstyle\sum_r \lambda_r C_{rS}\bigr), \tag{7b}$$

and g a positive function of S.

Some important remarks about the Pitman-Koopman formula (7):

a. A hypothesis that only stated what the sufficient statistics are would not determine the density $\mathrm{p}(\boldsymbol{\lambda} \mid I)$ or the function g in the formula above. The hypotheses we are going to compare thus contain additional information besides sufficiency.

b. The formula can be interpreted this way: our degree of belief about $\boldsymbol{S}$ is given by the degree of belief we would have if we knew the values of the sufficient statistics $\overline{\boldsymbol{C}}$ for an unlimited number of time bins, mixed over our uncertainty about the values themselves:

$$\mathrm{p}(\boldsymbol{S} \mid I) = \int \mathrm{d}\overline{\boldsymbol{C}}\; \mathrm{p}(\overline{\boldsymbol{C}} \mid I)\, \mathrm{p}(\boldsymbol{S} \mid \overline{\boldsymbol{C}}, I). \tag{8}$$

c. Formula (7) is obtained from the latter one by the one-to-one reparametrization

$$\overline{C}_r(\boldsymbol{\lambda}) = \sum_S C_{rS} \frac{g_S}{Z(\boldsymbol{\lambda})} \exp\bigl(\textstyle\sum_r \lambda_r C_{rS}\bigr) \equiv \partial_{\lambda_r} \ln Z(\boldsymbol{\lambda}) \tag{9}$$

This equation cannot be solved explicitly for $\boldsymbol{\lambda}$ in terms of $\overline{\boldsymbol{C}}$. The parametrization in terms of $\boldsymbol{\lambda}$ has several special properties:

c.1. The quantities (λ_r) can assume any values independently of one another, whereas the limit statistics $\overline{\boldsymbol{C}}$ have interdependent ranges.

c.2. The expression for $\mathrm{p}(\boldsymbol{S} \mid \overline{\boldsymbol{C}}, I)$ can be written explicitly in terms of $\boldsymbol{\lambda}$, but not in terms of $\overline{\boldsymbol{C}}$.

c.3. If some λ_r vanishes then the corresponding statistic is *irrelevant* – the corresponding term indeed disappears from the exponential in formula (7).

Remark **??** shows that the absolute value of each parameter, $\lvert\lambda_r\rvert$, somehow quantifies the degree of relevance of the corresponding statistic, zero meaning complete irrelevance. These absolute values can therefore be used as parameters for our continuum of hypotheses about how much each statistics is sufficient.

Bibliography

('de X' is listed under D, 'van X' under V, and so on, regardless of national conventions.)

Andersen, E. B. (1970): *Sufficiency and exponential families for discrete sample spaces*. J. Am. Stat. Assoc. **65**^{331}, 1248–1255.

Barankin, E. W., Maitra, A. P. (1963): *Generalization of the Fisher-Darmois-Koopman-Pitman theorem on sufficient statistics*. Sankhyā A **25**^{3}, 217–244.

Darmois, G. (1935): *Sur les lois de probabilité à estimation exhaustive*. Comptes rendus hebdomadaires des séances de l'Académie des sciences **200**, 1265–1266.

Denny, J. L. (1967): *Sufficient conditions for a family of probabilities to be exponential*. Proc. Natl. Acad. Sci. (USA) **57**^{5}, 1184–1187.

Fraser, D. A. S. (1963): *On sufficiency and the exponential family*. J. Roy. Stat. Soc. B **25**^{1}, 115–123.

Hipp, C. (1974): *Sufficient statistics and exponential families*. Ann. Stat. **2**^{6}, 1283–1292.

Koopman, B. O. (1936): *On distributions admitting a sufficient statistic*. Trans. Am. Math. Soc. **39**^{3}, 399–409.

Lauritzen, S. L. (1974): *Sufficiency, prediction and extreme models*. In: **barndorffnielsenetal1974**, 249–269. With discussion. Repr. without discussion in **lauritzen1974_r1974**.

— (1984): *Extreme point models in statistics*. Scand. J. Statist. **11**^{2}, 65–83. See also discussion and reply in **barndorffnielsenetal1984**.

— (1988): *Extremal Families and Systems of Sufficient Statistics*. (Springer, Berlin). First publ. 1982.

Pitman, E. J. G. (1936): *Sufficient statistics and intrinsic accuracy*. Math. Proc. Camb. Phil. Soc. **32**^{4}, 567–579.

Porta Mana, P. G. L., Rostami, V., Torre, E., Roudi, Y. (2018): *Maximum-entropy and representative samples of neuronal activity: a dilemma*. Open Science Framework doi:10.17605/osf.io/uz29n, bioRxiv doi:10.1101/329193, arXiv:1805.09084.

Porta Mana, P. G. L., Torre, E., Rostami, V. (2015): *Inferences from a network to a subnetwork and vice versa under an assumption of symmetry*. bioRxiv doi:10.1101/034199.