The Pitman-Koopman theorem and informationally relevant quantities

research notes

Y. Roudi
<yasser.roudi@ntnu.no>

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

23 June 2018; updated 25 June 2018

Application of the concept of sufficiency and the Pitman-Koopman theorem to the definition and measurement of 'cooperativity' among neurons

> I believe that the greatest unsolved problem in statistics is communicating the subject to others.
>
> N. Davies (Copas et al. 1995 p. 445)

1 From qualitative to quantitative statements

We often have the need of quantifying a qualitative property or phenomenon, for example to ascertain whether it does or doesn't occur in some situations. This quantification is usually difficult because the property itself is only vaguely defined in our minds.

At times we sweep this difficulty under the carpet by choosing a quantitative measure that 'feels right' and proceeding with the calculations. It's obvious that the quantitative results we obtain this way are mostly smoke: despite the deceiving presence of numbers, our results are only qualitative, since our starting point was vague and our quantitative measure wasn't unique. Which results would we have obtained if we had started from a slightly different qualitative intuition or used a slightly different quantitative measure?

Other times we choose several quantitative measures that 'feel right', proceed with the calculations, and then compare the quantitative results obtained from all these measures. Results obtained this way are more robust – for example if all the chosen measures lead to the conclusion that the property is there – but unfortunately still not unequivocally quantitative.

In science there are successful examples of qualitative definitions that, formulated as a set of desiderata, lead to unique quantitative measures or theories. The most impressive example is probably that of the probability calculus: simple intuitive desiderata about reasoning in situations of uncertainty, such as those examined by Pólya (1971; 1949; 1954; 1968), turn out to have a unique mathematical translation: the laws of the probability calculus (Cox 1946; Paris 2006; Jaynes 2003 ch. 2; Snow 1998; 2001; Dupré et al. 2009; Terenin et al. 2017; see also Aczél 2004). Another example is Shannon's entropy (Shannon 1948 § 6) as a measure of uncertainty, although Shannon's desiderata explicitly include the use of addition as the mathematical representation of 'combination' (cf. Jaynes 2003 § 11.3, p. 347), a requirement which is not universally agreed upon. In physics, semi-quantitative invariance requirements can often be mathematically translated into group-invariance statements leading to fully fledged theories; classical examples being Einstein's requirements (Einstein 1905 § 2) leading to the Lorentz transformations and special relativity, or Noether's theorem (Noether 1918; anticipated for the Euclidean group by Cosserat et al. 1909).

✠ to be continued

'A bewildering variety of statistical measures has been reported in the literature on spike-train analysis. Unfortunately, there has been little theoretical justification for the use and interpretation of many of these techniques' (Moore et al. 1966 p. 494)

2 Sufficient statistics and the Koopman-Pitman theorem

Suppose we have a sequence, possibly infinite, of observations with outcomes in a set K. We represent the proposition 'The ith observation yields outcome k' by $O_k^{(i)}$, with $k \in K$.

A typical inference problem is to forecast the outcomes of the $(m+1)$th up to the $(m+n)$th observations given that we know the outcomes of the 1st up to the mth observations and given some additional knowledge or assumptions denoted by I. Note that we aren't assuming that these $\{1, \dots, m+n\}$ observations are in temporal order: our problem could be to infer past, unknown observation from present, known ones. Our forecast is expressed by the probabilities

$$\mathrm{P}(O_{k_{m+n}}^{(m+n)}, \dots, O_{k_{m+1}}^{(m+1)} \,|\, O_{k_m}^{(m)}, \dots, O_{k_1}^{(1)}, I), \tag{1}$$

where the comma represents the conjunction ('∧') of the propositions. We will also use the simplified notation $\mathrm{p}(k_{m+n}, \dots, k_{m+1} \,|\, k_m, \dots, k_1, I)$.

In general the full set of observed data $(k_1, \dots, k_m)$ is important for this inference. But in particular situations it may happen that only specific properties of the data are *sufficient* to our inference, the remaining details being *irrelevant*. By 'irrelevant' we mean that our inferences aren't changed if we forget about those details or if we don't know them in the first place. Let's represent the relevant properties by a map $s(k_1, \dots, k_m)$, called a *sufficient statistic*, with values in some space S. The relevance and irrelevance are expressed by the equalities

$$\mathrm{p}(k_{m+n}, \dots, k_{m+1} \,|\, k_m, \dots, k_1, I) = \mathrm{p}[k_{m+n}, \dots, k_{m+1} \,|\, s(k_1, \dots, k_m), I] \qquad \text{for all } m, n, \text{ and } (k_i). \tag{2}$$

Typical examples of sufficient statistics are the number of known observations m together with their mean $(k_1 + \dots + k_m)/m$, or second moments $(\boldsymbol{k}_1\boldsymbol{k}_1{}^\intercal + \dots + \boldsymbol{k}_m\boldsymbol{k}_m{}^\intercal)/m$, or range $\max_i(k_i) - \min_i(k_i)$; for example

$$s(k_1, \dots, k_m) \coloneqq \Bigl(m, \tfrac{1}{m} \textstyle\sum_i k_i\Bigr). \tag{3}$$

The number m is usually implied and left out of the definition.

