Ignorance priors and transformation groups

← Back to Chapters

Comments on 12.4

I find the discussion in 12.4 somewhat off the mark. "Statistical decision theory and bayesian analysis" by James O. Berger is a better reference. The summary below is largely based on it, particularly sections 3.3.2 and 6.6.

The word "invariance" presupposes a group action. The most natural setting is that in which a group G acts on the space X in which we get data.

Example 1: (location-scale in 1D) Take $X = \mathbb{A}^1$ the affine line, and G the group of (orientation-preserving) affine transformations of X. As is common (for example in computer graphics), we represent elements of X as vectors $\vec{x} = \begin{pmatrix} x \\ 1 \end{pmatrix}$ 2 (the line $y = 1$ inside $\mathbb{R}^2$) and then is indeed affine-linear. The product in G is then:

$$\begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a' & b' \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} aa' & ab' + b \\ 0 & 1 \end{pmatrix}$$

This is a non-commutative group. We have the following pair of group homomorphisms (known as a short exact sequence):

$$\mathbb{R} \hookrightarrow G \twoheadrightarrow \mathbb{R}_+$$

Here $\mathbb{R}$ is the real numbers under addition, and $\hookrightarrow$ sends a number to a translation by that amount; on the other hand $\mathbb{R}_+$ is positive reals under multiplication, and $\twoheadrightarrow$ sends an affine map to its stretching factor.

These maps do not depend on any particular way of representing X and G. On the other hand, there are "backward maps" $\mathbb{R}_+ \to G$ and $G \to \mathbb{R}$, but those are depend on additional choices, like the choice to represent X and G in terms of vectors

Example 2: (Location-scale in higher dimensions) When $X = \mathbb{A}^n$ the location-scale group G is the subgroup of those affine transforms of X whose linear part is a pure rescaling by a (positive) factor. We then have $\mathbb{R}^n \hookrightarrow G \twoheadrightarrow \mathbb{R}_+$.

We are interested in distribution of the data, i.e. in probability distributions over X. Thus we consider a collection of $\mathcal{P}$ of distributions over X.

Defintion: The collection $\mathcal{P}$ is said to be invariant under the action of G if for any $p \in P$ and any $g \in G$ the pushforward distribution $g_* p$ is also in $\mathcal{P}$.

Since $(g \cdot h)_* p = g_*(h_* p)$, this means that G acts on $\mathcal{P}$ as well. If $\mathcal{P}$ is a parametric family, parametrized by space Θ then weconclude that G acts on Θ.

Example 1 continued:

a) Let $\mathcal{P}$ be the collection of all normal distributions on $\mathbb{A}^1$. Note that in order to talk about the collection $\mathcal{P}$ specifying the origin is not necessary. This collection is a invariant under the action of G. If we pick an origin, then we can use mean μ and standard deviation σ as parameters for $\mathcal{P}$. We can also use representation of G by matrices that we have discussed above. Then $\Theta = \mathbb{R} \times \mathbb{R}_+$ is the parameter space and G acts on by sending (μ, σ) to $(a\mu + b, a\sigma)$.

(Together with x being sent to $ax + b$ this appears as 12.30 in Jaynes.)

b) Let $\mathcal{P}$ be the collection of all normal distributions on $\mathbb{A}^1$; after choice of the origin this is the family with pdfs $\frac{1}{\pi\gamma\left[1+\left(\frac{(x-x_0)}{\gamma}\right)^2\right]}$. We no longer have mean or standard deviation available as a parameters, but we do have location x_0 and scale γ. They transform under G by the same formulas as (μ, σ) did before.

c) Let $\mathcal{P}$ be the collection of all mixures of normal distributions on $\mathbb{A}^1$. After picking the origin, this is the collection of distributions which can be written as $p = \sum_i^m w_i \mathcal{N}(\mu_i, \sigma_i^2)$ for some $m \in \mathbb{N}$ and $w_i > 0$ with $\sum w_i = 1$. This collection is invariant under G. When m is fixed the supbfamily $\mathcal{P}_m$ it is parametric, with parameters being $3m$ dimensional vectors (μ_i, σ_i, w_i). If m is not fixed, however, the collection $\mathcal{P}$ is not parametric in the usual sense of the word.

Example 3: Consider the family $\mathcal{P}$ of all Gamma distributions on $X = \mathbb{R}_+$, with pdfs $\Gamma_{(k,\theta)}(x) = \frac{1}{\Gamma(k)\theta}(\frac{x}{\theta})^{k-1} \exp(-\frac{x}{\theta})$. Here $G = \mathbb{R}+$ acts on X by multiplication, $\mathcal{P}$ is invariant, and the action of $a \in G$ sends (k, θ) to $(k, a\theta)$.
