Sufficiency, ancillarity, and all that

← Back to Chapters

A few questions.

1) Does there exist an "antisymmetric kernel theory" akin to that of RKHS theory? I.e. when is an antisymmetric function $K(\theta, \theta')$ given by applying $\phi : \Theta \to V$ to θ and θ' and then evaluating antisymmetric bilinear form ω on the images, i.e. $K(\theta, \theta') = \omega(\phi(\theta), \phi(\theta'))$.

2) Is 8.6 of that form?

3) Is that of use for anything?

Random things about Fisher information.

We assume various regularity conditions.

0) If $p_\theta(x)$ is a family of probability distributions,

$E[\frac{\partial \log p_\theta}{\partial \theta_i}] = \int \frac{\partial p_\theta}{\partial \theta_i} \frac{1}{p_\theta} p_\theta dx = \int \frac{\partial p_\theta}{\partial \theta_i} dx = \frac{\partial}{\partial \theta_i} \int p_\theta dx = 0$

1) We define $I(\theta) = E[\frac{\partial \log p(x|\theta)}{\partial \theta_i} \frac{\partial \log p(x|\theta)}{\partial \theta_j}]$ to be the covariance of the derivative of score function. Then under some regularity conditions,

$$I(\theta) = -E[\frac{\partial \log p(x|\theta)}{\partial \theta_i \partial \theta_j}],$$

minus the Hessian of the score.

Proof: Abusing notation by writing $f_i(\theta) = \frac{\partial f}{\partial \theta_i}$ and p for $p(x|\theta)$ we have

$(\log p)_i = \frac{p_i}{p}$,

$(\log p)_{ij} = \left(\frac{p_i}{p}\right)_j = \frac{p_{ij}}{p} - \frac{p_i p_j}{p^2} = \frac{p_{ij}}{p} - (\log p)_i (\log p)_j$

So we just need to show that $E[\frac{p_{ij}}{p}] = 0$. This is similar to the previous item. Here goes:

$\int_X \frac{p_{ij}}{p} p dx = \int_X p_{ij} dx = (\int_X p dx)_{ij} = 1_{ij} = 0$

2) From Wikipedia: Suppose X has pdf $f_\theta(X)$, and $T = T(X)$ of θ has pdf $g_\theta(T)$. Let $I(\theta)$ be the Fisher information of f_θ and $J(\theta)$ be the Fisher information of g_θ. Then $I(\theta) \geq J(\theta)$ (meaning difference is positive semidefinite) with equality if and only if T is sufficient.

Proof:

Write

$f_\theta(X) = p_\theta(X|T)g_\theta(T)$

$(\log f_\theta(X))_i = (\log p_\theta(X|T))_i + (\log g_\theta(T))_i$

By the factorization theorem, T is sufficient precisely when $p_\theta(X|T)$ is constant (in θ) i.e. the first term of the right hand side is zero. In that case the two Fisher information matrices are indeed identical. The result we are after follows from the fact that the two terms on the right hand side are uncorrelated. Since they have zero mean (by the 0th item), we covariance is expectation of product. We compute by conditioning on T and using item 0 again:

$$E_X[(\log p_\theta(X|T))_i(\log g_\theta(T))_j] =$$

$$E_T[E_{X|T}[(\log p_\theta(X|T))_i(\log g_\theta(T))_j]]$$

$$E_T[(\log g_\theta(T))_j E_{X|T}[(\log p_\theta(X|T))_i]]$$

$$E_T[(\log g_\theta(T))_j 0] = 0$$

as wanted.

Remark: This last computation is related to the one in Chain rule for mutual information.

Section 8.10.1

This line of reasoning about "disjunction of elementary 'atomic' propositions" seems equally inappropriate as a criticism of either Bayesean or frequentist probability, as long as one keeps in mind distinction between events (which have probabilities, but are often not 'atomic'), and outcomes (which are 'atomic', but often have no probability). See section 8.11 instead.

Section 8.12.4

This is exactly the wrong example to complain about "isolated clever tricks" – setting up and analysing a Markov chain is a fairly general method to solve similar probability problems, well connected to other key areas of probability theory. This serves as an ironic illustration of a deeper point - many clever tricks

when well understood become powerful methods, much more powerful indeed than straightforward but uninspiring computations.

There is less disagreement here than may at first appear. I'm all for "general mathematical techniques which will work not only on our present problem, but on hundreds of others"; it's just that your current "general technique" may solve a given problem, but not explain what is going on in it (Paul Zeitz calls this "How vs. Why"). A clever trick may lead you to a better general theory, closer to answering the "why" question — as indeed the Peter and Paul coin tossing example illustrates. So, no to gamesmanship, yes to bringing the game to the next level.