There may be physical or biological reasons why only some statistics of past data are sufficient for our inference. This sufficiency may also be only approximate. We can hypothesize that one or another statistics are sufficient – call these hypotheses $H_1, H_2, \dots$ – and, given some data D, the probability calculus tells us which hypotheses among these are more probable:

$$\mathrm{P}(H_i \,|\, D, I) = \frac{\mathrm{P}(D \,|\, H_i, I)\, \mathrm{P}(H_i \,|\, I)}{\sum_i \mathrm{P}(D \,|\, H_i, I)\, \mathrm{P}(H_i \,|\, I)}. \tag{4}$$

To calculate these probabilities we need to assign numerical values to the probabilities of the data given the hypotheses, $\mathrm{P}(D \,|\, H_i, I)$.

The Koopman-Pitman-Darmois theorem (Koopman 1936; Pitman 1936; Darmois 1935; see also later analyses: Hipp 1974; Andersen 1970; Denny 1967; Fraser 1963; Barankin et al. 1963) ✠ add comments about these references and about the discrete-variable case

Bibliography

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

Aczél, J. (2004): *The associativity equation re-revisited*. Am. Inst. Phys. Conf. Proc. **707**, 195–203.

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.

Chatfield, C. (1995): *Model uncertainty, data mining and statistical inference*. J. Roy. Stat. Soc. A **158**^{3}, 419–444. See also discussion in Copas, Davies, Hand, Lunneborg, Ehrenberg, Gilmour, Draper, Green, et al. (1995).

Copas, J. B., Davies, N., Hand, D. J., Lunneborg, C. E., Ehrenberg, A. S., Gilmour, S. G., Draper, D., Green, P. J., et al. (1995): *Discussion of the paper by Chatfield [Model uncertainty, data mining and statistical inference]*. J. Roy. Stat. Soc. A **158**^{3}, 444–466. See Chatfield (1995).

Cosserat, E., Cosserat, F. (1909): *Théorie des corps déformables*. (A. Hermann et fils, Paris). http://ebooks.library.cornell.edu/cgi/t/text/text-idx?c=math;idno=06420001.

Cox, R. T. (1946): *Probability, frequency, and reasonable expectation*. Am. J. Phys. **14**^{1}, 1–13. http://algomagic.org/ProbabilityFrequencyReasonableExpectation.pdf.

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.

Dupré, M. J., Tipler, F. J. (2009): *New axioms for rigorous Bayesian probability*. Bayesian Anal. **4**^{3}, 599–606.

Einstein, A. (1905): *Zur Elektrodynamik bewegter Körper*. Ann. der Phys. **17**, 891–921. Transl. in Einstein (1989a), Doc. 23, pp. 140–171.

— (1989a): *The Collected Papers of Albert Einstein. Vol. 2: The Swiss Years: Writings, 1900–1909. (English translation)*. (Princeton University Press, Princeton). Transl. of Einstein (1989b) by Anna Beck and Peter Havas.

— (1989b): *The Collected Papers of Albert Einstein. Vol. 2: The Swiss Years: Writings, 1900–1909*. (Princeton University Press, Princeton). Ed. by John Stachel. Transl. in Einstein (1989a).

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.

Jaynes, E. T. (2003): *Probability Theory: The Logic of Science*. (Cambridge University Press, Cambridge). Edited by G. Larry Bretthorst. First publ. 1994. https://archive.org/details/XQUHIUXHIQUHIQXUIHX2, http://www-biba.inrialpes.fr/Jaynes/prob.html, http://omega.albany.edu:8008/JaynesBook.html.

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

Moore, G. P., Perkel, D. H., Segundo, J. P. (1966): *Statistical analysis and functional interpretation of neuronal spike data*. Annu. Rev. Physiol. **28**, 493–522.

Noether, E. (1918): *Invariante Variationsprobleme*. Ges. d. Wiss. Nachrichten (Gött.) Math.-Phys. Kl. **1918**, 235–257. http://cwp.library.ucla.edu/articles/noether.trans/german/emmy235.html. Transl. in Noether (1971).

— (1971): *Invariant variation problems*. Transport Theory Stat. Phys. **1**^{3}, 186–207. http://cwp.library.ucla.edu/articles/noether.trans/english/mort186.html. Transl. by M. A. Tavel of Noether (1918).

Paris, J. B. (2006): *The Uncertain Reasoner's Companion: A Mathematical Perspective*, reprint. (Cambridge University Press, Cambridge). See also Snow (1998).

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

Pólya, G. (1949): *Preliminary remarks on a logic of plausible inference*. Dialectica **3**$^{1-2}$, 28–35.

— (1954): *Mathematics and Plausible Reasoning: Vol. I: Induction and Analogy in Mathematics*. (Princeton University Press, Princeton).

— (1968): *Mathematics and Plausible Reasoning: Vol. II: Patterns of Plausible Inference*, 2nd ed. (Princeton University Press, Princeton). First publ. 1954.

— (1971): *How To Solve It: A New Aspect of Mathematical Method*, 2nd ed. (Princeton University Press, Princeton). First publ. 1945.

Shannon, C. E. (1948): *A mathematical theory of communication*. Bell Syst. Tech. J. **27**^{3,4}, 379–423, 623–656. https://archive.org/details/bstj27-3-379, https://archive.org/details/bstj27-4-623, http://math.harvard.edu/~ctm/home/text/others/shannon/entropy/entropy.pdf.

Snow, P. (1998): *On the correctness and reasonableness of Cox's theorem for finite domains*. Comput. Intell. **14**^{3}, 452–459.

— (2001): *The reasonableness of possibility from the perspective of Cox*. Comput. Intell. **17**^{1}, 178–192.

Terenin, A., Draper, D. (2017): *Cox's theorem and the Jaynesian interpretation of probability*. arXiv:1507.06597. First publ. 2015.